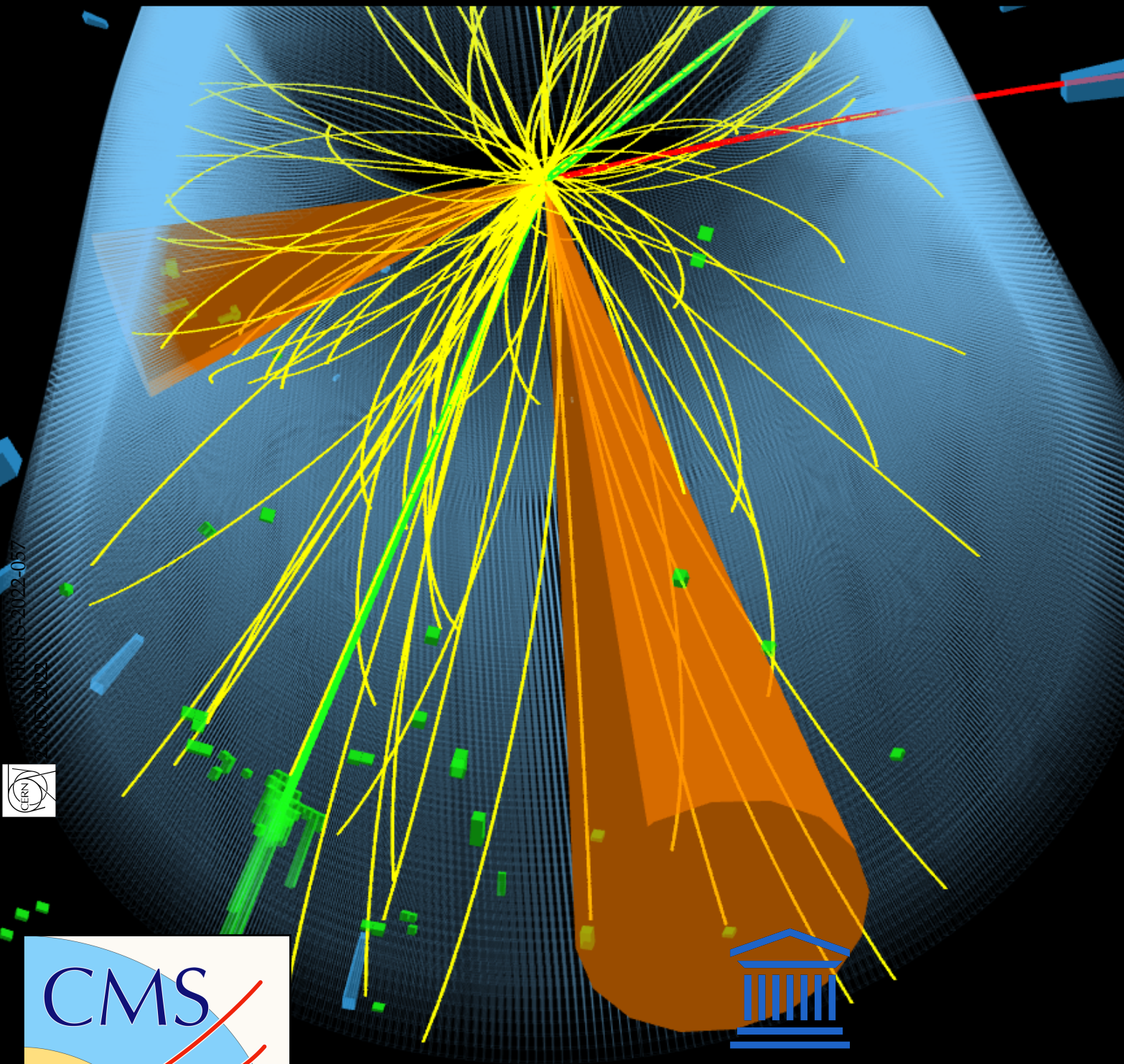
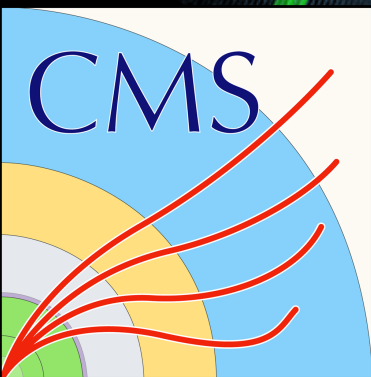


Searches for undiscovered processes using the
multilepton final state in proton-proton
collisions at CMS

Willem Verbeke



CMSSW_515-2022-057
2022-05-2022



UNIVERSITEIT
GENT

Cover image: Display of a proton-proton collision event collected with the CMS detector that is highly likely to originate from single top quark production in association with a Z boson.



Searches for undiscovered processes using the multilepton final state in proton-proton collisions at CMS

Willem Verbeke

Promotor: Didar Dobur

Dissertation submitted in fulfillment of the requirements for the degree of
Doctor (Ph.D.) in Science: Physics

Ghent University
Faculty of Sciences
Department of Physics and Astronomy
2016-2022

Members of the examination committee

Promotors

prof. dr. Didar Dobur (Universiteit Gent, *promotor*)

prof. dr. Filip Moortgat (CERN, *copromotor*)

Chair

prof. dr. Jutho Haegeman (Universiteit Gent)

Committee

prof. dr. Dirk Ryckbosch (Universiteit Gent)

dr. Pieter Everaerts (University of Wisconsin)

prof. dr. Miha Nemevsek (University of Ljubljana)

prof. dr. Nick Van Remortel (Universiteit Antwerpen)

prof. dr. Archisman Ghosh (Universiteit Gent)

Abstract

*“ You summoned me, Odin,
to tell what I recall
of the oldest deeds
of gods and men.”*

Unknown, *Poetic Edda*

The standard model (SM) of particle physics represents our most fundamental knowledge of nature. The model leaves several pressing questions unanswered, hinting at the existence of new physics. This dissertation is a small part of the large CMS Collaboration effort to answer these questions.

Three distinct searches for undiscovered processes in 13 TeV proton-proton collision data collected with the CMS detector at the CERN LHC are presented. The searches target diverse physics, but are experimentally closely related. All of the searches use events with multiple charged leptons in the final state, with a particular focus on events with three leptons.

The first search being presented is a search for sterile Majorana neutrinos in events with three charged light leptons, electrons or muons. It is the first sterile neutrino search in the multilepton final state at the LHC. This final state enables the study of sterile neutrinos in the mass range below 40 GeV, a first at a hadron collider. The sensitivity extends downwards to sterile neutrino masses of 1 GeV. In addition, the search marks the first time an analysis is able to probe sterile neutrino masses beyond 500 GeV, achieving sensitivity all the way up to 1.2 TeV. No excess over the SM hypothesis is found, and mass-dependent exclusion limits are placed on the square of the mixing parameter between sterile neutrinos and electron and muon neutrinos. The limits range from $1.2 \cdot 10^{-5}$ to 1 in the sterile neutrino mass range 1 GeV–1.2 TeV.

Second is a search for single top quark production in association with a Z boson (tZq), an extremely rare SM process with an extraordinary sensitivity to new physics. Being about 50 times more rare than Higgs boson production in 13 TeV proton-proton collisions, tZq production remained undiscovered prior to the work presented here. The search analyzes events with three light leptons and at least one jet identified as originating from a b hadron decay. The use of a machine learning based lepton identification algorithm and a redesigned analysis strategy massively increase the sensitivity of the search compared to those that came before. This results in the first observation of tZq production with an observed (expected) significance of 8.2 (7.7) standard deviations over the background-only hypothesis. The tZq production cross section is measured with a relative uncertainty of 15%, more than twice as precise as any earlier measurement.

Lastly, a search for the direct electroweak production of charginos and neutralinos, superpartners of the electroweak gauge and Higgs bosons, is presented. The range of possible chargino and neutralino production and decay modes is vast. Events with three or more leptons, including up to two hadronically decaying τ leptons, as well as events with two light leptons of the same sign are analyzed. Events are divided into a large number of orthogonal categories by their lepton flavor and sign content. Each event category is sensitive to particular manifestations of supersymmetry. Events with three light leptons, at least two of which form an opposite-sign and same-flavor pair, provide the bulk of the search sensitivity to several important signal models. Unfortunately, the majority of SM background events also falls under this category. Several parametric neural networks are used to separate signal

and background events in this region, vastly increasing the reach of the search over earlier results. This is the first time parametric neural networks are used in such a search. The signal extraction strategy in other event categories is also entirely redesigned compared to earlier work in CMS, resulting in a far greater sensitivity to all studied models of superpartner production. No excess of events over the SM expectation is observed, and chargino masses up to 1450 GeV are excluded under certain conditions. This is currently the overall highest excluded mass for an electroweak superpartner, under any production or decay hypothesis.

Samenvatting

Het standaardmodel van de deeltjesfysica geeft onze meest fundamentele kennis van de natuur weer. Desalniettemin blijven verschillende belangrijke vraagstukken onbeantwoord in het standaardmodel, wat doet vermoeden dat er nieuwe fysica bestaat. Dit proefschrift maakt een deel uit van een groter geheel: de poging van het CMS samenwerkingsverband tot het ontdekken van fysica buiten het standaardmodel.

Drie verschillende onderzoeken naar heden onontdekte processen in data van 13 TeV proton-protonbotsingen, verzameld met de CMS detector aan de CERN LHC, worden gepresenteerd. De onderzoeken richten zich op diverse fysische processen, maar zijn experimenteel nauw met elkaar verwant. Elk van de drie onderzoeken gebruikt protonbotsingen die resulteren in meerdere geladen leptonen, met een bijzondere focus op botsingsevenementen met drie leptonen.

Ten eerste wordt een zoektocht naar steriele Majorananeutrino's in botsingsevenementen met drie lichte leptonen, elektronen of muonen, getoond. Dit is het eerste onderzoek aan de LHC dat botsingsevenementen met drie leptonen gebruikt om het bestaan van steriele neutrino's na te gaan. Dit maakt het mogelijk om voor de allereerste keer het bestaan van steriele neutrino's met massa's kleiner dan 40 GeV te onderzoeken in een hadronversneller. Massa's zo klein als 1 GeV kunnen worden bestudeerd. Bovendien is dit het eerste resultaat dat in staat is om steriele neutrino's met massa's groter dan 500 GeV te verkennen. Het bereik van de zoektocht loopt tot steriele neutrino massa's van 1.2 TeV. Geen teveel aan evenementen over de veronderstelling dat enkel het standaard model bestaat wordt geobserveerd. De resultaten worden bijgevolg gebruikt om massa-afhankelijke bovenlimieten te plaatsen op de eventuele vermenging van steriele neutrino's met elektron- en muonneutrino's. De limieten variëren van $1.2 \cdot 10^{-5}$ tot 1 voor steriele neutrino massa's in het interval 1 GeV–1.2 TeV.

Hierna wordt een onderzoek naar de simultane productie van een enkel topquark met een Z-boson (tZq) gepresenteerd. Dit is een extreem zeldzaam standaardmodelproces, ongeveer 50 keer zeldzamer dan de productie van een Higgsboson in 13 TeV protonbotsingen. Vanwege zijn zeldzaamheid bleef tZq productie onontdekt voorafgaand aan de resultaten die in dit proefschrift uiteengezet worden. Botsingsevenementen met drie lichte leptonen en minstens één jet die geïdentificeerd is als komende van een b-hadron, worden geanalyseerd. Door gebruik te maken van artificiële intelligentie voor het identificeren van leptonen, en de strategie van de analyse te herontwikkelen, wordt het bereik van het onderzoek enorm vergroot ten opzichte van eerdere resultaten. Voor de eerste keer wordt tZq productie geobserveerd, met een geobserveerde (verwachte) significantie van 8.2 (7.7) standaardafwijkingen boven de veronderstelling dat er enkel een achtergrond is. De werkzame doorsnede van tZq productie wordt gemeten met een relatieve onzekerheid van 15%, meer dan twee keer kleiner dan de onzekerheid in alle eerdere metingen.

Ten slotte wordt een zoektocht naar de directe elektrozwakke productie van chargino's en neutralino's, superpartners van de elektrozwakke ijk- en Higgsbosonen, gepresenteerd. Het aantal mogelijke manieren waarop chargino's en neutralino's kunnen vervallen is enorm. Daarom worden evenementen met drie of meer leptonen, ook deze die tot twee hadronisch vervallende τ -leptonen bevatten, en evenementen met twee lichte leptonen met dezelfde lading geanalyseerd. De botsingsevenementen worden onderverdeeld in een groot aantal orthogonale categorieën aan de hand van de leptontypes en ladingen die ze bevatten. Elke categorie is gevoelig aan bepaalde mogelijke manifestaties van supersymmetrie. Evenementen met drie geladen leptonen, waarvan er minstens twee een paar vormen met tegengestelde ladingen en hetzelfde leptontype, verschaffen het merendeel van gevoeligheid van

de zoektocht aan een aantal belangrijke modellen. Helaas vallen de meeste achtergrondevenementen van het standaardmodel ook onder deze categorie. Daarom worden hier verschillende parametrische neurale netwerken gebruikt om het signaal van de achtergrond te onderscheiden, hetgeen het bereik van de zoektocht enorm uitbreidt. Dit is de eerste keer dat parametrische neurale netwerken worden gebruikt in dit soort onderzoek. De methodes om het signaal van de achtergrond te onderscheiden worden ook in alle andere categorieën van botsingsevenementen herontwikkeld ten opzichte van eerder werk. Dit brengt een veel groter bereik met betrekking tot alle onderzochte modellen van superpartnerproductie met zich mee. Nogmaals wordt er geen teveel aan evenementen gevonden, dus worden limieten gezet op de productie van chargino's en neutralino's. Onder bepaalde voorwaarden worden charginomassa's tot 1450 GeV uitgesloten, tot op heden de globaal hoogst uitgesloten massa voor elektrozwakke superpartners onder welke hypothese dan ook.

Acknowledgements

*“A torch is lit by another
and burns till it’s burned out;
a fire is kindled by another fire.
A man becomes wise
by speaking with other men,
but foolish by keeping to himself.”*

Unknown, *Poetic Edda*

The journey that culminated in this dissertation spanned a number of years over the course of which I interacted with many people. The work and the people along the way shaped me professionally, but also on a deeply personal level. I want to express my heartfelt gratitude to those that have positively impacted my research and my life the most significantly. I start by thanking the former group, and then move on to the latter.

I commence by thanking the jury and the chair of the examination board. You took the time to read this manuscript and thoroughly review my work, thank you for that.

Next, I want to thank Didar Dobur. As the supervisor of both my PhD and my master thesis, you presented me with many unique research opportunities. I would not have been a part of some of the fascinating projects in this dissertation without you. While there was often strife between us early on in my PhD, I think everything worked out well in the end. I appreciate all the support you gave my work over the years. Thank you for being the catalyst for this whole experience!

Then I want to thank Dirk Ryckbosch. You gave me the opportunity to do a summer internship at CERN before I started my master thesis. I do not exaggerate when I say that this has drastically impacted the course of my life. Being at CERN opened my eyes to how interesting experimental physics, data analysis, and even software development are. Fundamentally, the things I became interested in that summer are still what I am working with today.

I also express my heartfelt gratitude to Lesya Shchutska. Working closely with you during both my master thesis and the first analysis of my PhD taught me most of the basics of CMS physics. It is from this basis that I set out to obtain the rest of the results presented in this dissertation. Not only that, but your tireless drive was highly motivational during the early period of my PhD.

Another person worth mentioning is Marco Peruzzi. You were an overall inspirational figure for me as a physicist. You were one of the few people that actively challenged and tried to improve my analysis methods during the later stages of my PhD. You always took the time to explain your ideas and teach me new things. You brought me to new insights and supported my work on many occasions. Aside from that, I think we have had some great times, at conferences, hiking, and at CERN. It really was a pleasure to work with you.

Other people whom I would like to thank for their advice and support of my work are Santiago Folgueras, Filip Moortgat, Pieter Everaerts, and Laurent Thomas. Each of you gave me valuable suggestions on different occasions and helped guide my work. I am thankful to have worked with all of you.

Carlos Erice Cid also deserves my gratitude. I am happy we worked together on the search for electroweak superpartner production presented at the end of this dissertation. I think we learned a

lot from each other and were a very effective duo in completing that search. Thank you for the great team work!

Next, I shall thank my office mates Illia, Tom, and Martina. Illia, I have amazing memories of the old days when we had the office for just the two of us. Aside from the fun we had, you helped me with a lot of practical problems when I was still an inexperienced master student, and I appreciate it. Tom, you were perhaps the only person in our research group as passionate about software development as I was. I enjoyed our collaboration in writing and maintaining a lot of the software used by the UGent CMS group. What I perhaps enjoyed even more was judging bad code together on a daily basis. Martina, we did not really see eye to eye in the beginning of our PhD projects, but at the end of the road I can say I consider you a close friend. Over the years you were the person at work I talked to the most, be it serious discussions, or more often gossip. Unfortunately, there were some occasions where we spent more time chatting in our office than actually doing research.

Another person I have to thank is Marek. Thank you for all the fun discussions and distractions during the pandemic. I never thought I would end up wasting countless hours watching videos of people reviewing old military food rations, but you managed to get me there.

Perhaps I am most grateful of all to my parents. The reasons are legion and obvious. You are ultimately who supported me all the way up to the beginning of my PhD, and even during the last few years when it was needed. Without you none of this would have happened. I sincerely hope that now that this dissertation is finished I will have more time to spend with both of you. I would also like to thank Sophie and Philip.

My grandparents have also had a great impact on my life, and it saddens me that both of my grandfathers passed away during the years that spanned this PhD. I am sure both of them would have been proud to see this dissertation being completed. I will never forget either of them.

Julia, I am incredibly lucky to have met you. Your support and patience have been invaluable during the writing of this dissertation, and I appreciate them immensely. We had some wonderful times over the last year. I do understand that it was not always easy with me spending my time writing this dissertation after coming home from work, and often not going anywhere during the weekends. But at every turn I felt your support and love. Thank you for being there, this is where the fun begins! I would also like to thank Per and Birgitta. You have been incredibly hospitable, and really made me feel welcome in your family. You deserve a special thank you for coming all the way to Belgium to see the defense of this dissertation.

Hendrik, I could not wish for a better brother or friend. I can not begin to thank you for anything specific, there would be too much to list. You have always been one of the most important people in my life, and that will not change. Thank you for all the good times, and may there be more of them for a long time to come!

Lennart, you are the oldest friend with whom I still have regular contact. We only met sporadically during most of our student years, but I am happy our friendship got reinvigorated during the pandemic. I really enjoy the countless evenings we spent playing Guild Wars together, and occasionally going out for drinks after it was possible again. Now that this dissertation is wrapped up I will have more time on my hands for some fun evenings.

My friends from university: Jesse, Simon, Pol, and Sander are also dear to me. My student years preceding this PhD would not have been nearly as fun without you, if I would have even ended up here at all. Who knew that going to the wrong classroom in the first week of university, or overhearing strangers in the auditorium talk about Starcraft could lead to long friendships. I am happy those coincidences occurred, and I hope we get to have more good times together in the future. Jesse, you deserve a special commendation for being an endless source of interesting discussions and juicy gossip. Let us keep it up!

I would not be able to complete this dissertation in good conscience without thanking Jaana. All but one of the years I spent doing active research for this dissertation I spent with you. I will always fondly look back on our time together. You were an incredible support during that period, and I will forever be grateful. Continuing this thread of thought, I also want to thank Virpi and Eetu. You let

me spend many holidays at your home in Raabe, and I always felt very welcomed by both of you. I am truly grateful for having gotten to know you.

I also feel like giving a shout-out to my friends from the CERN School of Computing: Oliver, Kimmo, Andrzej, and Maciej. That school was the greatest trip among all the conferences and schools I attended during my PhD. It is incredible to see that after three years our chat is still alive and well. Let us keep it up and maybe even relive the good old days on another trip!

Notations and conventions

“Most basic ideas of science are essentially simple and can usually be expressed in a language that everyone understands.”

Albert Einstein

Several conventions are used throughout this dissertation and are listed here. Natural units in which the speed of light and the reduced Planck constant are set to unity are used without exception. This greatly simplifies the notation when presenting particle masses and momenta.

Neutrinos are technically leptons, but the text often uses the term lepton while solely referring to charged leptons. This is a pragmatic stance driven by the fact that the dissertation puts a heavy emphasis on experimental signatures with multiple charged leptons, while neutrinos can not be directly detected with the CMS detector. When statements are made that also regard neutrinos, this is clearly indicated. In every other case the reader should interpret the word lepton as referring to charged leptons.

In a similar fashion, particle names such as electron or muon are used as umbrella terms that implicitly cover charge conjugation. This means that a positron also falls under the term “electron”, unless it is clear from the context that it is not so. The same holds for antimuons, τ antileptons, and antiquarks. The charge of a particle is also often called its “sign”, the reason being that only the signs of particle charges are directly measured by the CMS detector.

Throughout the dissertation particles names might refer to a particular physical state or to a reconstructed particle candidate. For instance when using the term “muon” this might mean a muon in the physical sense, or a particle that is reconstructed in a collision event and identified to be a muon with high probability. Which of the two is meant will be clear from the context. Virtual particles are represented by a $*$ superscript, and brackets indicate that a particle might be either real or virtual. For example the notation $W^{(*)}$ covers both real and virtual W bosons.

All figures in the dissertation have references unless they were made by me or a colleague that took part in the publication process of one of the presented analyses. Figures which I did not produce myself are clearly indicated as such in the acknowledgement sections of each chapter. All figures displaying numeric quantities are made in either ROOT [1, 2] or MATPLOTLIB [3]. Illustrations, including Feynman diagrams, are made using the TIKZ [4] package. All histograms included in the dissertation that do not come from work published by others, with the exception of Fig. 7.3 and 7.4 include overflow bins.

Contents

1	Introduction	1
2	Standard model of particle physics and beyond	5
2.1	Standard model particle zoo	6
2.2	Theoretical fundamentals	7
2.2.1	Quantum electrodynamics	7
2.2.2	Nonabelian gauge transformations	9
2.2.3	Quantum chromodynamics	10
2.2.4	Electroweak unification and spontaneous symmetry breaking	11
2.2.5	Fermion masses	14
2.2.6	Mixing matrices	15
2.3	Open problems	16
2.3.1	Gravity	16
2.3.2	Dark matter	16
2.3.3	Dark energy	17
2.3.4	Matter-antimatter asymmetry	17
2.3.5	Neutrino masses	18
2.3.6	Hierarchy problem	19
2.3.7	Discrepancies in high energy physics measurements	19
2.4	This work in context	20
3	Large Hadron Collider	23
3.1	Necessity of colliders	23
3.2	Cross sections and luminosity	24
3.3	Particle colliders	25
3.4	Large Hadron Collider	27
3.5	Simulating collision events	31
4	CMS detector	35
4.1	Coordinate system	37
4.2	Missing transverse momentum and impact parameters	38
4.3	Superconducting magnet	38
4.4	Tracking detector	40
4.4.1	Pixel detector	40
4.4.2	Silicon strip tracker	41
4.5	Electromagnetic calorimeter	42
4.6	Hadron calorimeter	43
4.7	Muon system	45
4.8	Trigger system	46
4.9	Detector Simulation	47

5	Event reconstruction	49
5.1	Particle flow reconstruction	49
5.1.1	Track and vertex reconstruction	50
5.1.2	Calorimeter clustering	55
5.1.3	Information combination	56
5.2	Jet clustering	57
5.3	Jet energy scale correction	58
5.3.1	Pileup correction	58
5.3.2	Response correction	59
5.3.3	Missing transverse momentum	60
5.4	Heavy flavor jet identification	60
5.5	Reconstruction and identification of hadronic τ lepton decays	64
6	Multilepton final state	67
6.1	Light lepton identification	68
6.1.1	Standard lepton identification techniques	69
6.1.2	Machine learning based lepton identification	72
6.1.3	Lepton identification efficiency measurement	73
6.2	Trigger selection	75
6.3	Standard model processes in the multilepton final state	75
6.3.1	WZ production	77
6.3.2	ZZ production	79
6.3.3	Photon conversions	79
6.3.4	Nonprompt leptons	80
6.3.5	Top quark associated production	80
6.3.6	Multiboson production	81
6.4	Nonprompt lepton background prediction	82
6.4.1	Cone correction	83
6.4.2	Loose selection tuning	84
6.4.3	Fake rate measurement	85
6.5	Systematic uncertainties	90
6.5.1	Experimental uncertainties	90
6.5.2	Theoretical uncertainties	92
7	Search for heavy neutral leptons	93
7.1	Introduction	93
7.2	Signal model	95
7.3	Event and object selection	96
7.3.1	Lepton selection and triggering	96
7.3.2	Jet selection	100
7.4	Signal extraction	100
7.5	Background estimation	106
7.6	Systematic uncertainties	108
7.7	Search region results	111
7.8	Interpretation	111
7.9	Summary and outlook	118
7.10	Acknowledgements	119

8	Observation of tZq production	121
8.1	Introduction	121
8.2	Event and object selection	123
8.2.1	Jet selection	123
8.2.2	Lepton identification	124
8.2.3	Event selection	127
8.3	Signal extraction	128
8.4	Background estimation	134
8.5	Systematic uncertainties	137
8.6	Observation and measurement	142
8.7	Summary and outlook	147
8.8	Acknowledgements	150
9	Search for charginos and neutralinos	151
9.1	Introduction	151
9.2	Signal models	155
9.2.1	$\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with slepton-mediated decays	156
9.2.2	$\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with W, H, and Z boson-mediated decays	157
9.2.3	Production of mass-degenerate neutralinos and charginos	157
9.2.4	Simulation and cross sections	159
9.3	Event selection	159
9.4	Signal extraction	164
9.4.1	Three light leptons with at least one OSSF pair	164
9.4.2	Three light leptons without an OSSF pair	183
9.4.3	Two same-sign light leptons	184
9.4.4	Three leptons with one or more τ_h candidates	185
9.4.5	Four or more leptons	188
9.5	Background estimation	188
9.5.1	Nonprompt leptons	189
9.5.2	Electron sign misidentification	191
9.5.3	Prompt leptons	192
9.6	Systematic uncertainties	194
9.7	Results	198
9.8	Interpretations	216
9.9	Summary and outlook	227
9.10	Acknowledgements	229
10	Summary and outlook	231
	Bibliography	235

Acronyms

ALICE A Large Ion Collider Experiment. 27

ATLAS A Toroidal LHC ApparatuS. 27, 29, 94, 95, 113, 114, 118, 121, 123, 145, 150, 152–154, 227, 232, 233

AUC area under ROC curve. 126, 131, 175, 176, 181

AVR adapive vertex reconstruction. 62

BDT gradient boosted forest. 65, 66, 70, 73, 83, 85, 87, 124–129, 131, 132, 134, 135, 137, 138, 140–143, 145–150, 159–161, 164, 168, 181, 228, 229

C charge conjugation. 6, 18

CERN European Organization for Nuclear Research. 1, 14, 27, 28

CKM Cabibbo–Kobayashi–Maskawa. 15, 60

CMS Compact Muon Solenoid. 1, 2, 6, 12, 14, 15, 27, 29–31, 35–38, 40, 42, 45–47, 49–52, 56, 57, 60, 62, 63, 65, 68, 71–73, 75, 76, 80, 94, 95, 113, 114, 118, 119, 121, 123–125, 127, 128, 134, 137, 139, 144, 145, 147, 150, 152–154, 159, 162, 172, 197, 227, 228, 231–233

CP charge conjugation parity. 6, 15, 17, 18

CPT charge conjugation parity time. 8, 19

CSC cathode strip chamber. 46, 53

DT drift tube. 45, 46, 53

EB ECAL barrel. 42, 44

ECAL Electromagnetic Calorimeter. 42–44, 49, 53, 55–57, 70, 90, 98, 114, 123, 124, 160

EE ECAL endcaps. 42

FSR final state radiation. 59, 79, 84, 92, 108, 141, 194

GSF Gaussian sum filter. 52, 53, 56, 162

HB HCAL barrel. 43, 44

HCAL Hadon Calorimeter. 43, 44, 49, 55–57, 59, 60, 70, 90

HE HCAL endcaps. 43, 44

HF forward HCAL. 44

- HLT** High Level Trigger. 46, 47
- HO** outer HCAL. 44
- ISR** initial state radiation. 59, 84, 92, 108, 141, 165, 173, 184, 197
- IVF** inclusive vertex finding. 62
- L1 trigger** Level-1 trigger. 46, 75
- LEP** Large Electron-Positron Collider. 18, 27, 93, 113, 114
- LHC** Large Hadron Collider. 1, 2, 5, 14, 23–25, 27–31, 35, 37, 38, 41–43, 45, 46, 49, 53, 55, 68, 91, 93–96, 118, 121, 123, 130, 149, 151–153, 155, 156, 185, 228, 231–233
- LHCb** LHC beauty. 27, 95
- LO** leading order. 34, 95, 159
- LSP** lightest supersymmetric particle. 151, 153, 155, 157, 163, 165, 168, 172, 182, 216, 217
- LSTM** long short-term memory. 63
- MC** Monte Carlo. 34, 58, 143, 144
- MSSM** minimal supersymmetric extension of the standard model. 155
- NLO** next-to-leading order. 34, 92, 95, 144, 159, 173, 174, 216
- NLSP** next-to-lightest supersymmetric particle. 155, 216, 217
- OSSF** opposite-sign and same-flavor. 77, 100–108, 128, 129, 135, 155, 163–165, 172, 182–185, 188, 189, 192, 194, 203, 205, 210, 211, 227, 228
- PDF** parton distribution function. 26, 92, 95, 108, 110, 140, 141, 197
- PF** particle-flow. 49, 50, 55–60, 68–71, 123, 159
- PMNS** Pontecorvo–Maki–Nakagawa–Sakata. 15
- PS** Proton Synchrotron. 27, 29
- QCD** quantum chromodynamics. 10, 11, 17, 31–34, 46, 61, 65, 66, 84, 85, 88, 92, 95, 108, 110, 140, 141, 144, 159, 161, 173, 174, 197, 198, 216
- QED** quantum electrodynamics. 7, 9–11, 14
- ROC** receiver operating characteristic. 126, 175
- RPC** resistive plate chamber. 46, 53
- SM** standard model of particle physics. 1, 2, 5–8, 11, 14–21, 23, 24, 27, 29, 31, 34, 38, 65, 67, 68, 75, 76, 93–95, 100–104, 106, 110, 111, 116, 118, 121, 134, 144, 145, 147, 149–153, 155, 156, 165–167, 169–173, 175, 178, 182–184, 187, 188, 190, 192, 195, 196, 198, 215, 216, 227, 228, 231–233
- SPS** Super Proton Synchrotron. 14, 27, 29

Acronyms

TEC Tracker EndCaps. 41

TIB Tracker Inner Barrel. 41

TID Tracker Inner Disks. 41

TOB Tracker Outer Barrel. 41

Chapter 1

Introduction

*“The words unfreeze in my mouth
and the phrases are tumbling
upon my tongue they scramble
along my teeth they scatter.”*

Elias Lönnrot, *Kalevala*

The standard model of particle physics (SM) is the culmination of a centuries old drive to understand the fundamental laws of nature. The theory describes the smallest constituents of matter and their interactions. It is the most fundamental theory of physics that has been experimentally verified. While the SM is often lauded as one of the greatest achievements of modern science, it is unlikely to be the final word in fundamental physics. Observations in astronomy, cosmology, and high energy physics itself, as well as theoretical arguments, indicate that there must be a deeper and more general theory than the SM. The quest to solve this puzzle and find physics beyond the SM has involved several generations of physicists over the last decades. This dissertation is a wave in an ocean of results that aim to shed light on the open questions in the SM. Chapter 2 briefly introduces the SM and brushes over the loose ends left by the theory.

Testing the SM at high precision and having the possibility to find what is beyond requires the measurement of a vast amount of particle interactions at the highest possible energies. At the time of writing the European Organization for Nuclear Research (CERN) Large Hadron Collider (LHC) is the foremost source of high energy particle interactions, colliding protons at the energy and intensity frontier. In recent years, the LHC delivered an unprecedented volume of proton-proton (pp) collision data at a center of mass energy of 13 TeV. This is the highest collision energy ever achieved in a laboratory, almost an order of magnitude higher than the energies realized with any other machine. The LHC is the ideal tool for exploring the SM and what might be beyond. The LHC itself, and the need for such a machine are described in Chapter 3.

Just having high energy particle interactions is not enough, they must be measured with great accuracy in the harsh environment of the intense LHC collisions. This is the job of the Compact Muon Solenoid (CMS) detector, one of two general purpose physics detectors that surrounds a point where the LHC beams collide. The work in this dissertation is a part of the larger CMS Collaboration effort to search for new physics and to thoroughly test the SM. All of the collision data used in this work was collected with the CMS detector. Chapter 4 describes the CMS detector and its characteristic features.

Translating the complex signals generated by a modern detector like CMS into a coherent picture of particle collisions requires advanced reconstruction and particle identification algorithms. The essential parts of the reconstruction and identification procedures relevant to this dissertation are laid out in Chapter 5. Starting at this chapter, the text purposefully dives into a deeper level of technical detail compared to the earlier chapters. This is motivated by the fact that the dissertation is focused on the analysis of the CMS data set, rather than on the design of the detector itself. After Chapter 5, the

more general overview of CMS physics ends and the text becomes focused on the particular research topics of the dissertation.

This dissertation is the end product of several years of my research as a member of the CMS collaboration. During this time I had a major involvement in four publications. In three out of these I was the main analyzer, the main editor of the paper, and the main author of the internal CMS documentation on the results. I made contributions to several more publications, mostly in the form of software frameworks which are still heavily used at the time of writing. Analyses where my contributions are minor or restricted to software are omitted from this dissertation. The presented analyses focus on collision events with multiple charged leptons in the final state. Leptons can be reconstructed and identified with outstanding precision because they do not hadronize. The starting point of the work towards this dissertation was a search for the direct electroweak production of supersymmetric particles in the 2016 pp collision data set collected with CMS [5]:

- CMS Collaboration. “Search for electroweak production of charginos and neutralinos in multilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV”. JHEP 03 (2018), p. 166. doi:10.1007/JHEP03(2018)166. arXiv:1709.05406 [hep-ex].

The contributions I made in this search were largely a continuation of the work I did during my master thesis. They revolved around providing a precise data-driven prediction of the main irreducible background to the search, SM WZ production. The most important work, such as the design of the general analysis strategy and producing the final results, was performed by others. Hereafter, I worked on three more searches in which I was the main analyzer. In the last two I was also the contact person. First came a search for the production of heavy sterile neutrinos in events with three charged leptons [6]:

- CMS Collaboration. “Search for heavy neutral leptons in events with three charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV”. Phys. Rev. Lett. 120 (2018), p. 221801. doi:10.1103/PhysRevLett.120.221801. arXiv:1802.02965 [hep-ex].

This was the first search for sterile neutrinos in the three lepton final state at the LHC, and the first search at a hadron collider sensitive to sterile neutrinos in the mass range below 40 GeV. Sterile neutrino masses beyond 500 GeV were experimentally probed for the first time, with the sensitive range of the analysis going all the way up to 1.2 TeV. The search was highlighted in the CERN Courier [7], and I had the opportunity to present the work at two conferences, one of them the major Rencontres de Moriond session on electroweak interactions and unified theories [8].

After completing this project I worked on a search, and eventual measurement, of single top quark production in association with a Z boson (tZq):

- CMS Collaboration. “Observation of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV”. Phys. Rev. Lett. 122 (2019), p. 132003. doi:10.1103/PhysRevLett.122.132003. arXiv:1812.05900 [hep-ex].

The tZq process is about 50 times more rare than H boson production in 13 TeV pp collisions and had never been observed at the time of the search. The analysis presented in this dissertation marked the first ever observation of the process. In this analysis I was the only analyzer. I designed the analysis strategy from beginning to end, though I had the benefit of hindsight from earlier CMS analyses and some advice from other researchers. The search was the headline of the CMS Collaboration news page for several months [9], and I presented this analysis and several others in a plenary presentation at the Rencontres de Moriond session on quantum chromodynamics and high energy interactions [10]. In addition I presented the results as a poster at the European Physical Society conference on high energy physics [11].

The final analysis performed for this dissertation is an updated version of the first one: A new search for the electroweak production of supersymmetric particles, this time using the entire Run 2 pp collision data set collected by CMS from 2016 to 2018 [12]:

- CMS Collaboration. “Search for electroweak production of charginos and neutralinos in proton-proton collisions at $\sqrt{s} = 13$ TeV”. JHEP 04 (2022), p. 147. doi:10.1007/JHEP04(2022)147. arXiv: 2106.14246 [hep-ex].

The analysis was the first to use parametric neural networks to search for the production of supersymmetric particles. This increased the sensitivity of the search compared to the earlier iteration by a greater extent than the four times larger data set did. Like the sterile neutrino search, this search was featured in the CERN Courier [13]. Here I was one of two main analyzers and designed the novel neural network based search strategy. This analysis fully supersedes the results in the first search mentioned in this chapter, and my contribution was far greater. For this reason, the older version of the search is not described in the dissertation.

Although the target physics of the analyses described in this text is markedly different, all of them use collision events containing multiple leptons in the final state. They therefore share a large degree of similarity. Common techniques and considerations that are applicable to the different analyses are described thoroughly in Chapter 6. The following three chapters give detailed descriptions of the specific analyses, while referring back to previous chapters where necessary. The analyses are presented in chronological order, as listed above. Chapter 7 describes the search for sterile neutrinos, Chapter 8 the search tZq , and finally Chapter 9 presents the search for direct electroweak production of supersymmetry. The technical sophistication increases in each chapter, reflecting my increased experience as time went on. The search in Chapter 7 uses fairly classical analysis methods, while the search in Chapter 9 is vastly more complex and uses techniques that were never used before in such a search. Chapter 8 falls somewhere in the middle in terms of complexity.

The descriptions of each of the analyses contains a higher level of detail than the corresponding journal publications. This is particularly the case for the sterile neutrino and tZq searches which were published in Physical Review Letters, necessitating short and dense analysis descriptions. This text aims to make the analyses reproducible to the extent possible, and the hope is that much of the description can be used as a template by other analyzers for doing similar searches in the future. To achieve these goals the content is often intentionally dense.

Each of the analysis chapters includes a short summary at the end. General conclusions and an outlook are given in Chapter 10.

Chapter 2

Standard model of particle physics and beyond

“The Sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work.”

John Von Neumann

Our current knowledge of fundamental particles and their interactions is encapsulated in the SM, the fundamental theory of particle physics. The SM is enormously successful: Its predictions match the data in experiments carried out by generations of physicists. The current incarnation of the theory has been around since the 1970s, although it took until 2012 for the last piece in its particle puzzle, the Higgs boson, to be discovered at the LHC [14, 15].

This chapter gives a superficial panorama of the SM. An overview of the particle content of the theory is given in Section 2.1. The theoretical foundations of the SM as a renormalizable quantum field theory are introduced in Section 2.2.

As it stands today, the SM unfortunately leaves a number fundamental questions unanswered. Some of these are theoretical in nature, while others are experimental observations in both astronomy and high energy particle physics that seem incompatible with the SM. These open questions are briefly presented in Section 2.3 and are the motivation behind the analyses done in the context of this dissertation and the dissertation itself. The work presented in this dissertation covers subjects with highly similar experimental signatures, but very different theoretical backgrounds. The theoretical discussion relating to each of the publications covered in the dissertation is therefore kept relatively short in this chapter. More details about each topic are included in the chapters that describe the corresponding analyses (chapters 7 to 9).

Section 2.4 summarizes where the work presented in this dissertation fits into the larger context of addressing the incompleteness of the SM.

2.1 Standard model particle zoo

“If I would remember the names of all these particles, I’d be a botanist.”

Enrico Fermi

All atoms are made up out of neutrons, protons (p), and electrons (e). In turn, neutrons and protons are composite particles consisting of three quarks [16, 17] encased in a cloud of gluons (g) and quark-antiquark pairs. These particles interact with each other via four forces: The strong force that holds nuclei together, the weak force that is responsible for radioactive decays, and the well-known forces electromagnetism and gravity. The weak and electromagnetic interactions are each parts of the unified electroweak interaction.

The quarks that make up protons and neutrons, and thus all everyday matter, are the lightest two. These are called the up (u) and down (d) quarks. Two heavier sets of quarks with otherwise similar properties to the u and d quarks exist: The strange (s) and charm (c) quarks, and the bottom (b) and top (t) quarks. Each of these sets of two quarks is called a generation, and the individual quark types are called flavors. The t quark is the heaviest particle in the SM, with a mass of approximately 172.8 GeV, similar to the atomic masses of heavy elements. Quarks are susceptible to all four of the fundamental interactions.

To the best of our current understanding electrons are fundamental particles. They belong to a class of particles called leptons. Unlike quarks, leptons do not feel the strong interaction. Heavier copies of the electron exist. They are in order of mass, the muon (μ) and the tau lepton (τ). These charged leptons are closely related to the three flavors of neutrinos, the e neutrino (ν_e), the μ neutrino (ν_μ) and the τ neutrino (ν_τ). In similar fashion to what is done for quarks, the three leptons types and their corresponding neutrinos are often referred to as generations. The neutrinos are electrically neutral, they only feel the weak interaction and gravity. This makes them incredibly hard to detect, as they mostly pass through matter without interacting. They are the lightest fermions in the SM, with masses so small they have not been measured. Only upper limits have been set on their masses. The phenomenon of neutrino oscillations in which one neutrino flavor turns into another while in flight [18, 19], indicates a nonzero mixing matrix between the neutrino flavor and mass eigenstates. In other words, neutrinos are massive particles. In what follows we take on the convention that the word lepton only refers to charged leptons. This is motivated by the fact that the experimental signatures focused on throughout this work always contain multiple charged leptons, while the CMS detector is unable to detect neutrinos.

Leptons, neutrinos, and quarks all have spin values of 1/2, meaning they are fermions and obey Fermi-Dirac statistics. The SM fermions have antiparticles sharing almost all of their properties, but with inverted charges. Particles and antiparticles are not completely equal however: Both the charge conjugation (C) [20] and charge conjugation parity (CP) [21, 22] symmetries are broken in the SM. Some more has to be said on this topic in Section 2.3, in the context of open problems in the SM.

Massless photons are the carriers of the electromagnetic interaction and are themselves electrically neutral. The electrically charged W^+ and W^- bosons (collectively called W bosons), and the neutral Z boson transmit the weak interaction. These gauge bosons are among the most massive particles, making the weak interaction very short ranged. Another defining property of the weak interaction is that interactions via W bosons violate parity (P) symmetry and change the flavor of the particles involved. Eight massless gluons carry the strong interaction. The gluons, W, and Z bosons bear the charges of their respective forces, meaning they are self-interacting. This is not the case for photons. All of the force carries are spin 1 bosons, hence Bose-Einstein statistics is applicable to them. More information about the gauge bosons and their interactions is given in Section 2.2.

The Higgs (H) boson [23–25] is a particle like no other. It is a heavy spin 0 scalar particle [26, 27], necessary for the preservation of gauge invariance in the SM given that the weak gauge bosons are massive. The fundamental fermions also acquire their masses through their couplings to the H boson,

although some possible caveats are discussed in Section 2.3.5. A bit more information on the nature of the H boson is given in Section 2.2.4.

An overview of the fundamental particles, the latest estimates of their masses, and their charges and spins is given in Fig. 2.1.

2.2 Theoretical fundamentals

The SM is mathematically formulated in the framework of quantum field theory [30, 31]. The basic objects in the framework are complex fields whose value depends on their position in space-time. Particles are excitations of the fields, and the central properties of each field are its behaviors under different transformations.

The Lorentz group represents the set of transformations that leave the space-time intervals defined with the Minkowski metric tensor invariant. Angular momenta determine the transformation behavior of fields under the Lorentz group. Fields can be categorized into scalars (spin 0), left-, and right-handed spinors (spin 1/2), vectors (spin 1), and higher spin states. Each type of particle is an excitation of a particular field, and the spin statistics theorem [32] links their spin to their nature as a boson or a fermion. Particles with half-integer spin values are fermions and obey Fermi-Dirac statistics, while integer spin particles are bosons that follow Bose-Einstein statistics.

The equations of motion for fields can be expressed in terms of a Lagrangian density, simply called Lagrangian in what follows. The link between the Lagrangian and the equations of motion is the principle of least action. The principle dictates that the action integral is stationary over the possible paths taken by the field in space-time. The action integral is written as:

$$S = \int \mathcal{L} d^4x. \quad (2.1)$$

For free fermions a Lorentz invariant Lagrangian can be written as:

$$\mathcal{L}_{\text{free fermions}} = i\bar{\psi}\not{\partial}\psi - m\bar{\psi}\psi, \quad (2.2)$$

in which ψ is a spinor under the Lorentz group, $\not{\partial} = \gamma^\mu \partial_\mu$ with γ^μ the four gamma matrices, and $\bar{\psi} = \psi^\dagger \gamma^0$. The Dirac equation that describes the relativistic motion of noninteracting fermions results from finding a stationary state in the action integral of Equation 2.1, after inserting the Lagrangian 2.2.

The Lagrangian 2.2 does not include any interaction terms between fermions. To introduce these, one forces the Lagrangian to be invariant under one or more local gauge transformations. Local in this context means that the phase shift of the gauge transformation depends on the position x in space-time. We start by deriving the Lagrangian of quantum electrodynamics (QED) by introducing gauge symmetry under the simple Abelian U(1) group. We then generalize this principle to nonabelian groups in Section 2.2.2, and use the result to derive the Lagrangians of the strong and unified electroweak interactions in sections 2.2.3 and 2.2.4 respectively.

2.2.1 Quantum electrodynamics

The theory of QED that describes electromagnetic interactions between particles can be formulated by constructing a Lagrangian that is invariant under local U(1) gauge transformations [33–36]. The fermion spinors transform under the U(1) group as:

$$\psi \rightarrow \exp[iq\chi(x)]\psi \quad (2.3)$$

$$\bar{\psi} \rightarrow \exp[-iq\chi(x)]\bar{\psi}, \quad (2.4)$$

where the phase factor χ is a scalar that depends on the space-time coordinate x , and we have explicitly factored out q which is a multiple of the fundamental electric charge e . Because of the x -dependence of χ , a transformation of the Lagrangian 2.2 under the U(1) group results in:

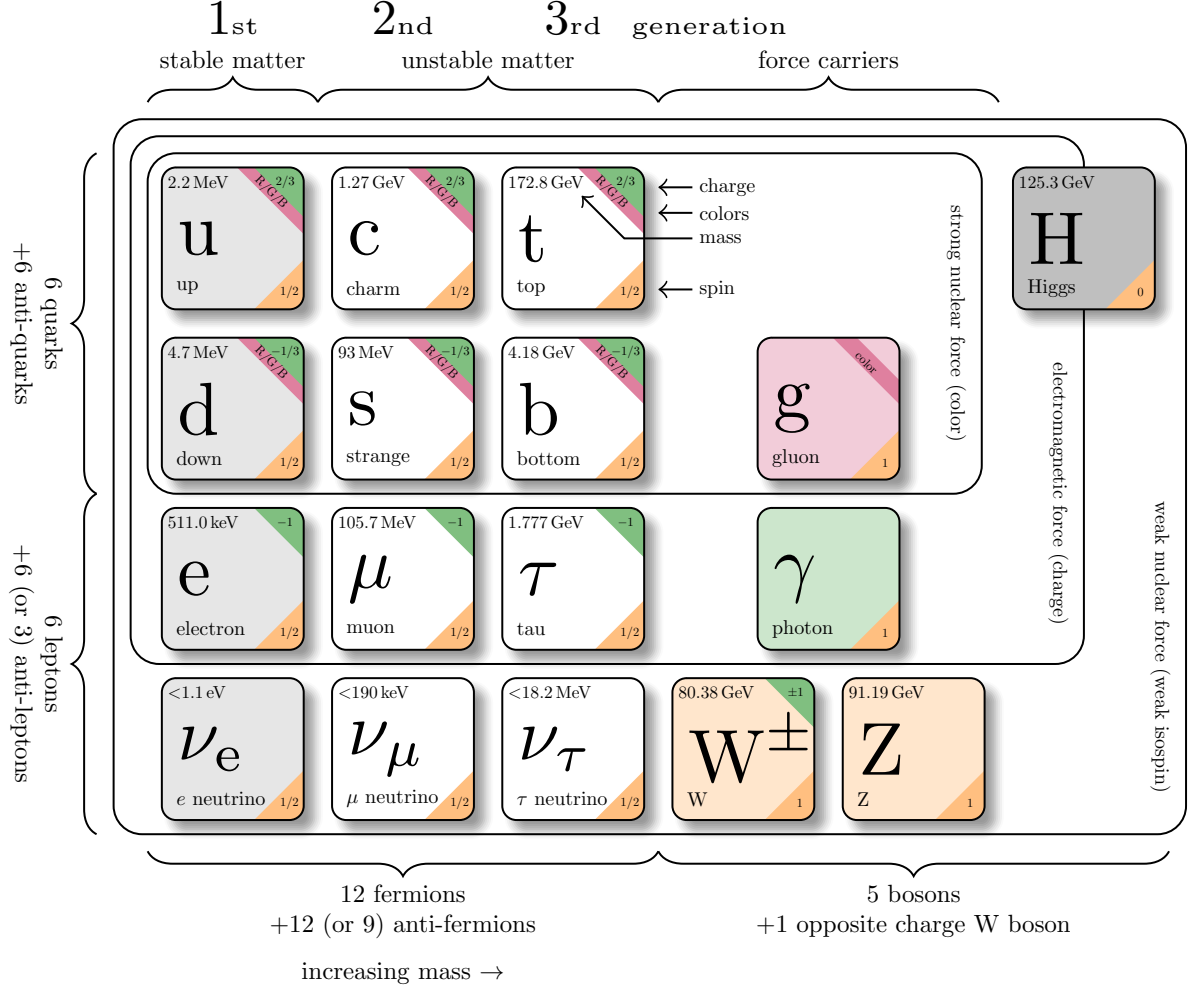


Figure 2.1: Overview of the particles in the SM. The figure is based on a template provided in [28]. Each square box corresponds to a particle and lists its name in the bottom and the corresponding symbol in the center. The symbols defined here are used throughout this dissertation to refer to each of the particles. All of the listed masses are the latest values from Particle Data Group [29] under the assumption of charge conjugation parity time (CPT) symmetry. This symmetry implies that particles and antiparticles have identical masses. For neutrinos no mass measurements are available so upper limits are given. Neutrinos might be either Dirac or Majorana fermions (Section 2.3.5). If the latter is true there are no antineutrinos. For the massless gluons and the photon no mass is indicated. Particle spins are given in the lower right of each box, marked in orange. The top right corner of each box contains the electric (color) charge of each particle in green (purple) if it is nonzero. Fermions are listed on the left, and bosons on the right. Each column of fermions corresponds to a particular generation, marked on top of the column. The figure is subdivided into three areas, indicating the interactions felt or carried by the particles.

$$\mathcal{L}_{\text{free fermions}} \rightarrow \mathcal{L}_{\text{free fermions}} - q\bar{\psi}\psi\phi\chi(x). \quad (2.5)$$

So the Lagrangian for noninteracting fermions is not invariant under local $U(1)$ gauge transformations. In classical electromagnetism the field equations can be written in terms of the electromagnetic four-potential A^μ , and the theory is invariant under the transformation:

$$A_\mu \rightarrow A_\mu - \partial_\mu\chi(x). \quad (2.6)$$

We can introduce a vector field A_μ that satisfies the same property. If we then introduce a term $q\bar{\psi}\not{A}\psi$ into Lagrangian 2.2 we obtain:

$$\mathcal{L}_{\text{symmetric}} = i\bar{\psi}\not{D}\psi - m\bar{\psi}\psi - q\bar{\psi}\not{A}\psi = i\bar{\psi}\not{D}\psi - m\bar{\psi}\psi. \quad (2.7)$$

Here we have introduced the gauge-covariant derivative $D_\mu = \partial_\mu + iqA_\mu$. The new term added to the Lagrangian to preserve gauge symmetry describes interactions between the fermion fields and the vector field A_μ . The latter can be associated to photons, the force carriers of electromagnetism, and the term $q\bar{\psi}\not{A}\psi$ can thus account for interactions between charged particles. To complete the Lagrangian the electromagnetic field tensor is introduced, which accounts for the equations of motion of free photons. In classical electromagnetism this tensor can be formulated in terms of the four-potential as:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (2.8)$$

Using the concepts introduced above we can write down the Lagrangian of QED as:

$$\mathcal{L}_{\text{QED}} = i\bar{\psi}\not{D}\psi - m\bar{\psi}\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \quad (2.9)$$

where the factor $1/4$ ensures Maxwell's equations can be retrieved from this Lagrangian. Lagrangian 2.9 forms the basis of QED, the theory that describes all electromagnetic interactions between particles. These interactions are the origin of chemistry. Because of the lack of interactions between photons themselves and the small electromagnetic coupling constant, extremely precise predictions can be made with QED.

Practical QED calculations are done with perturbation theory. Each term in a perturbative expansion of a particle interaction can be pictorially represented by a Feynman diagram. Each diagram represents the amplitude of a process, and a set of rules link parts of each Feynman diagram to certain terms in the amplitude. The coupling constant of QED, directly linked to the fundamental charge e , is smaller than one. This makes perturbative calculations convergent, with higher order terms providing ever smaller corrections. Some more will be said about this in the context of the strong interaction in Section 2.2.3, where the practical circumstances are unfortunately less favorable.

While Feynman diagrams can be used as mathematical tools for calculations in perturbation theory, they are used throughout this dissertation as aides to represent the particles involved in a certain process.

2.2.2 Nonabelian gauge transformations

To introduce the strong and weak interactions into a Lagrangian, the procedure introduced in Section 2.2.1 to make a Lagrangian gauge invariant must be extended to the nonabelian $SU(N)$ groups with N larger than one [37]. The group $SU(N)$ consists of the transformations that can be represented by unitary $N \times N$ matrices. To be able to apply an $SU(N)$ transformation to the Lagrangian we arrange the fermion spinors into multiplets as follows:

$$\Psi_{\text{SU}(N)} = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_N \end{bmatrix}, \quad (2.10)$$

and replace the spinors in Lagrangian 2.2 by such multiplets. The resulting Lagrangian is not invariant under local $\text{SU}(N)$ gauge transformations. Invariance can be restored by replacing the derivative by the following gauge covariant derivative:

$$D_\mu = \partial_\mu + i\frac{g}{2}\tau^a A_\mu^a, \quad (2.11)$$

where each τ^a is a generator of the $\text{SU}(N)$ group, an implicit summation over all $N^2 - 1$ generators is assumed, g is the coupling constant of the interaction being described, and the factor $1/2$ is introduced by convention. Instead of the one gauge field in QED, one gauge field A_μ^a is introduced for each generator.

Because $\text{SU}(N)$ gauge transformations do not commute, the transformation properties of the gauge fields necessary for a gauge invariant fermionic part of the Lagrangian become more complex. Each field must transform as:

$$A_\mu^a \rightarrow A_\mu^a - \partial_\mu \chi^a(x) - gf^{abc}\tau^a \chi^b(x)A_\mu^c, \quad (2.12)$$

with f^{abc} the structure constants of the $\text{SU}(N)$ group. For each of the $N^2 - 1$ gauge fields we must introduce a bosonic term like the one describing free photons in Equation 2.9. But given the transformation in Equation 2.12, the equivalent of the field tensor $F_{\mu\nu}$ for the new gauge fields must become:

$$G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc}A_\mu^b A_\nu^c, \quad (2.13)$$

to retain gauge invariance. The final term in Equation 2.13 includes different vector fields and represents interactions between the gauge bosons that were introduced to enforce the gauge symmetry. This is a notable difference with the Lagrangian obtained for QED through the Abelian $\text{U}(1)$ group, which contains no such term. Photons do not couple to each other, but gauge bosons of an $\text{SU}(N)$ gauge theory do.

2.2.3 Quantum chromodynamics

The theory of the strong interaction, quantum chromodynamics (QCD), can be formulated in terms of a gauge symmetry of the Lagrangian of quarks under the $\text{SU}(3)$ gauge group. Quark spinors are organized into triplets that transform under the $\text{SU}(3)$ group. The different entries in the triplet are referred to as colors of a certain quark flavor. The gauge-covariant derivative of QCD is directly given by Equation 2.11 in which eight gauge fields appear. The gauge fields that are introduced are the gluons that carry the strong interaction [38, 39].

Like photons, gluons are massless particles. But unlike photons, gluons interact among themselves following Equation 2.13. Gluon self interactions make QCD vastly more complex than QED. Not only does the presence of self interactions lead to far more terms at each order of a perturbative expansion, it results in a large effective coupling constant α_s at low energy scales. This makes perturbative calculations in low energy QCD impossible. A theoretical treatment of this effect is beyond the scope of this dissertation, but some intuition is provided below. The coupling constant α of a force is defined in terms of the charge g appearing in Equation 2.11 as:

$$\alpha = \frac{g^2}{4\pi}. \quad (2.14)$$

The coupling constant of each interaction depends on the square momentum transfer scale Q^2 of the interaction. In the simplest case, QED, the evolution of the coupling constant can be understood by the fact that each vertex in a Feynman diagram has an infinite series of higher order correction terms. In QED higher order terms increase the scattering amplitude, and the contribution of the higher order terms becomes larger at larger Q^2 . The QED coupling constant can be measured indirectly by experimentally determining the cross section of a process. The measured cross section includes all terms in the perturbative expansion of the process. One can define an effective coupling constant at the Q^2 value of the measurement, which absorbs the effects of all higher order diagrams. Interactions occurring at larger momentum transfer scales have higher effective coupling constant values because the contribution of higher order terms grows with Q^2 .

A more intuitive way to understand this is by imagining a measurement of the electron charge by scattering charged particles off an electron. Electron-positron pairs spontaneously appear and annihilate in a quantum field theory vacuum. An electron polarizes these pairs, and the positive charges of the vacuum fluctuations are on average closer to the electron than the negative charges. This effect screens the electron charge. The higher the energy of the scattering probe, the closer it can get to the electron and the smaller the screening effect is. With a higher energy probe particle a larger electron charge, or equivalently QED coupling constant, is measured.

Gluon self interactions make the evolution of the QCD coupling constant as a function of Q^2 fundamentally different to that of the QED coupling constant. It can be shown that vector boson loop diagrams give contributions to scattering amplitudes that are of opposite sign to those given by fermion loops. In QED only fermion loops are present, while QCD has both fermion and boson loops because of gluon self interactions. This gluon antiscreening effect is larger than the screening effect of quarks, and the QCD coupling constant decreases as a function of Q^2 . At sufficiently low Q^2 values α_S becomes larger than one, and perturbative calculations stop converging. Measurements of α_S as a function of the momentum transfer scale are shown in Fig. 2.2. The large QCD coupling constant makes low energy QCD calculations enormously complex. Phenomenological methods are needed in this regime since calculations from first principles are too complex [40].

Free quarks have never been observed. For this reason it is hypothesized that only color neutral states can appear as free particles. Quarks must either be bound in mesons, which consist of quark-antiquark pairs or baryons made up of three (anti)quarks of different colors. Mesons and baryons are collectively called hadrons. The fact that no free colored states exist causes the strong interaction to be short-ranged, as opposed to electromagnetism which has an infinite range. No analytic proof for this exists as of now. The energy needed to separate quarks is found to approximately increase linearly with the distance between them. At a certain distance between two quarks, the gluon field between them contains sufficient energy to produce new quark-antiquark pairs, which bind to the original quarks. This process is known as hadronization. In collider experiments this manifests itself by the appearance of jets of hadrons in the detector. Quarks produced in a collision can travel away from each other until new quark-antiquark pairs are produced from the energy between them. This process cascades to create more hadrons until there is no longer sufficient energy to produce additional quark-antiquark pairs. More details about this are given in Section 3.5.

2.2.4 Electroweak unification and spontaneous symmetry breaking

In addition to the strong and electromagnetic forces, the SM also describes the weak interaction. Experimentally, the weak interactions is found to violate parity [20], and to contain both flavor-changing charged current interactions and flavor-conserving neutral current interactions [42]. The short range of the weak interaction hints at the responsible gauge bosons being massive. These properties make the weak interaction conceptually rather different from the strong and electromagnetic interactions, both of which are mediated by massless gauge bosons which can not change the flavor of a particle.

The interaction terms in the QED and QCD Lagrangians transform as vectors under Lorentz transformations. Only a few possible forms of interaction terms are allowed by Lorentz invariance,

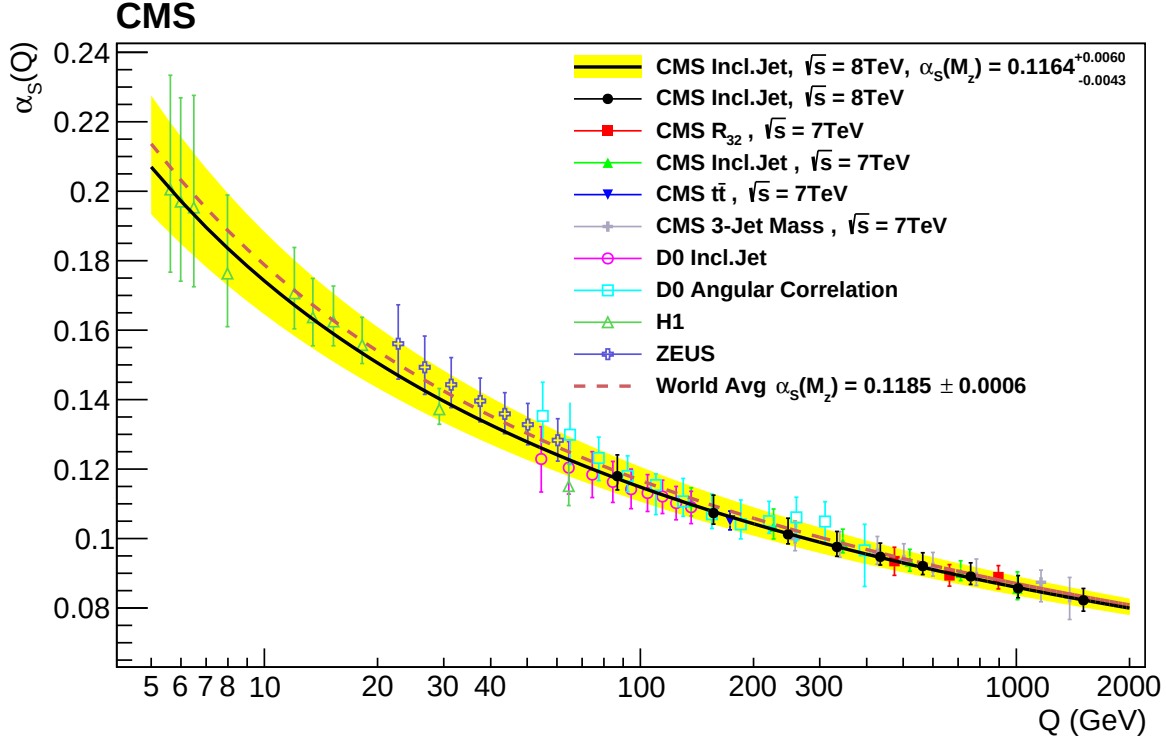


Figure 2.2: The strong coupling constant α_s as a function of the momentum transfer scale Q . The black line and yellow band represent α_s and its associated uncertainty as measured by CMS [41]. A number of measurements from different experiments are also included in the figure.

and it turns out that the parity violation of the charged-current weak interaction is well modeled by an interaction term which has a vector component and a negative axial vector component [43, 44]. Such an interaction term can be obtained by splitting fermion spinors into their left and right chiral components as:

$$\psi = \frac{1 - \gamma_5}{2} \psi + \frac{1 + \gamma_5}{2} \psi = \psi_L + \psi_R, \quad (2.15)$$

and having only the left-handed components take part in the interaction. We used the chirality operator γ_5 , which is defined as $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$.

The charged-current weak interaction changes left-handed leptons into left-handed neutrinos and left-handed up-type quarks into left-handed down-type quarks and vice versa. This can be explained by arranging left-handed particles into doublets that transform under the SU(2) gauge group [45, 46]:

$$\Psi_{L \text{ lepton}} = \begin{bmatrix} \nu_{\ell L} \\ \ell_L \end{bmatrix} \quad (2.16)$$

$$\Psi_{L \text{ quark}} = \begin{bmatrix} u_L \\ d_L \end{bmatrix} \quad (2.17)$$

It turns out that for a complete picture of the weak interaction only introducing an SU(2) gauge symmetry is not enough, and a unification with the electromagnetic interaction is required. We therefore impose both an SU(2) and U(1) symmetry on the Lagrangian. Right-handed fermions are taken to transform trivially under the SU(2) gauge transformations, while all fermions are susceptible to the U(1) gauge transformations. Following Equation 2.11 we introduce a SU(2) $\otimes$ U(1) gauge-covariant derivative:

$$D_\mu = \partial_\mu + ig \frac{\sigma^a}{2} W_\mu^a + ig' \frac{Y}{2} B_\mu, \quad (2.18)$$

in which σ^a are the three Pauli matrices that are the generators of the SU(2) group, and a quantity Y called the weak hypercharge is introduced. The covariant derivative brings with it four gauge fields: Three W_μ^a fields, one for each generator of SU(2), and one field B_μ originating from the U(1) gauge symmetry. Linear combinations of two of the W_μ^a gauge fields represent charged-current interactions that change the flavor of the particles involved. These correspond to the physical W^+ and W^- bosons. The third W_μ^a field and the B_μ field conserve the flavor of the interacting particles and correspond to the Z boson and the photon in a way that becomes clear below.

The electroweak gauge bosons are massive, but bare mass terms of the form $m^2 A^\mu A_\mu$ for a gauge field A^μ are not gauge invariant considering Equation 2.6. A mechanism for making a U(1) gauge field massive while still retaining gauge invariance was introduced in [23–25]. It requires the introduction of a complex scalar field with a nonzero vacuum expectation value. This facilitates the introduction of massive gauge fields without loss of gauge symmetry, at the cost of inserting a scalar field into the theory. By extending this procedure to an SU(2) doublet of complex scalars, a consistent theory of electroweak interactions can be formulated [47]. The scalar doublet and the corresponding terms in the Lagrangian are written as:

$$\Phi = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{bmatrix} \quad (2.19)$$

$$\mathcal{L}_{\text{scalar}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2, \quad (2.20)$$

where $\lambda > 0$ and $\mu^2 < 0$. The former condition ensures that the potential term in the Lagrangian is bounded from below, while the latter results in a nonzero minimum. The minima of the potential are the points on the hypersurface that satisfies:

$$\Phi^\dagger \Phi = -\frac{\mu^2}{\lambda} \equiv \frac{v^2}{2}, \quad (2.21)$$

in which we introduced the vacuum expectation value v . Without loss of generality a direction for the vacuum state of the field can be chosen, causing the spontaneous breaking of the SU(2) gauge symmetry. The scalar doublet Φ can be rewritten in terms of excitations of the vacuum state. There are four scalar fields, but it can be shown that only that scalar field corresponding to the direction of the nonzero vacuum state is physical. The other three fields can be gauged away. We call the physical field h , and choose the nonzero vacuum state along the real component of the lower entry in the doublet Φ . The h field corresponds to a massive neutral scalar boson, the H boson. Using the representation of the Pauli matrices used in [31], the first term in Lagrangian 2.20 becomes:

$$\begin{aligned} (D_\mu \Phi)^\dagger (D^\mu \Phi) &= \frac{1}{2} (\partial_\mu h)^\dagger (\partial^\mu h) \\ &+ \frac{g^2}{8} (W_\mu^1 + iW_\mu^2) (W^{1\mu} - iW^{2\mu}) (v + h)^2 \\ &+ \frac{1}{8} (gW_\mu^3 - g'B_\mu) (gW^{3\mu} - g'B^\mu) (v + h)^2. \end{aligned} \quad (2.22)$$

This equation contains several mass terms in which the gauge fields couple to themselves. Diagonalization of the mass terms reveals two massive charged gauge bosons: W^+ and W^- . These are linear combinations of two of the SU(2) gauge fields. Two neutral mass eigenstates are formed from linear combinations of the third SU(2) gauge field and the U(1) gauge field. One of them is massless and corresponds to the photon, while the other one is massive and represents the Z boson [48, 49].

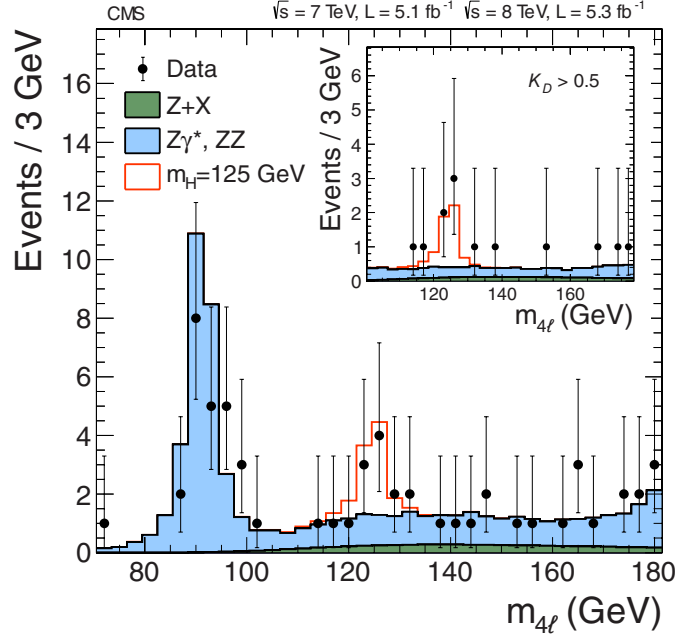


Figure 2.3: Observed and expected event yields in the CMS H boson search as a function of the four lepton invariant mass in events with four leptons. An excess of events over the background-only hypothesis, compatible with a SM H boson with a mass of 125 GeV, can be seen [15].

In summary, a gauge symmetry under $SU(2) \otimes U(1)$ transformations and an $SU(2)$ doublet of scalar fields with a nonzero vacuum expectation value are introduced in the Lagrangian. This results in a unified theory of the weak and electromagnetic interactions, collectively electroweak interactions. Massive gauge bosons are present in this theory without gauge symmetry being violated, and a massive scalar boson is introduced. The theory holds up to experiment, with all predicted particles having been discovered. The W [50, 51] and Z [52, 53] bosons were discovered in the 1980s at the CERN Super Proton Synchrotron (SPS). The H boson was the last particle in the SM to be discovered, and its observation is currently the crowning achievement of the LHC physics program [14, 15]. Fig. 2.3 and 2.4 show the results of two of the most important channels in the CMS discovery of the H boson. Several more decay channels were analyzed and combined to make the discovery.

2.2.5 Fermion masses

While introducing the Lagrangians for free fermions and QED in equations 2.2 and 2.9, bare mass terms $m\bar{\psi}\psi$ were included without further consideration. After grouping the left-handed fermion components into $SU(2)$ doublets, such mass terms violate gauge symmetry and can no longer be included in the Lagrangian. Quarks and charged leptons are experimentally known to be massive, so alternative mass terms need to be introduced into the SM Lagrangian. The solution is once again provided by the scalar doublet introduced in Equation 2.19. For leptons and down-type quarks we can write gauge invariant mass terms as:

$$\mathcal{L}_{\text{lepton and down-type quark masses}} = -y\bar{\Psi}_L\Phi\psi_R - y^*\bar{\psi}_R\Phi^\dagger\Psi_L, \quad (2.23)$$

where the mass of each fermion is determined by the coupling constant y . To obtain a gauge invariant mass term for up-type quarks, $\tilde{\Phi} = i\sigma^2\Phi^*$ is introduced, which lets us write the following:

$$\mathcal{L}_{\text{up-type quark mass}} = -y\bar{\Psi}_L\tilde{\Phi}\psi_R - y^*\bar{\psi}_R\tilde{\Phi}^\dagger\Psi_L. \quad (2.24)$$

The scalar doublet introduced to allow for massive gauge bosons in the electroweak Lagrangian also provides gauge invariant fermion mass terms. The fermion mass terms written above are called

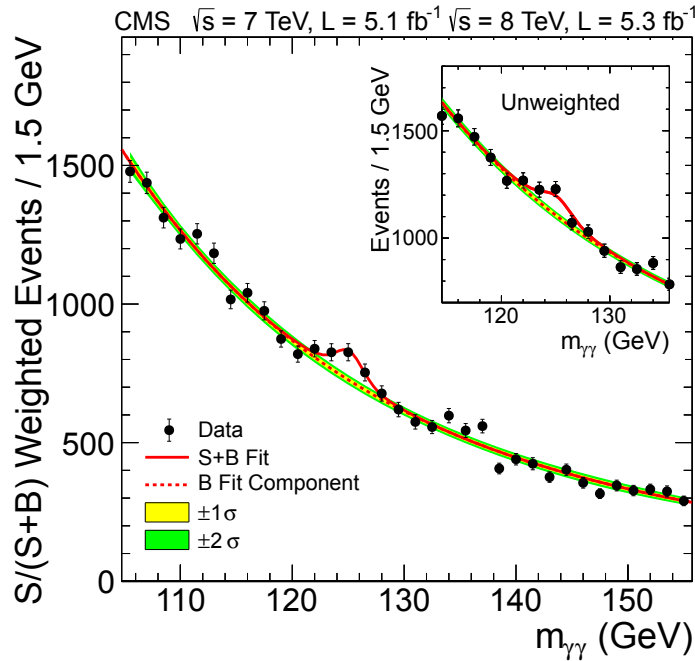


Figure 2.4: Observed and expected weighted event yields in the CMS H boson search as a function of the diphoton invariant mass in events with two photons. An excess of events over the background-only hypothesis, compatible with a SM H boson with a mass of 125 GeV, can be seen [15].

Yukawa couplings. While the electroweak theory gives a prediction for the W and Z boson masses in terms of the vacuum expectation value and the coupling constants of the electroweak forces, it contains no predictive power for fermion masses. The fermion masses, or equivalently their Yukawa coupling constants y , are free parameters of the theory and have to be measured.

Mass terms of the form predicted here imply that heavier fermions have stronger coupling to the H boson. This provides an experimental means for probing the correctness of the mass terms that were introduced: The decay rates of the H boson must be higher to heavier fermions. Measurements of the H boson production cross section in its decay modes to b quarks [54, 55] and τ leptons [56–59] are compatible with the SM fermion mass terms.

Given the large top quark mass, H bosons can not decay to the top quark, but measurements of the production cross section of top quark pair production in association with an H boson ($t\bar{t}H$) also indicate that the top quark owes its mass to the Yukawa couplings as predicted by the SM [60–63].

Note that there was no mention of neutrino masses in this section, even though it has been experimentally established that neutrinos do in fact have masses. More on this in Section 2.3.5.

2.2.6 Mixing matrices

Thus far, it was left open which up-type and down-type quark states actually appear in the SU(2) doublets of Equation 2.17. It turns out that these are not the mass eigenstates which we identify as the different quark flavors, but admixtures of them. This mixing is described in the unitary Cabibbo–Kobayashi–Maskawa (CKM) matrix [64, 65]. The couplings between different generations of quarks are relatively small, but nonzero. This means that an s quark can decay to an u quark, a c quark to a d quark and so on. Aside from mixing the weak interaction doublets and mass eigenstates the CKM matrix also introduces CP violation into the quark sector of the SM through complex phases [65].

The Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix [66, 67] is a similar unitary mixing ma-

trix for the leptonic sector of the SM. It describes the mixing of the different neutrino mass eigenstates into the states appearing in the weak interaction doublet of Equation 2.16. The fact that there is such a mixing was observed through neutrino oscillations [18, 19], which also imply that neutrinos are massive particles.

2.3 Open problems

The SM, briefly introduced in the preceding sections, has proven itself to be remarkably successful over the years. The predictions of the theory have been found to match experimental observations in particle physics in countless experiments and at great precision. Despite its monumental success, many clues point towards the conclusion that the SM is not a complete theory of nature. A host of phenomena and experimental observations can not be explained adequately, or at all, by the SM. Some of the most pressing questions that are left unanswered by the SM are succinctly summarized in sections 2.3.1 to 2.3.7. Each section gives a brief indication of the relationship between the presented problem and the work in this dissertation.

2.3.1 Gravity

Perhaps the most blatant hole in the SM is that it describes only three of the four known fundamental forces. Gravity is not accounted for in any way. General relativity, the modern theory of gravity, is formulated in the framework of classical physics and seems to be fundamentally incompatible with the quantum mechanical SM. General relativity itself also seems to be incomplete as it predicts singularities in curvature of space-time at the center of black holes. Singularities in a model are a strong indication of its inadequacy in a particular regime. As of now there is no universally accepted quantum theory of gravitation. Because of the weakness of the gravitational force, quantum gravity effects are only expected to become relevant at the Planck energy scales of about 10^{28} eV. This is far out of reach of the highest energy colliders, and will in all likelihood be for many generations to come. The Planck energy scale is also orders of magnitude higher than the highest energy cosmic rays that have been observed [68]. Experimental evidence on quantum gravity is unlikely to become available in the foreseeable future, and this topic is not addressed further in this dissertation.

2.3.2 Dark matter

Other strong hints at the incompleteness of the SM come from astronomical and cosmological observations. Known stable matter consists of the light SM fermions, but mounting evidence suggests the existence of prodigious quantities of matter that interacts via neither the electromagnetic nor strong interactions. This hypothetical matter is called dark matter and is not accounted for in the SM. Given their low masses, SM neutrinos alone can not make up this weakly interacting matter. Early hints at the existence of unobserved matter in the universe came from studies of stellar velocity dispersion [69–73] and galactic rotation curves [74–77].

The aforementioned results might also be addressed by theories of modified dynamics [78], but observational evidence for dark matter has since emerged from a variety of other sources. The mass distributions of galaxies and clusters can be estimated by their gravitational lensing of light, and the observed mass distributions are not compatible with that of the luminous matter alone [79]. A particularly striking example is the Bullet Cluster, two colliding galaxies in which a large mass component is found not to overlap with the gas or stars visible in either galaxy [80]. Surveys of the cosmic microwave background also give strong indications for the existence of dark matter. Ordinary and dark matter affect the cosmic microwave background differently. In the early universe the former interacted strongly with radiation while the latter only affected it through gravity. A decomposition of the anisotropies in the cosmic microwave background into their angular power spectrum can be used to estimate the dark matter density of the universe [81, 82]. The observed spectrum is difficult to reconcile with theories of modified dynamics. All in all dark matter is thought to make up more

than 84% of the total matter density of the universe [29, 83]. The SM only accounts for the remaining fraction.

All current evidence for dark matter is indirect. Dark matter is presumably made up out of as of yet undiscovered particles. Experiments of many kinds have sought to provide direct evidence for dark matter, but so far without success. Several types of dark matter searches are common. Searches for the production of dark matter particles at colliders try to identify collision events compatible with one or more dark matter models [29, 84]. Direct detection experiments try to observe the scattering of galactic dark matter off atomic nuclei [29, 85]. Finally, there are a host of searches for axions, a hypothetical particle introduced to explain the CP symmetry of QCD [86, 87]. Axions are possible dark matter candidates, and most dedicated searches try to detect them in strong magnetic fields via their expected coupling to photons [88].

Two of the publications exhibited in this dissertation try to address the question of the origin of dark matter. Chapters 7 and 9 both present direct searches for hypothetical particles, and each of them could have provided answers to the dark matter riddle. In the first analysis, an unstable heavy sterile neutrino is sought. This heavy sterile neutrino itself is not a dark matter candidate, but its existence might indicate it is one of several sterile neutrinos. The lightest sterile neutrino could make up much of the dark matter in the universe [89]. The other search for new physics aims to unearth the direct electroweak production of supersymmetric particles. In many of the targeted models the lightest supersymmetric particle is an excellent dark matter candidate.

2.3.3 Dark energy

Observational evidence suggests that the universe is expanding and that the rate of expansion is accelerating. The first indications came from type-Ia supernovae [90, 91], which are standard candles for cosmic distance measurements because of their consistent peak luminosity. Since then, many cosmological observations have confirmed the accelerating expansion of the universe [92]. The fact that the expansion accelerates despite the force of gravity either implies that the vacuum has a nonzero energy density, or that general relativity must be modified at cosmological scales [93].

Current estimates indicate that if the vacuum energy density is to explain the acceleration, it must make up about 70% of the energy density in the universe [29, 81, 82]. While this value seems large, and dark energy appears to be the dominant component of the universe, the value can not be reconciled with the far greater one predicted by the SM. The vacuum energy density can be computed in quantum field theory, and the result is a value about 120 orders of magnitude larger than that necessary to match cosmological observations [94].

Supersymmetry might resolve the cosmological constant problem of the SM. The theory extends the SM by predicting a fermionic partner to every SM boson and vice versa for SM fermions. These superpartners give contributions to the vacuum energy density that are equal and opposite to those of their SM counterparts, if superpartners and SM particles have identical masses [95]. Unfortunately no superpartners have yet been discovered, which means that if they exist they are almost certainly more massive than their SM equivalents. More details about supersymmetry are given in Chapter 9, where a dedicated search for supersymmetry is presented. Through this search the dissertation tries to at least indirectly contribute to addressing the mystery of the vacuum energy of the universe.

2.3.4 Matter-antimatter asymmetry

The observable universe appears to have an abundance of matter and an almost complete lack of antimatter [96]. The reason for this is not yet understood, and the SM offers no suitable explanation. Several conditions must have held true in the early universe for matter to become ample, and antimatter to disappear [97].

Firstly, baryon number conservation must be violated [98]. Quarks have a positive baryon number, antiquarks a negative one, and that of all other particles is zero. The SM strictly conserves baryon

number. There is no known interaction that converts a quark into an antiquark or into any other state with a different baryon number. The SM therefore clearly does not fulfill this first condition.

In addition to baryon number violation, C and CP violation are necessary criteria [98]. Without violation of C symmetry, any process that violates baryon number to create additional quarks would be counteracted by a process creating antiquarks. Similarly, CP symmetry would imply that the production rate of left-handed quarks and right-handed antiquarks would be equal, and vice versa for the opposite handedness. The SM violates both C [20] and CP symmetry [21, 22]. Current limits on the violation of baryon number are however too stringent for the known CP violation rate of the SM to be able to explain the matter-antimatter asymmetry in the universe [99–101].

To address the problem of the matter-antimatter asymmetry in the universe, a source of baryon number violation and additional sources of CP violation are needed. The existence of sterile neutrinos might provide an answer to this question [89, 102–104]. Chapter 7 searches for these particles and slightly more details are provided there.

2.3.5 Neutrino masses

In the SM, fermions obtain their masses through the Yukawa couplings introduced in equations 2.23 and 2.24. A priori, the SM only includes fermion states that transform nontrivially under at least one of its gauge symmetries. States that only undergo trivial transformations do not interact with any of the SM forces and would be virtually impossible to detect. The weak interaction SU(2) doublet in Equation 2.16 does not include right-handed neutrinos or leptons. Unlike right-handed leptons, right-handed neutrinos also transform trivially under the SM U(1) gauge transformations. Right-handed neutrinos are therefore not present in the SM. This directly implies that the SM contains no neutrino mass terms of the form of Equation 2.24. Loop level neutrino mass terms necessarily violate lepton number and are consequently not allowed in the SM [29]. Meanwhile, the observation of neutrino oscillations establishes that neutrinos have nonzero masses [18, 19, 105].

Extensions of the SM that add additional neutrinos must add sterile neutrinos, meaning they do not interact via the SM forces. The Z boson decay width to neutrinos can be measured as the difference between the total Z boson decay width and the observed Z boson decay widths to charged leptons and hadrons. The ratio of the Z boson decay width to leptons and neutrinos then gives a precise estimate of the number of neutrinos taking part in electroweak interactions. The Z boson decay width has been precisely measured in electron-positron collisions at the Z boson resonance provided by the Large Electron-Positron Collider (LEP). The result is that there must be exactly three neutrinos interacting with the SM electroweak force, the known SM neutrinos ν_e , ν_μ , and ν_τ [106].

The simplest way to extend the SM with a sterile neutrino would be to introduce a right-handed neutrino as a component of a Dirac spinor field which also includes a left-handed SM neutrino. This would allow one to write a mass term similar to that in Equation 2.24. Right-handed neutrinos would be singlets under all the SM gauge groups and therefore feel none of the SM forces. Their only interaction with the SM would be through potential mixing with other neutrino flavors (Section 2.2.6) and their exceptionally weak coupling to the H boson.

This approach is not the only solution and leaves loose ends. First of all, there is no particular reason for Equation 2.24 to be the only mass term. If sterile neutrinos are introduced one could write a bare mass term that does not involve the scalar field Φ in the Lagrangian without violating gauge symmetry. Such a mass term would have to involve neutrino fields that are equal to their charge conjugate. These are known as Majorana neutrinos. A bare neutrino mass term would violate lepton number. Hypothesizing that no such mass term is present, and thus only allowing massive Dirac neutrinos, would be equivalent to introducing an ad hoc requirement for lepton number conservation into the theory. In the SM the situation is notably different, lepton number is conserved as a direct consequence of gauge invariance if no gauge singlets are present [29].

A second problem with only having a Yukawa mass term with a sterile Dirac neutrino is that neutrino masses are eerily small. Neutrino masses have never been directly measured, but there are upper limits from several sources. Direct measurements of beta decays in tritium nuclei indicate that

the $\bar{\nu}_e$ mass¹ must be below 1.1 eV [29, 107]. Direct limits on the mass of ν_τ are the least stringent, namely 18.2 MeV [29, 108]. Cosmological observations and the hot big bang model place indirect constraints on the neutrino masses: The sum of the SM neutrino masses must be below 0.12 eV. All limits are given at 95% confidence level.

Measuring about 173 GeV, the top quark mass is at least twelve orders of magnitude larger than the lowest upper limits on the SM neutrino masses. As mentioned in Section 2.2.5 experimental results indicate that top quarks and other relatively heavy fermions do in fact get their masses from Yukawa couplings as predicted by the SM. If right-handed Dirac neutrinos were to be introduced in the SM, it still provides no indication of why the Yukawa couplings of neutrinos would be so small compared to those of other SM fermions.

The existence of sterile Majorana neutrinos could address these problems. No artificial symmetries would have to be introduced in the SM to accommodate neutrino masses, and the small mass values would have an explanation: It turns out that if a Majorana mass term with a mass far exceeding the electroweak symmetry breaking scale is introduced into the SM, diagonalization of the neutrino mass terms results in three very light neutrinos. This effect is known as the seesaw mechanism [105, 109–118].

Chapter 7 describes a search for sterile Majorana neutrinos that tries to directly address the open question of the origin of neutrino masses.

2.3.6 Hierarchy problem

At the time of writing, the combination of the available H boson mass measurements results in a value of 125.25 ± 0.17 GeV [29]. In the SM, the H boson mass is highly sensitive to quantum corrections from any particle that interacts with the Higgs field. The loop correction terms to the square of the H boson mass are proportional to the square of the ultraviolet energy cutoff scale of the theory. This cutoff scale is in practice the energy scale in which the SM is superseded by a more fundamental theory. In the worst case it might be the Planck scale in which quantum gravity effects are expected to become relevant (Section 2.3.1), but it could also be the energy scale of grand unification or another yet unknown model [119].

All of these scenarios have in common that the measured H boson mass seems to be too light. The bare H boson mass term μ in Equation 2.20 must cancel the gigantic mass corrections with immense precision to result in the measured H boson mass that is many orders of magnitude smaller than the correction terms. If the Planck scale is the cutoff scale of the model, this implies that the cancellation must be exact up to 30 digits behind the comma, a miraculous coincidence at face value. This question is referred to as the hierarchy problem. The hierarchy problem can also be formulated as the question of why the weak interaction is so many orders of magnitude stronger than gravity, and why their mass scales lie so far apart.

Supersymmetry is one of the potential solutions to the hierarchy problem. The H boson mass correction terms for bosons and fermions have opposite signs. If supersymmetry is unbroken, then each boson has a fermionic partner of equal mass and vice versa. All correction terms to the H boson mass would cancel exactly, and the hierarchy problem is resolved. The condition of unbroken supersymmetry does not seem to hold in practice since no superpartners have been observed, and the exclusion limits on the masses of many of them have been pushed into the TeV regime. Nonetheless, supersymmetry would still make the degree of fine tuning needed to obtain the measured H boson mass vastly smaller. More details are given in Chapter 9 where we search for the direct production of electroweak superpartners.

2.3.7 Discrepancies in high energy physics measurements

The problems in the SM laid out over the previous sections are all either theoretical in nature, or based on indirect measurements from astronomy and cosmology. There are a few recent instances in

¹If CPT symmetry is not conserved, neutrinos and antineutrinos might have different masses.

which direct high energy physics measurements have resulted in tensions with the SM.

Firstly, the best measurement of the anomalous magnetic dipole moment of the muon seems at odds with the SM. The anomalous magnetic dipole moment of a particle is defined as the total sum of higher order corrections to the tree-level magnetic dipole moment. New measurements of the anomalous muon magnetic dipole moment have pushed down the uncertainty to about half a part per million, revealing a discrepancy with the SM at a significance of 4.2 standard deviations [120]. This is the largest discrepancy that has yet been observed between a direct SM prediction and a single measurement.

The definite origin of the discrepancy remains unclear. One source could be corrections to the muon anomalous magnetic dipole moment that involve currently undiscovered particles. One of the candidate theories is supersymmetry [121]. Alternatively, and perhaps more likely, systematic uncertainties in the theoretical computations might be the culprit, and the SM might in fact be able to accurately predict the anomalous muon magnetic dipole moment [122].

Another source of tension with the SM are measurements of b hadron decays to leptons. Some of those decays seem to violate the lepton universality of the SM, which dictates that electrons, muons, and τ leptons are identical aside from their mass. Whenever a particle decays into leptons, the SM expectation is therefore that all lepton flavors occur at about the same rate in the final state. Differences arising from the lepton masses can be significant in meson decays, but are well predicted by the SM. Ratios of the measured decay rates of B mesons (mesons containing a b quark and a lighter flavor (anti)quark) to different lepton flavors do not match the SM expectations. In decays of B mesons to excited D mesons (mesons containing a c quark and a lighter flavor (anti)quark), leptons, and neutrinos, too many events with τ leptons are observed. This excess is reported by three different experiments [123–125], and when the results are combined the significance of the discrepancy with the SM expectation might be more than five standard deviations [126]. More recently, measurements of B meson decays to kaons, leptons, and neutrinos have also revealed a violation of lepton universality with a significance of 3.1 standard deviations [127]. These observed violations of lepton universality, which strictly holds in the SM, might indicate that b meson decay rates are modified by new physics.

Recently, the mass of the W boson has been measured with a relative uncertainty of a bit more than one part in ten thousand in data from Tevatron proton-antiproton collisions [128]. This is the most precise measurement of the W boson mass to date, and the measurement seems to be at odds with the SM. The mass of the W boson can be computed in the SM given the measurements of a number of other quantities, and the uncertainty in this prediction is of a similar size to that in the measurement mentioned above [29]. The measured W boson mass is larger than the SM prediction, and the significance of the discrepancy between the values amounts to 7.0 standard deviations. This indicates that the SM prediction of the W boson mass must be reevaluated, or that the mass is modified by new physics effects [129]. Here it must be noted that older, less precise, W boson mass measurements are compatible with the SM prediction [130–132]. Supersymmetric scenarios with light superpartners might provide an explanation for the high W boson mass [133]. Searches in the multilepton final state, such as the one presented in Chapter 9, could be ideal to probe these scenarios.

None of the issues related to the anomalous muon magnetic dipole moment, b meson decays, or the W boson mass are directly addressed in this dissertation.

2.4 This work in context

It is in the context of the incompleteness of the SM that the work presented in this dissertation finds its place. Three separate publications are covered in this work. The first (Chapter 7) is a search for the production of sterile neutrinos. The particles whose existence is probed in the aforementioned analysis might provide answers to three of the open problems of the SM: The matter-antimatter asymmetry in the universe, the smallness of neutrino masses, and the origins of dark matter. More details are given in Chapter 7.

The last publication described in the dissertation (Chapter 9) describes a search for the direct

electroweak production of supersymmetric particles. Supersymmetry is one of the most well-studied extensions of the SM and has the potential to provide a dark matter candidate, solve the hierarchy problem, unify the coupling constants of the SM gauge groups, and perhaps even address the large vacuum energy of the SM. The introduction of Chapter 9 goes into more detail about supersymmetry.

Chapter 8 does not discuss a search for new physics, but a search and measurement of a SM process that remained unobserved at the time the presented analysis was carried out. The production of a single top quark in association with a Z boson is about 50 times more rare than the production of an H boson in 13 TeV pp collisions, and the process extraordinarily sensitive to modifications to the SM. This analysis does not attempt to directly address any questions left open by the SM, but could have yielded another experimental discrepancy with the SM. In that sense the aim of the publication at its outset was similar to some of the measurements discussed in Section 2.3.7, but the results match the SM expectation.

Chapter 3

Large Hadron Collider

*“The cabin was built aright
the shelter put in its place:
a hundred men joined the walls
a thousand worked on the roof
building this cabin
laying down the floor.”*

Elias Lönnrot, *Kalevala*

3.1 Necessity of colliders

The majority of the particles introduced in Section 2.1 do not exist under earthly conditions. Studying these heavy particles and their interactions requires a source of such particles. For an interaction between light particles to produce massive particles, the invariant mass of the interacting system needs to exceed the sum of the masses of the heavy particles being produced. The invariant mass of a system of n particles with momenta $\vec{p}_i$ and energies E_i is a Lorentz invariant quantity given by:

$$m = \sqrt{\left(\sum_{i=1}^n E_i\right)^2 - \left(\sum_{i=1}^n \vec{p}_i\right)^2} \equiv \sqrt{s}. \quad (3.1)$$

The invariant mass is equivalent to the energy in the center of momentum frame, and its square is often called s .

In the early days of particle physics research particles were discovered in experiments with cosmic rays. More recently cosmic rays with tremendous energies have been observed, sometimes exceeding several EeV [134, 135]. Such occurrences are however exceedingly rare. If an incoming particle collides with another particle at rest the square of the center of momentum energy is given by :

$$s = m_a^2 + m_b^2 + 2m_b E_a \approx 2m_b E_a, \quad (3.2)$$

with a referring the incoming particle, and b to the target particle. The last approximation is valid if the energy of the incoming particle is much larger than the masses of both interacting particles. Reaching the LHC center of momentum energy of 13 TeV in a laboratory frame collision between protons, would require an incoming proton with an energy of close to 10^{17} eV. On Earth only several tens of thousands of cosmic rays with an equal or larger energy are expected per square kilometer per year [29]. Only a minute fraction of these collisions lead to interesting interactions in which heavy particles are produced. The rate of cosmic rays is thus by far insufficient to study most SM processes, let alone search for new physics at the TeV scale and beyond.

To increase the rate of incoming particles of a particular type or energy beyond what can be found in cosmic rays, artificial particle accelerators are necessary. If the accelerated beam is used in a fixed target experiment, i.e. made to collide with material at rest, the resulting center of mass energy rises as the square root of the beam energy following Equation 3.2. The highest beam energy currently achieved is 6.5 TeV in the LHC, which would lead to a center of mass energy of about 116 GeV on a proton target. This is not enough to produce many of the heavy SM particles.

To achieve the highest possible energies in a laboratory, particle colliders in which beams are accelerated in opposite directions and made to collide head on, are used. For such collisions the center of momentum energy is given by:

$$s = (E_a + E_b)^2 = 4E^2, \quad (3.3)$$

where the last equality only holds for two beams of the same energy. Far higher center of momentum energies can be achieved in this way than in fixed target collisions because of the linear increase of $\sqrt{s}$ with the beam energy. To perform accurate measurements of the properties of heavy SM particles, or to discover new massive particles, colliders are the tool of choice. They allow us to achieve high center of mass energies, while also allowing for a far higher collision intensity at a given energy than can be observed in high energy cosmic rays [49, 136].

In the rest of this dissertation we will use the terms center of momentum and center of mass energy interchangeably. This is motivated by the implicit assumption that all mentioned collisions are between particles of the same mass.

3.2 Cross sections and luminosity

To produce a particle or state with a certain mass we require a high enough center of mass energy as discussed in the preceding section. However, only achieving a certain center of mass energy is not enough to study heavy particles and their interactions. In order for a process to be studied in a collider, it has to occur with a nonnegligible rate. The number of events from a particular physical process expected to occur in a collider over a period of time is given by:

$$N_{\text{events}}^{\text{process}} = \sigma \int \lambda(t) dt = \sigma L, \quad (3.4)$$

with σ the cross section of the process, and λ and L respectively standing for the instantaneous and integrated luminosity. A cross section represents the quantum mechanical probability that a certain interaction will occur, and is a fundamental property of the process itself. The instantaneous luminosity, on the other hand, is related to the flux of incoming beam particles, and is a characteristic of the collider. The quantity that ultimately decides the expected number of events at a given cross section is the instantaneous luminosity integrated over the time a collider operates, the integrated luminosity. Many SM processes, such as the production of an H boson or even more so single top quark production in association with a Z boson, discussed in Chapter 8, have very small cross sections. They are expected only to occur a handful of times in billions of high energy pp collisions, or even less. The cross sections of the mentioned processes, and many others in pp collisions, can be found in Fig. 6.2 later in this dissertation. To be able to study such processes a collider with a high luminosity is required.

In summary, the luminosity and center of mass energy are the two most important traits of a collider. The center of mass energy determines what particles or processes can be produced in the collider, while the luminosity determines if the process will occur at a sufficient rate, if at all, to be studied.

Particle beams in colliders are typically composed of several bunches of particles that travel through the machine. The instantaneous luminosity of two colliding beams in a circular collider is given by:

$$\lambda = \frac{N_1 N_2 f_{rev} n_b}{2\pi \sqrt{s_{1x}^2 + s_{2x}^2} \sqrt{s_{1y}^2 + s_{2y}^2}} \cdot F \cdot W, \quad (3.5)$$

where 1 and 2 refer to the interacting beams, N represents the number of particles in a bunch, n_b the number of bunches, f_{rev} the revolution frequency of a bunch around the collider, and s the extent of a bunch in the horizontal (x) or vertical (y) direction perpendicular to the beam axis (the coordinate system is defined in detail in Section 4.1). The luminosity is reduced by a factor F due to the crossing angle at the interaction point and a W due to the transverse offset at the collision point [137, 138].

3.3 Particle colliders

Particle accelerators ordinarily consist of a vacuum pipe, which can be either circular or linear, with accelerating cavities that boost particles. Circular accelerators require a magnetic field to bend the particle paths along the trajectory of the accelerator. Linear accelerators, on the other hand, are limited by their length and can not reuse the same beam of particles.

The type of particle being accelerated largely determines the sort of measurements and discoveries that can be made. The practical candidates are protons, electrons, and their antiparticles because they are stable. More recently the idea of muon colliders has been introduced [139], but this goes beyond the scope of this dissertation.

Electron-positron colliders provide a clean collision environment. The colliding particles annihilate and have a well determined center of mass energy, the sum of the beam energies. This type of collider led to many important discoveries such as one of the two J/ψ meson discoveries [140] and the observation of the τ lepton [141]. Despite the past success with electron-positron colliders, pursuing the ever higher collision energies needed for new discoveries hits a wall with this kind of accelerator. In circular accelerators the energy loss due to synchrotron radiation rises with the fourth power of the beam energy, and the inverse of the bending radius of the accelerator. The energy loss is also inversely proportional to the fourth power of the mass of the particle being accelerated. Because of the low electron mass, the beam energy that can be practically achieved in an electron-positron collider is limited by the power of the accelerating cavities [137]. In this case building larger circular accelerators only leads to meager increases in center of mass energy. Linear accelerators do not suffer from this problem, but have other constraints. Their collision energy is limited by their length, and it is harder to sustain high intensity collisions.

Because of the inverse proportionality of the synchrotron radiation energy loss with the fourth power of the particle mass, heavier particles allow for higher beam energies. A proton is roughly 2000 times more massive than an electron and as a result the accelerating cavities are not a limiting factor in proton or antiproton acceleration. The momentum of an accelerated particle and the strength of the magnetic field in the accelerator determine the radius of the particle trajectory. The accelerator size and the magnetic field strength are therefore the limiting factors in hadron accelerators. To achieve the highest possible collision energies one would want a collider as large as possible, with a bending magnetic field as strong as possible [137]. Historically hadron accelerators have become bigger and bigger, with increasingly powerful magnetic fields. The LHC is the current culmination of this trend. An overview of different particle accelerators and their center of mass energies is given in Fig. 3.1. While linear accelerators present certain advantages for electron-positron collisions, they are not practical for colliding hadrons.

While proton-proton, or proton-antiproton colliders allow for much higher center of mass energies than electron-positron colliders, they have the added complication that hadrons are composite particles consisting of valence and sea quarks, as well as gluons. In inelastic hadron collisions, two partons, each carrying a fraction of the momentum of their parent hadron, interact as depicted in Fig. 3.2. The interaction between the partons takes place at a center of mass energy:

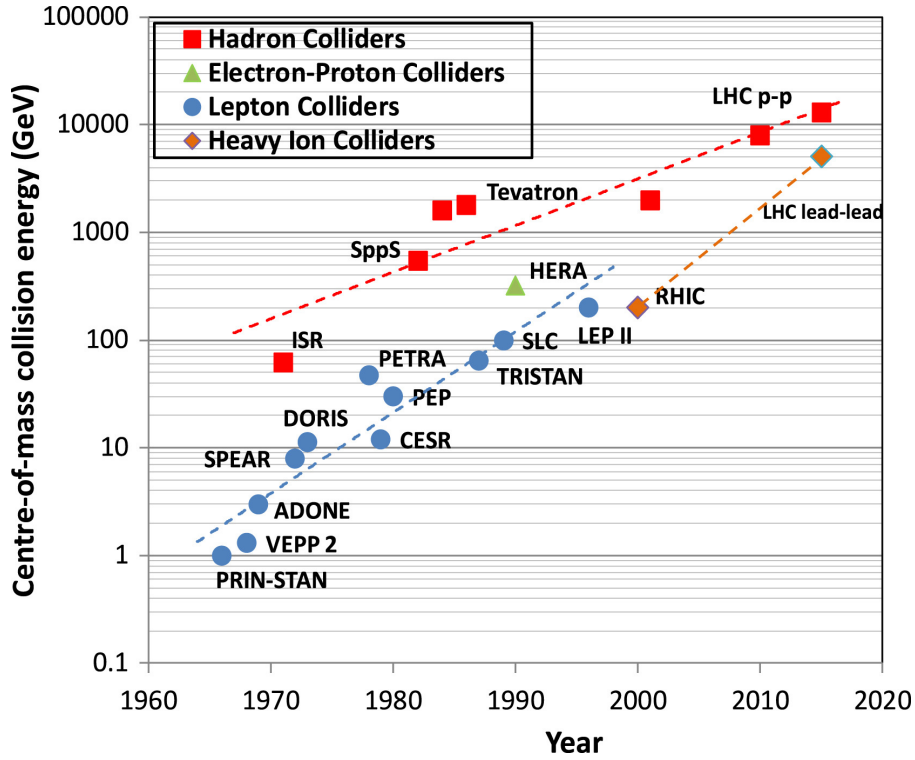


Figure 3.1: Center of mass energy of particle accelerators over time. Hadron colliders can be seen to achieve far higher center of mass energies than lepton colliders [142].

$$\sqrt{s'} = \sqrt{x_1 x_2 s}, \quad (3.6)$$

in which s is the center of mass energy of the hadron collision and x_1 and x_2 are the momentum fractions carried by the two interacting partons. This means that only particles with masses up to $\sqrt{s'}$ can be produced in this collision. The effective center of mass energy of the interacting partons in a hadron collision is a priori unknown, and depends on the parton distribution functions (PDFs) that describe the probability density that a certain type of parton carries a particular fraction of the total hadron momentum [143]. The PDFs are experimentally determined from a combination of measurements, primarily using deep-inelastic scattering [29].

In Proton-antiproton collisions the valence (anti)quarks, which tend to carry a larger momentum fraction than the other hadron components, can annihilate to form heavy particles. Following the effective center of mass energy in Equation 3.6 this presents a significant advantage compared to proton-proton collisions where particle production requires gluons or sea quarks. Proton-antiproton machines lie at the basis of important discoveries such as the first observations of the W [50, 51] and Z bosons [52, 53] at the Super Proton-Antiproton Collider, and the discovery of the top quark at the Tevatron [144, 145]. Unfortunately producing large amounts of antiprotons is challenging, so proton-proton colliders can reach much higher collision luminosities and might be the better option to discover rare processes.

Hadron colliders are often tentatively dubbed “discovery machines” because of the properties listed above. They achieve the highest possible center of mass energies, and produce parton collisions covering a continuum of different masses. Such colliders have the potential to discover heavy new particles of unknown mass, while operating at a fixed center of mass energy.

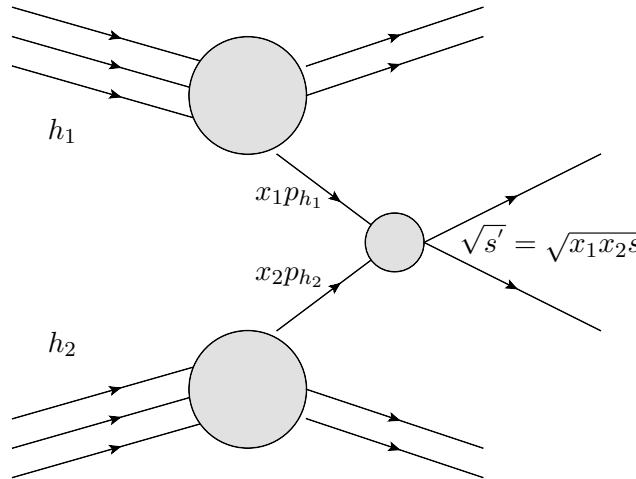


Figure 3.2: Diagram depicting an inelastic hadron-hadron collision. Two partons, respectively carrying fractions x_1 and x_2 of the momentum of their parent hadron interact. This results in a parton collision at a center of mass energy of $\sqrt{s'} = \sqrt{x_1 x_2 s}$, in which $\sqrt{s}$ the center of mass energy of the hadron collision.

3.4 Large Hadron Collider

The LHC is a 26.7 km long hadron collider, located in the old LEP tunnel. It is the largest hadron collider ever built, and is used for both pp and heavy ion collisions. This dissertation is based on data from LHC pp collisions, so this collision configuration is the only focus in what follows.

The collider was designed to collide protons at a center of mass energy of 14 TeV with a luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, adequate for the discovery of a SM-like H boson with a mass up to 1 TeV, and possibly to unearth physics beyond the SM at the TeV scale. This is the highest center of mass energy ever reached in a collider, and the highest luminosity reached in a hadron collider. The LHC operates at both the energy and intensity frontier of laboratory based particle physics experiments.

Before protons are injected into the LHC they are accelerated by a sequence of pre-accelerators at the CERN accelerator complex, shown in Fig. 3.3. The protons start as hydrogen atoms of which the electrons are stripped by an electric field. Acceleration of the protons begins in a linear accelerator called Linac 2. The protons are subsequently injected into the Booster, after which they are further accelerated by the circular Proton Synchrotron (PS), and at a later stage the SPS. When proton beams reach an energy of 450 GeV they are finally injected into the LHC.

The LHC is made up of eight arcs and eight straight sections. Four of the straight sections contain beam crossings and experimental installations. Two general purpose high luminosity experiments, CMS and A Toroidal LHC ApparatuS (ATLAS), are located at two diametrically opposed straight sections. Two other straight sections contain the LHC beauty (LHCb) and A Large Ion Collider Experiment (ALICE) experiments, as well as the injection points for proton beams coming from the SPS [137, 138].

The dipole magnets that keep protons in orbit determine the maximum beam energy that can be reached within the confines of the LHC tunnel. To reach the design center of mass energy of 14 TeV, the LHC is equipped with superconducting magnets cooled down with superfluid helium to temperatures of 1.9 K, colder than intergalactic space. These magnets have a peak dipole strength of 8.33 T, slightly in excess of what would strictly be required. Two proton beams are accelerated in opposite directions in the LHC. The same magnet system and vacuum can not be used for both beams. This problem is solved within the limited size of the available tunnel by using twin bore magnets with two sets of magnetic coils and beam channels contained in a single cryostat and mechanical encasement. This structure is shown in Fig. 3.4, and the LHC contains 1232 such magnets for bending the beam.

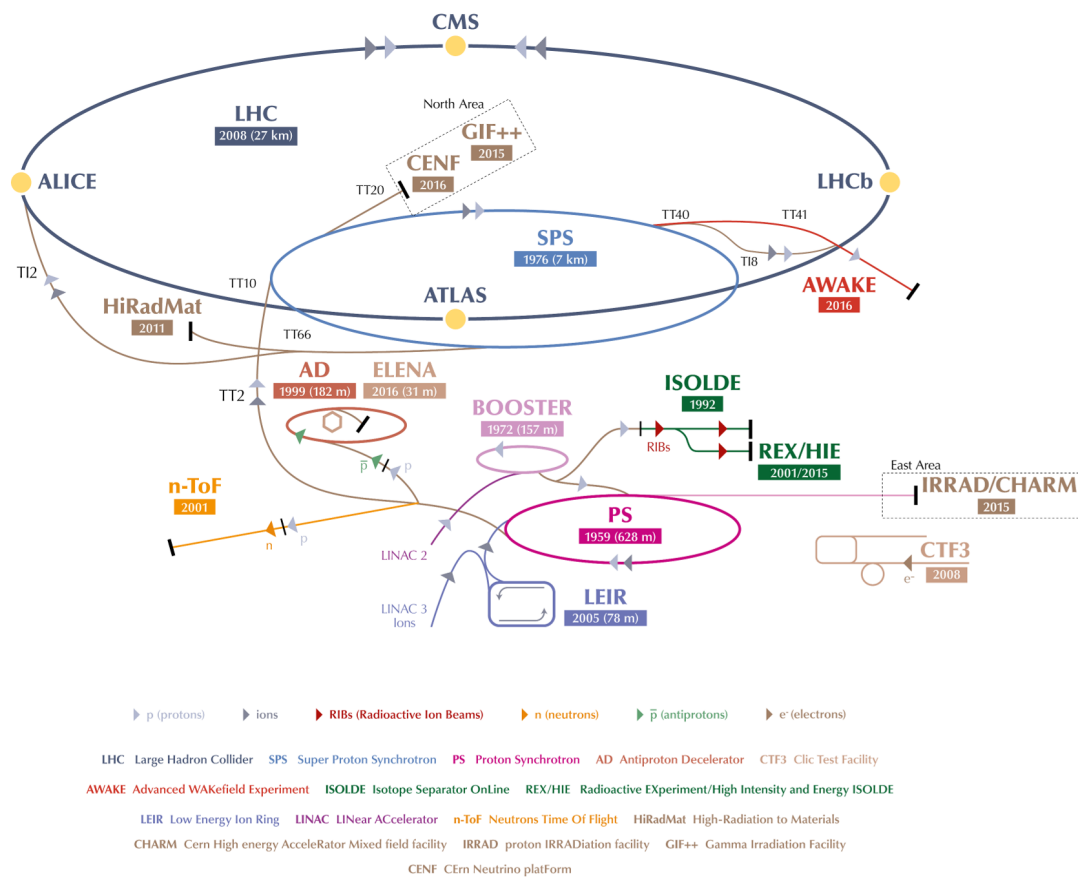


Figure 3.3: Accelerator complex at CERN. Starting at Linac 2, protons travel through the drawn chain of accelerators before finally being injected into the LHC [146].

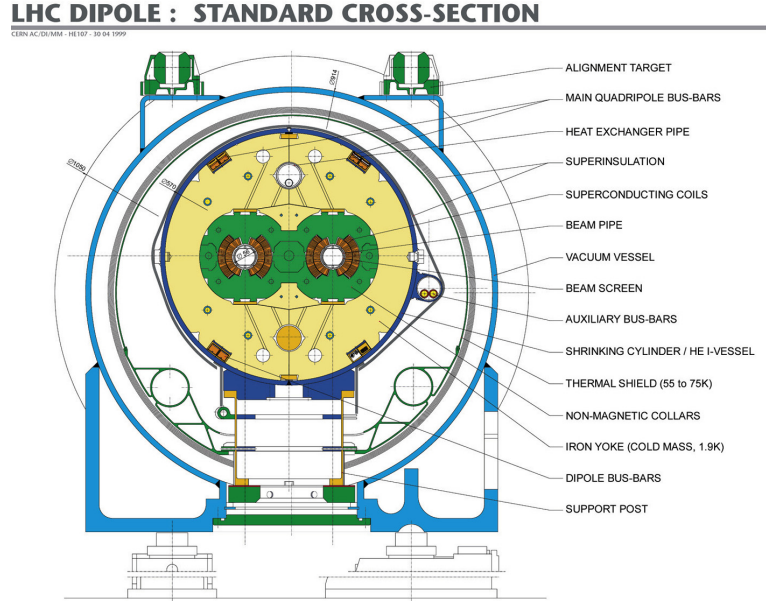


Figure 3.4: Cross section of an LHC dipole magnet [147].

Aside from the dipole magnets, the LHC contains many other multipole magnets for correcting and stabilizing the beams [137, 138].

The LHC design luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ is what is necessary to discover a 1 TeV SM-like H boson. This requirement forced the LHC to be a pp collider, as reaching the prerequisite antiproton intensity for such a luminosity is currently not feasible. By design, the accelerated protons would be grouped into 2808 bunches of $1.15 \cdot 10^{11}$ protons, separated by approximately 7.5 m. This means a bunch crossing occurs every 25 ns at the interaction points [137, 138].

In practice protons were collided at energies of 7 and 8 TeV in 2011 and 2012 respectively. The center of mass energy was increased to 13 TeV for pp collisions taking place from 2015 to 2018, collectively called Run 2. The data used for the results presented in this dissertation are all collected at 13 TeV collision energies. During Run 2 luminosities of above $2 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, twice the design value, were achieved as shown in Fig. 3.5. The LHC luminosity is so high that each bunch crossing tends to lead to multiple inelastic pp collisions. The mean number of interactions per bunch crossing is given by:

$$\langle n \rangle = \lambda \cdot \sigma^{\text{inelastic pp}} \cdot \Delta t, \quad (3.7)$$

where Δt is the time between bunches, 25 ns, and the inelastic pp interaction cross section $\sigma^{\text{inelastic pp}}$ at $\sqrt{s} = 13 \text{ TeV}$ is 68.6 mb [148]. At a luminosity of $2 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ this gives a mean of about 34 interactions per bunch crossing. The distribution of the mean number of interactions per bunch crossing observed in the CMS detector for each year of Run 2 data taking is shown in Fig. 3.6.

The instantaneous luminosity in the LHC is not constant during operation, but drops gradually after the beams start colliding¹. The main source of luminosity loss comes from the collisions in the two high luminosity experiments ATLAS and CMS. Other sources of decreasing luminosity are particle losses due to intrabeam scattering, and a build up of emittance due to intrabeam scattering and collisions with residual gas. The expected time it takes for the luminosity to reduce by a factor $1/e$ is about 15 hours [138], and in practice LHC fills during Run 2 lasted about this long in the absence of hiccups. After the beams are dumped, the LHC magnets have to be ramped down and new beams can be injected. The full LHC beam injection takes several cycles of beam acceleration by the PS and SPS. Once the beam is injected into the LHC its energy is ramped up to the nominal collision

¹Technically the instantaneous luminosity is often kept constant for a while by not operating at the maximum instantaneous luminosity right away. This is called luminosity leveling.

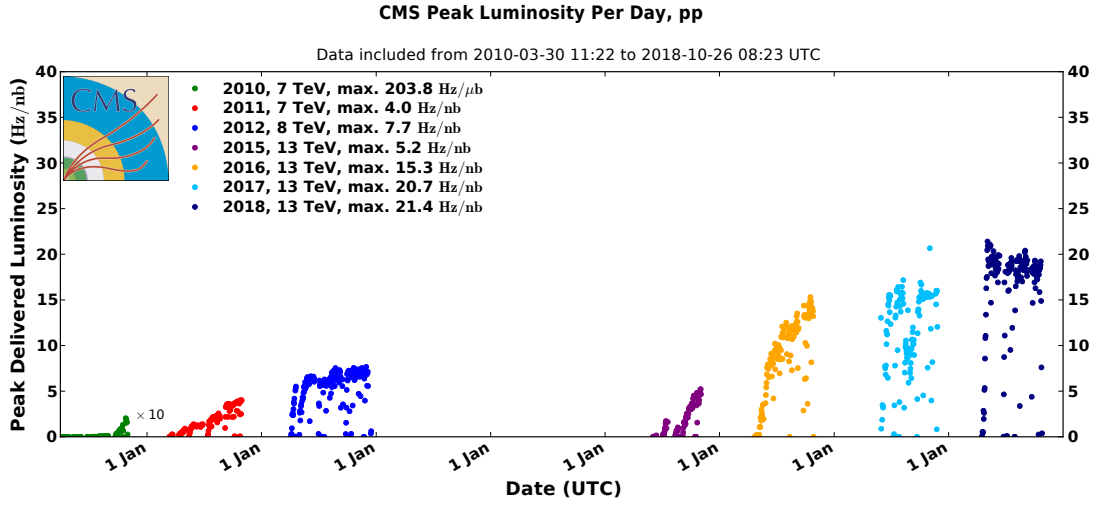


Figure 3.5: Peak instantaneous luminosity delivered by the LHC as recorded by the CMS detector through the years of operation [149].

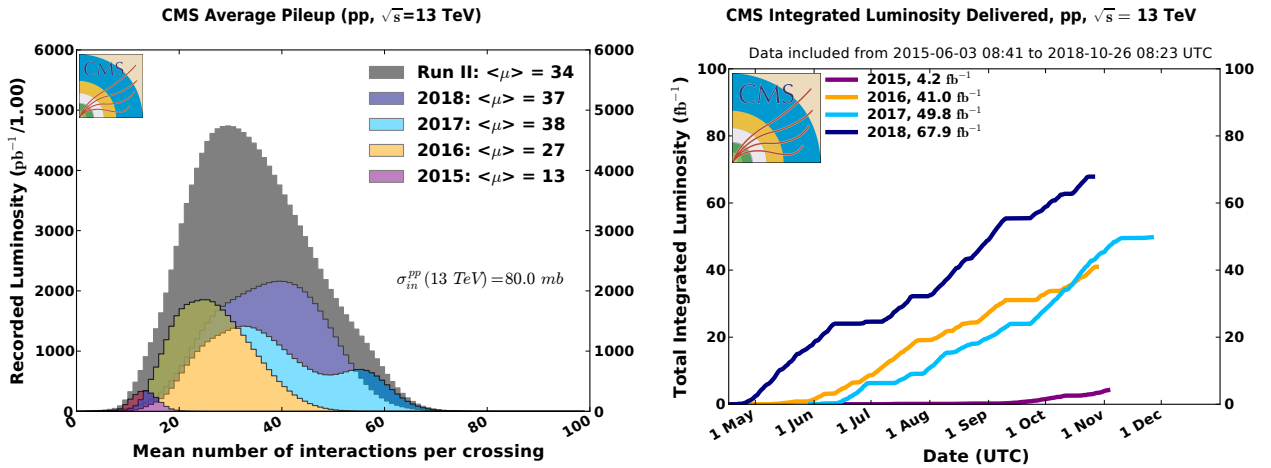


Figure 3.6: Distribution of the mean number of interactions per bunch crossing (left) and integrated luminosity delivered by the LHC (right) measured with the CMS detector for each year of pp collision data taking [149].

energy. The theoretical minimum time between fills is 70 minutes. In practice the time between fills often lasts several hours because not every proton injection is successful. During Run 2 the LHC steadily delivered collisions as shown in Fig. 3.6, amounting to a total integrated luminosity of 160 fb^{-1} . This is an unprecedented data volume for hadron collisions and enables the study of very rare SM processes, as well as searches for new physics with far greater sensitivity than those that came before the LHC. The analyses presented in this work are examples of this. These analyses use parts or all of the data collected by the CMS detector from 2016 to 2018, corresponding to a total integrated luminosity of 137 fb^{-1} . The integrated luminosity of the CMS data set used for analysis is smaller than what the LHC delivered because not all parts of the CMS detector are always operating at peak performance. Only the parts of the data set in which all CMS subdetectors performed adequately are used in the analyses presented here.

3.5 Simulating collision events

The precise simulation of collision events is a crucial ingredient in modern collider experiments. Simulation is necessary both for modeling SM expectations, and the expectations of new physics to design targeted analyses. Additionally, simulations are used to make future projections, predicting the sensitivity of next generation colliders or detectors in various measurements and searches. The analyses in this dissertation all heavily use simulation to estimate the event yields from irreducible background sources, to optimize the analysis strategies, and even to design data-driven background estimation processes. This section gives a superficial overview of the simulation of collision events and lists the software packages and tools used throughout the dissertation. Here we are only concerned with the simulation of a particle collision before the collision products interact with a detector. The simulation of a detector and the particles traversing it is another essential aspect of making realistic predictions and is succinctly covered in Section 4.9. For simplicity, we keep the text focused on pp collisions, but most of the discussion also applies to other types of particle collisions.

The vast majority of LHC pp collisions have a low momentum transfer between partons and do not result in physics of interest. Physics analyses at the LHC typically only focus on a particular final state in which a process of interest occurs. The final state of interest often occurs only once in millions or billions of pp interactions. The analyzed data set therefore only contains a minute and heavily biased sample of pp collisions. This is manifestly the case in this dissertation in which we focus on events with three or more leptons from the decay of electroweak bosons. The CMS pp collision data set collected in three years of 13 TeV operation contains several tens of thousands of these events, while the LHC provides more than a billion inelastic pp interactions per second. It is therefore impractical to simulate general pp collision events with all possible final states and interpret the results. This would require simulating an amount of collisions that is computationally intractable and can also not be stored. Instead, different physics processes are simulated separately so that those relevant to a particular analysis can be modeled with small statistical uncertainties. To estimate the total expected event yields, simulated events of all relevant processes are normalized by their corresponding cross sections and then added together.

The simulation of a single collision event can be subdivided into several distinct stages [150]. These stages are illustrated in Fig. 3.7 and described below. In most analyses, including those presented in this dissertation, events of interest contain a high momentum transfer between two of the partons in the colliding protons. This is called a hard scattering process, and often results in the creation of massive particles such as electroweak bosons or top quarks, which promptly decay into lighter particles. At high interaction energies, or equivalently small distances, QCD is perturbative (Section 2.2.3). This implies that hard scattering interactions between the partons can be modeled with fixed order computations in perturbation theory. Simulating the hard scattering process is the first step of collision event generation.

Perturbative calculations of hard scattering processes in hadron collisions can use the following factorization equation for computing the cross section to obtain a final state of interest n :

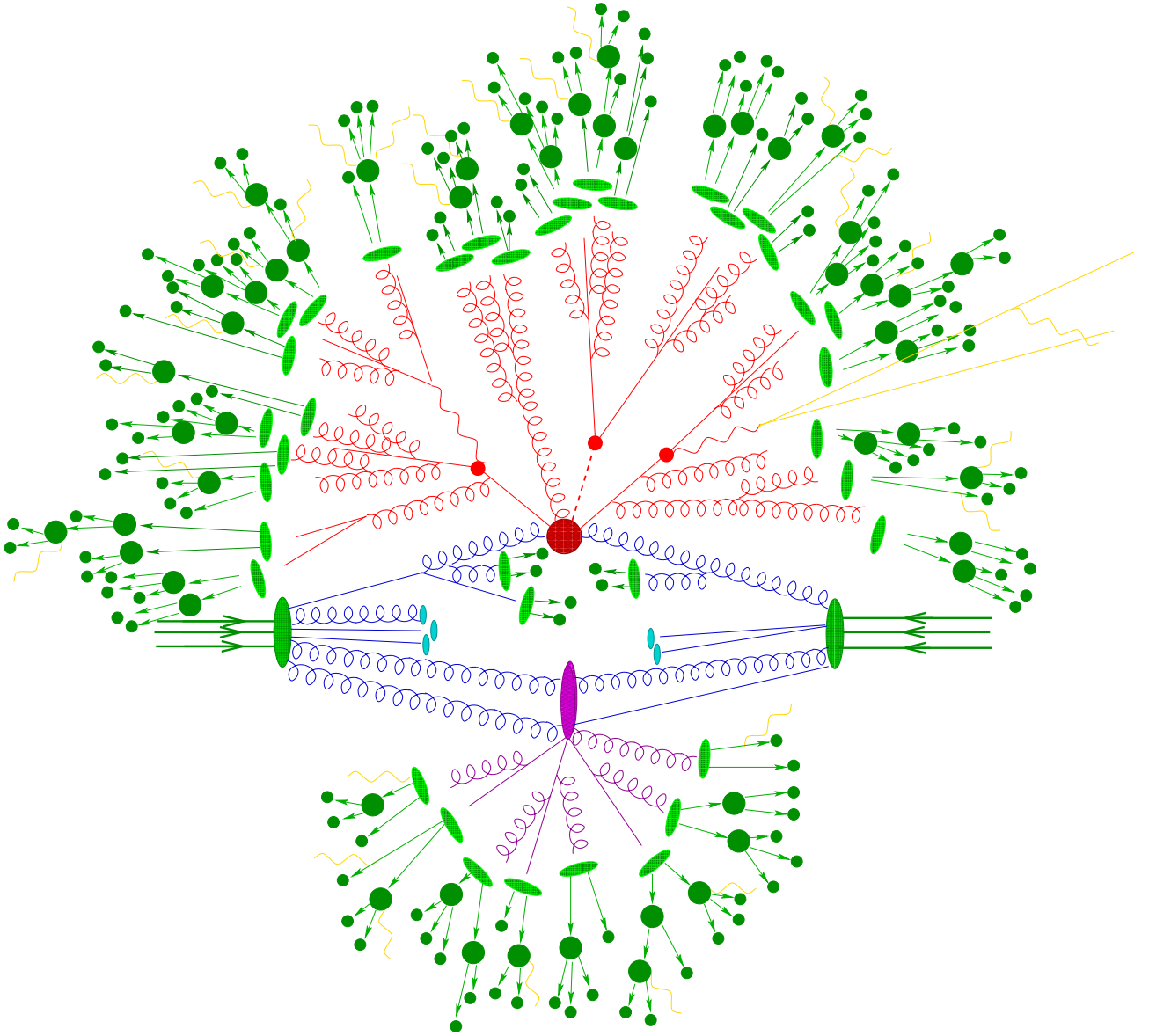


Figure 3.7: Illustration of the different stages of a pp collision resulting in the associated production of a top quark pair and an H boson. Two incoming protons from the particle beam, drawn as green ovals, collide. Interactions occur between constituent partons of the protons, drawn in blue. Two of the partons interact with a high momentum transfer, producing two top quarks and an H boson. The top quarks radiate gluons, drawn as red helices, before decaying via the emission of W bosons, the red sine waves. Quarks and gluons, drawn as red solid lines, continue to radiate more and more gluons, resulting in an extensive parton shower. The parton shower ends locally wherever photons, leptons, or neutrinos, all drawn in yellow, are created. When quarks and gluons in the parton shower reach the nonperturbative QCD energy regime, they hadronize. The resulting heavy hadrons, drawn as light green ovals, undergo decays until only photons, leptons, neutrinos, and light hadrons are left. Hadrons are represented as ovals instead of circles to illustrate the effects of time dilation in the laboratory reference frame where the collision is observed. Aside from the red hard scattering, the constituent partons of the initial protons undergo several softer interactions which are drawn in purple. This is the underlying event, which alters the observed kinematics of the hard scattering and produces soft final state particles that might overlap with those originating from the hard scattering [151].

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int d\sigma'_{ab \rightarrow n}(\mu_F, \mu_R) f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) \quad (3.8)$$

$$= \sum_{a,b} \int_0^1 dx_a dx_b \int d\Phi_n f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) \frac{1}{2s'} |\mathcal{M}_{ab \rightarrow n}|^2(\Phi_n; \mu_F, \mu_R). \quad (3.9)$$

Here a and b represent the possible initial state particle types in each colliding hadron h_1 and h_2 , x the momentum fraction of the mother hadron carried by a parton, f the parton distribution functions of the hadrons, μ_F and μ_R the QCD factorization and renormalization scales, and $d\sigma'_{ab \rightarrow n}$ the fully differential parton level cross section for producing the final state n from the initial state partons a and b . In the second step the equation is rewritten in terms of the differential phase space element $d\Phi_n$ and the matrix element $\mathcal{M}_{ab \rightarrow n}$ of the interaction, which can be obtained by summing all Feynman diagrams up to a desired order. The matrix element itself can be split into sub matrix elements for different parton helicities h_i and colors c_j as:

$$|\mathcal{M}_{ab \rightarrow n}|^2(\Phi_n; \mu_F, \mu_R) = \sum_{h_i, c_j} |\mathcal{M}_{ab \rightarrow n}|^2(\Phi_n, h_i, c_j; \mu_F, \mu_R). \quad (3.10)$$

When simulating collision events one typically samples random initial state parton types, helicities, and color states [152].

Depending on their nature, the final state particles from a hard scatter interaction continue to radiate photons and gluons after their creation. Unlike photons, gluons have self interactions because of the nonabelian nature of QCD (Section 2.2.3). This implies that radiated gluons themselves continue to radiate gluons in a rapidly expanding cascade. This phenomenon is known as a parton-shower, an evolution from a high energy scale to lower and lower energies through the radiation of an increasing number of gluons. The modeling the parton shower is the second stage in the simulation of collision events.

Showering continues until the nonperturbative regime of QCD is reached. At this point colored gluons and quarks in the collision hadronize into colorless hadrons. Predictions can no longer be directly made from first principles in this regime, and phenomenological models must be used. The work presented here exclusively uses different versions of the PYTHIA program to simulate the parton showering and hadronization stages of collision events. To describe hadronization, PYTHIA uses a string model in which the potential between two quarks increases linearly with their distance. This is motivated by the fact that the QCD field lines between two quarks attract each other because of gluon self interactions. When two quarks travel apart with relatively high momentum, they can reach a point where the energy stored in the gluon string is large enough to create a new quark-antiquark pair. This eventually results in the formation of hadrons [40]. Many of the hadrons resulting from the hadronization of high energy quarks and gluons are heavy and unstable. These hadrons decay, and their decay products might in turn decay as well. This chain of decays continues until all hadrons in the collision event are relatively light and stable on the timescale of interactions with the detector. The modeling of the hadronization and the subsequent decays of the formed hadrons is the third essential step in simulating collision events.

In the hard scattering process typically only two partons interact with a high momentum transfer. This leaves behind colored proton remnants. These fragments also undergo hadronization. In addition, other partons inside the proton often undergo interactions, albeit usually with a lower momentum transfer than the hard scattering interaction. The combination of the hadronization of proton fragments and secondary interactions between partons in the protons is called the underlying event. The underlying event alters the kinematics of the particles from the hard scattering process. In addition, it results in low momentum particles that can have small spatial separations from those originating from the hard scattering. An accurate simulation of the underlying event is yet another crucial ingredient in simulating realistic collisions events [153].

The generation of a complete pp collision is in essence a Markov chain. Each stage in the simulation chain depends on the one that came before it, and at each step random numbers are generated with Monte Carlo (MC) methods to decide what happens. The extensive use of MC techniques in collision event generation is the reason simulated events are often referred to as MC events.

In this dissertation all samples of simulated collision events are produced by using a dedicated program to do the matrix element computations of the hard scattering interactions. All important SM processes are simulated at next-to-leading order (NLO) in perturbative QCD. An overview of SM processes that are of particular relevance to the work in this dissertation is given in Section 6.3. Diboson production processes are simulated with the POWHEG v2 generator [154–157]. The only exception to this is the gluon fusion process resulting in two Z bosons, which is simulated with MCFM [158]. The simulation of all other SM processes uses the MADGRAPH5_aMC@NLO program [159, 160]. Version 2.2.2 (2.4.2) of MADGRAPH5_aMC@NLO is used to simulate collision events corresponding to 2016 (2017 and 2018) data-taking conditions. Some minor backgrounds that make up less than 1% of the total background in each of the presented analyses, are simulated at leading order (LO) in perturbative QCD. The same versions of the MADGRAPH5_aMC@NLO generator as listed above are used in these simulations.

The modeling of the parton shower and hadronization stages of collision events under 2016 (2017 and 2018) data-taking conditions is handled by the PYTHIA 8.212 (8.230) [161] generator with the CUETP8M1 [162, 163] (CP5 [164]) underlying event tune. When combining dedicated matrix element computations with a parton shower simulator, double counting of partons must be eliminated. This is resolved with the FxFx [165] (MLM [166]) matching scheme in simulation samples generated at NLO (LO) precision in perturbative QCD. The parton distribution function sets used in the simulated collision events are NNPDF 3.0 [167] (3.1 [168]) at NLO (next-to-NLO) for simulated collisions under 2016 (2017 and 2018) data-taking circumstances.

Chapter 4

CMS detector

“When one relies on sight to perceive the world, it is like trying to stare at the galaxy through a crack in the door.”

Kreia, Knights of the Old Republic II: The Sith Lords

An ambitious high luminosity and high energy collider like the LHC requires equally advanced detectors to achieve its physics goals. The CMS detector is a general-purpose particle detector surrounding one of the interaction points where the LHC’s beams collide [169]. It has a cylindrical design, consisting of a barrel section and two endcaps, to cover as much of the solid angle around the LHC collisions as possible. The detector is structured as a series of concentric layers of subdetectors around the interaction point. Each layer aims to provide detailed measurements of the properties of particular types of particles emerging from the interaction point. The tracking detector is situated closest to the interaction point, followed by the calorimeters, and the muon system [137, 143]. The combined information of all these subdetectors is used to identify and reconstruct the particles that traverse the detector, the topic of Chapter 5.

Several characteristic features allow CMS to accomplish the physics goals of the LHC. The detector has a superconducting solenoid magnet providing a 3.8 T axial magnetic field. This facilitates the precise measurement of the momenta of muons and other charged particles. Fully silicon based inner tracking detectors make the efficient reconstruction of charged hadrons in the dense collision environment possible. Meanwhile, an electromagnetic calorimeter based on scintillating crystals provides a good energy resolution for photons and electrons [169]. A schematic representation of the CMS detector is shown in Fig. 4.1.

Given the pp inelastic cross section [148], more than one billion such interactions are expected to occur each second while the LHC is colliding protons at a luminosity of $2 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [149] and a center of mass energy of 13 TeV. This event rate has to be reduced to the order of 1000 events per second to make storage and subsequent analysis of the data possible. This poses a formidable challenge which CMS tackles with a two-tiered trigger system [169].

Section 4.1 introduces the coordinate system typically used at CMS, and throughout this dissertation. Section 4.2 defines two important quantities used to describe collision events. A brief description of each of the subsystems of the CMS detector follows. Section 4.3 introduces the superconducting magnet, Section 4.4 covers the tracking detectors, sections 4.5 and 4.6 respectively describe the electromagnetic and hadronic calorimeters, and the muon system is presented in Section 4.7. The trigger system is introduced in Section 4.8. Lastly, the simulation of the CMS detector is succinctly presented.

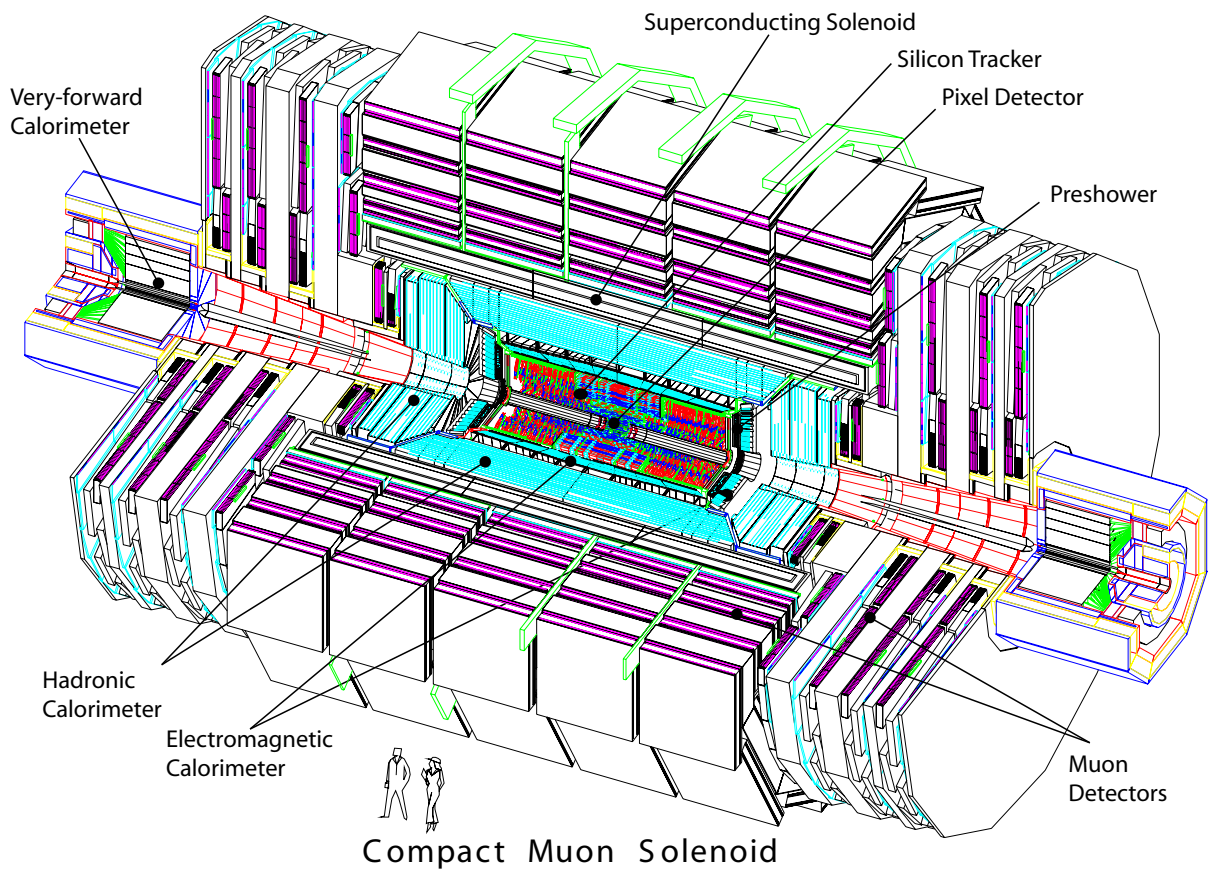


Figure 4.1: View of the CMS detector showing the different subdetectors [170].

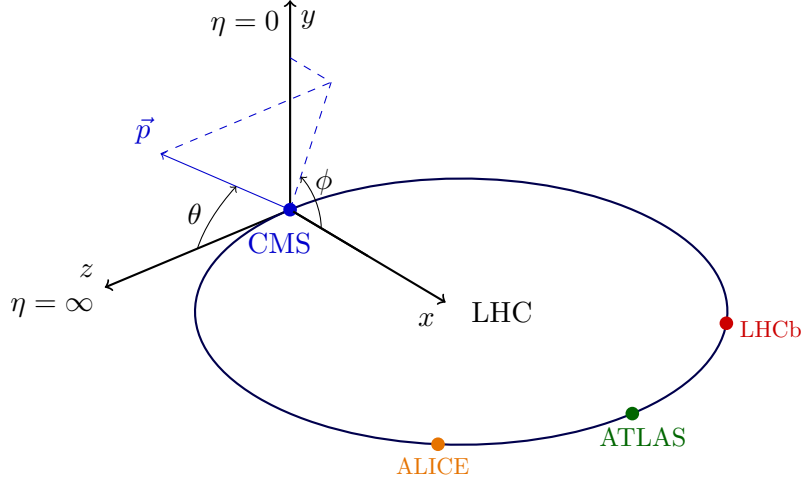


Figure 4.2: Coordinate system used in the measurement of particles traversing the CMS detector. The vector $\vec{p}$ represents the momentum of a particle originating from the center of the CMS detector. All coordinate symbols are defined in Section 4.1.

4.1 Coordinate system

An appropriate coordinate system in which particles crossing the CMS detector are measured must be defined. Here we introduce the coordinate system commonly used in the analysis of CMS data, and throughout this dissertation. An illustration of the coordinate system is given in Fig. 4.2.

The origin of the coordinate system is defined as the nominal beam interaction point at the center of CMS. The x axis points from CMS to the center of the LHC, the y axis points upwards from the center of the earth towards the sky, and the z axis is tangential to the direction of the beam at the interaction point. The coordinate system is right-handed. The azimuthal angle ϕ is defined from the positive x axis around the beam direction, and the polar angle θ is measured starting from the positive z axis. Vector components in the xy plane are called “transverse”, while “longitudinal” refers to vector components along the Z direction. The transverse (p_T) and longitudinal (p_L) momenta of a particle are given by:

$$p_T = p \cdot \sin \theta \quad (4.1)$$

$$p_L = p \cdot \cos \theta. \quad (4.2)$$

When the term transverse momentum is used in the rest of the dissertation this might refer to the scalar p_T introduced above, or a vector in the transverse plane. The meaning will be clear from the context.

The center of momentum frame of partons that interact in an inelastic pp collision is a priori unknown. The coordinate system must make it easy to handle boosts along the proton beam axis. The rapidity of a particle, defined as:

$$y = \frac{1}{2} \ln \frac{E + p_L}{E - p_L}, \quad (4.3)$$

with energy E , is suitable for this purpose. Differences in rapidity have the convenient property of being invariant under longitudinal transformations. Additionally, the particle production rate in hadron colliders does not drastically change as a function of the particle rapidity [171]. The rapidity of a particle however depends on its energy, making it impractical as a coordinate for expressing particle directions. Instead, the approximation of rapidity in the massless limit, pseudorapidity, is used:

$$\eta = \frac{1}{2} \ln \frac{|\vec{p}| + p_L}{|\vec{p}| - p_L} = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right]. \quad (4.4)$$

Pseudorapidity only depends on the polar angle, and given the large collision energies at the LHC, the massless limit holds well for final state particles. Consequently, pseudorapidity ranges are approximately invariant under longitudinal boosts. Purely transverse particles have $\eta = 0$, while $|\eta|$ approaches infinity as particles become aligned with the direction of the LHC beams. Angular separations between particles are usually expressed in terms of the quantity ΔR :

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}, \quad (4.5)$$

which is nearly invariant under longitudinal boosts. The kinematics of particles are typically given in terms of their p_T , η , ϕ , and energy.

4.2 Missing transverse momentum and impact parameters

Before starting the description of the CMS detector, some concepts which are used repeatedly in the rest of the dissertation are introduced. A good handle on these quantities is an important design consideration for the detector.

Collision events of interest typically involve a high degree of momentum transfer between two partons (Section 3.3). These partons generally do not have equal momentum. When neutrinos or hypothetical particles that only interact weakly are produced in a collision event, they almost invariably escape detection. It is therefore impossible to know what the center of momentum of an observed collision is in the longitudinal direction because weakly interacting particles might have been produced. In the transverse direction however, the total momentum is approximately zero since the transverse momenta of interacting partons are negligibly small compared to their longitudinal momenta. This means that the momentum balance in the transverse plane can be used to infer the presence of escaping particles. The negative vectorial sum of the transverse momentum components of all observed particles in an event is called the missing transverse momentum ($\vec{p}_T^{\text{miss}}$):

$$\vec{p}_T^{\text{miss}} = - \left[\sum_{x \in \text{observed particles in event}} \vec{p}_T(x) \right]. \quad (4.6)$$

Fig. 4.3 illustrates the concept. The magnitude of the missing transverse momentum is referred to as p_T^{miss} . It is an important quantity in the search for many hypothetical particles, as well as in the measurement of SM processes, particularly those that involve leptonically decaying W bosons. All of the analyses presented in this dissertation use $\vec{p}_T^{\text{miss}}$ and other quantities that depend on it. Equation 5.5 in the next chapter introduces a refined version of the simple $\vec{p}_T^{\text{miss}}$ definition given in Equation 4.6. The correction ensures that the detector response to hadronic jets is properly taken into account.

Another important notion is that of an impact parameter. The impact parameter of a particle with respect to a collision vertex is the closest approach of the extrapolated particle trajectory to the vertex. It can be used to identify the decay products of particles that have a significant lifetime. If a particle travels a macroscopic distance through the detector and decays, its decay products usually have positive impact parameter values. Impact parameters are, among other things, used in the identification of b quark jets and leptons that do not originate from the decay of electroweak bosons. For an illustration we refer to Fig. 5.5 in Chapter 5.

4.3 Superconducting magnet

The CMS detector is named after its superconducting solenoid magnet. The magnet provides a homogeneous axial magnetic field of 3.8 T inside of its coil. It is 12 m long, with a diameter of 6 m, and stores up to $2.6 \cdot 10^9$ J of energy. The tracking detectors and calorimeters are housed inside the solenoid. The magnetic flux is returned through a 10^7 kg steel return yoke, which simultaneously serves

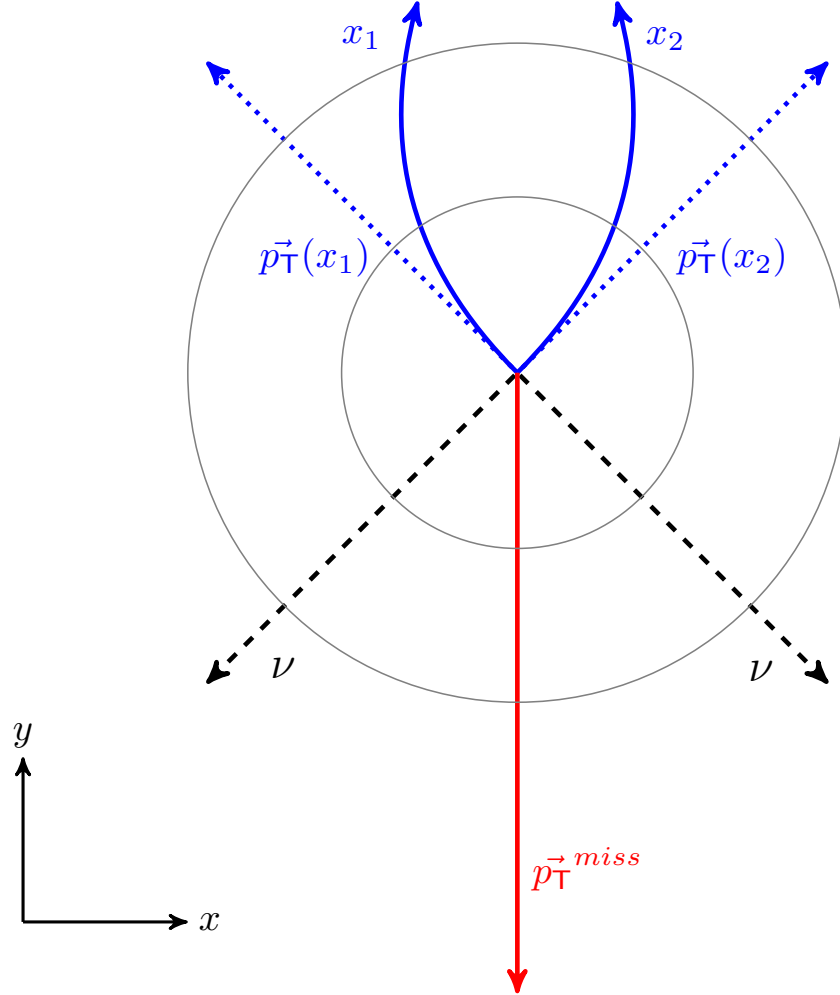


Figure 4.3: Illustration in the transverse plane of the missing transverse momentum vector $\vec{p}_T^{miss}$ in an event with two observed charged particles x_1 and x_2 (blue solid lines), and two neutrinos that escape detection (black dashed lines). The negative vectorial sum of the transverse momenta of the two charged particles (blue dotted lines) forms $\vec{p}_T^{miss}$ (red solid line). From momentum conservation in the transverse plane it can be inferred that escaping particles, the two neutrinos, were present in the event.

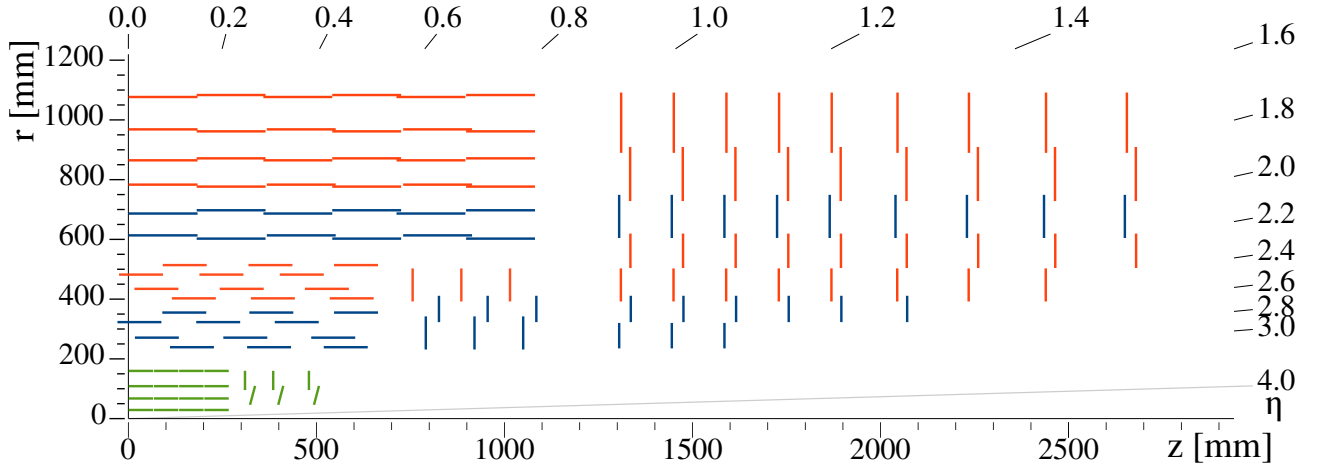


Figure 4.4: Layout of one quarter of the CMS tracking detector as of 2017 data taking. The pixel detector is shown in green, and single- and double-sided strip modules in red and blue. Prior to 2017, the pixel detector consisted of three layers in the barrel and two disks in the endcap [173].

as the absorber material for the muon system which is located outside of the solenoid. The return yoke ensures a magnetic field strength in the muon system that is sufficient for a good momentum resolution. A detailed description of the CMS magnet can be found in [169].

4.4 Tracking detector

The tracking detector is the first part of the CMS detector encountered by a particle traveling from the beam interaction point. This subdetector aims to provide a set of space point measurements that probe the trajectory of charged particles, enabling a precise reconstruction of their path through the detector. Because the solenoid magnet provides a uniform magnetic field across the volume of the tracking detector, a precise measurement of the charge and momentum of a particle can be derived from its reconstructed track. At luminosities of $2 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, more than 1000 particle tracks are expected per bunch crossing, with a new bunch crossing occurring every 25 ns. To cope with this immense particle flux, the tracking detector is highly granular, has a fast response time, and is resistant to the intense radiation environment. These goals are achieved in CMS by having the largest all-silicon tracking detector in the world at the time of writing. The amount of material used is as small as possible to limit the effects from multiple scattering, bremsstrahlung, and photon conversions [169, 172]. The tracking detector consists of a pixel detector and silicon strip tracker, both schematically drawn in Fig. 4.4. More details about these two components of the tracking detector are provided below in sections 4.4.1 and 4.4.2. The reconstruction of particle trajectories from the spatial measurements provided by the tracking detector is discussed in Section 5.1.1.

4.4.1 Pixel detector

Given the large particle flux at small radii the innermost part of the tracking detector must be pixelated in order for the occupancy to remain low. The pixel detector consists of pixel cells of size $100 \times 150 \mu\text{m}^2$, chosen to achieve a similar track resolution in both the (r, ϕ) and z coordinates. The small pixel size makes for a typical occupancy of 10^{-4} per cell.

The proximity of the pixel detector to the interaction point minimizes the effects of multiple scattering and extrapolation uncertainties in reconstructed particle trajectories. The measurements from the pixel detector are therefore essential for seeding tracks that are to be reconstructed, and for identifying secondary vertices from long lived particles such as b hadrons. The pixel detector covers the longitudinal range $|\eta| < 2.5$, and extended from radial distances of 4.4 cm to 10.2 cm prior to 2017

data taking. Back then it consisted of three layers in the barrel and two disks in each endcap. In this way the pixel detector provided three high precision space point measurements per particle [169].

The original pixel detector was designed to operate at a luminosity up to just $1 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, half of what the LHC managed to attain during Run 2. For this reason a new, lighter, pixel detector was installed after 2016 data taking. The size of the pixels remained identical, but the amount of pixels in the detector was roughly doubled. Meanwhile, the amount of material in the tracking detector volume was approximately halved. The new detector has four layers in the barrel and three disks in each endcap, and extends up to a radial distance 16.0 (16.1) cm in the barrel (endcap) region. The reduced material budget and additional measurement layers increase the resolution of all parameters of reconstructed particle tracks. A direct result is more efficient vertex and track reconstruction, with lower misidentification rates [174]. At a luminosity of $2 \cdot 10^{34} \text{ cm}^{-2}$, the first pixel layer in the barrel has a hit efficiency of 97.5%, while other layers have a hit efficiency surpassing 99%. The spatial resolution of the pixel detector is about one tenth of the size of the individual pixels [175]. More details on this are given in Section 5.1.1.1.

4.4.2 Silicon strip tracker

The hit rate density in the tracking detector decreases with the square of the distance from the interaction point. This makes it feasible to use silicon strips instead of pixels at a sufficient distance from the interaction point. The silicon strip detector extends from radii of 20 to 116 cm, and has the same longitudinal coverage as the pixel detector, up to $|\eta| < 2.5$. In the barrel, the silicon strips run parallel to the beam axis and provide measurements of the r and ϕ coordinates of traversing particles. The endcaps hold concentric rings of strips that go radially outwards along their long axis. In this way each strip provides a measurement of the ϕ and z coordinates of traversing particles. This configuration is illustrated in Fig. 4.4. Strip modules partially overlap each other and the distance between them, or “pitch”, is always 25% of the width of a strip. The strip tracker consists of several distinct parts, succinctly described below [176].

The Tracker Inner Barrel (TIB) and Tracker Inner Disks (TID) consist of four layers in the barrel and four disks in each detector endcap, extending up to radial distances of 55 cm from the interaction point. Strips with a length of about 117 mm and a pitch of 80 to 120 μm (100 to 141 μm) are used in TIB (TID). With these strip dimensions the maximal occupancy is 2-3%. The TIB and TID provide up to four position measurements along the trajectory of a traversing particle, with a resolution ranging from 13 to 38 μm . The resolution decreases with increasing strip pitch [169, 177].

The Tracker Outer Barrel (TOB) has six layers of strips, reaching up to a radial distance of 116 cm from the interaction point, and up to ± 118 cm in the longitudinal direction. A larger strip pitch, 122 to 180 μm , and length, 183 mm, are used here. Nevertheless, the resulting occupancy is only about 1%, due to the reduced hit density further from the interaction point. Six (r, ϕ) measurements are provided by TOB per traversing particle, with a resolution of about 18 to 47 μm [169, 176, 177].

The Tracker EndCaps (TEC) cover both forward regions at $124 < |z| < 282$ cm and radial distances of 22.5 to 113.5 cm. They consist of nine disks in each endcap, every disk made up of 7 concentric rings of radial silicon strips. The TEC strips have pitches ranging from 97 to 184 μm , and lengths of 85 to 200 mm. Up to nine (z, ϕ) measurements per traversing particle are provided by the TEC, with a resolution that is almost identical to that of the TOB [169, 177].

The first two layers of TIB and of TOB, as well as rings one, two, and three of TEC consist of double-sided strip modules. In such a double-sided strip module, the second module is rotated with respect to the first in the plane of the module. The rotated module layer provides a measurement of z (r) coordinate in the barrel (disks). The resolution in the additional coordinate measurement is one order of magnitude worse than the resolution in the primary coordinate. Each particle track is expected to be sampled at least nine times by the strip tracker, if the particle is within $|\eta| < 2.4$. Four or more of these measurements are expected to be from double-sided strip modules [169].

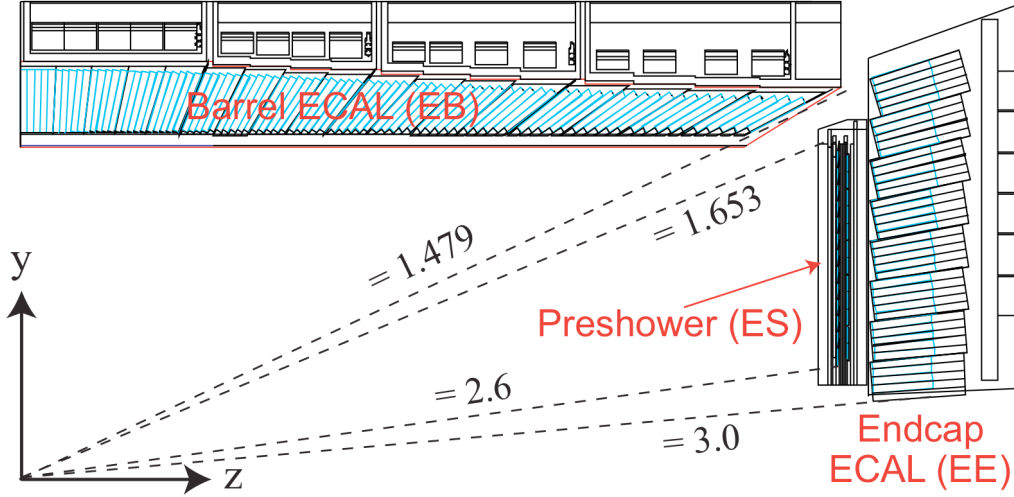


Figure 4.5: Transverse cross section of a quarter of the CMS ECAL detector [178].

4.5 Electromagnetic calorimeter

The Electromagnetic Calorimeter (ECAL) identifies and absorbs photons and electrons. It is located around the tracking detector and consists of a barrel and two endcaps that each have a preshower detector. A schematic of the ECAL is drawn in Fig. 4.5. Upon entering the ECAL, an energetic electron radiates bremsstrahlung photons, while photons convert into electron-positron pairs by interacting with the material in the ECAL. The subsequent photons and electron-positron pairs continue to interact with the calorimeter in the aforementioned way until they no longer have sufficient energy. This process gives rise to an electromagnetic shower [169].

One of the driving requirements in the design of the ECAL was the ability to discover the H boson in its decay two photons. This requires a granular and radiation resistant detector with a fast response time. These criteria are met by constructing a calorimeter out of lead tungstate (PbO_4) crystals. The small Molière radius of the material, meaning it has the tendency to produce collimated electromagnetic showers, facilitates a high granularity. In addition, lead tungstate has a high density and short radiation length of 0.89 cm, which makes it possible to build a compact calorimeter that can contain high energy electromagnetic showers. The response time of the crystals is also adequate. Their scintillation time is similar to the time between LHC bunch crossings. About 80% of the scintillation light in the ECAL is emitted within 25 ns from the initialization of a shower [169].

The ECAL barrel (EB) has a pseudorapidity coverage of $|\eta| < 1.479$, and has a 360-fold (170-fold) granularity in ϕ (η), bringing the total number of crystals up to 61200. Every crystal covers an area of approximately 0.0174×0.0174 in (η, ϕ) space, and has a length of 230 mm, corresponding to 25.8 radiation lengths. An avalanche photodiode is mounted on the back of each crystal to capture and amplify scintillation light from absorbed particles [169].

Pseudorapidities within $1.479 < |\eta| < 3.0$ are covered by the ECAL endcaps (EE). Each of the endcaps contains 7324 crystals, with cross sections of $28.6 \times 28.6 \text{ mm}^2$ in the front and $30 \times 30 \text{ mm}^2$ in the back. The constant surface area in Cartesian coordinates means the EE granularity becomes coarser in (η, ϕ) space with increasing pseudorapidity. The crystals all have a length of 220 mm, or 24.7 radiation lengths. The photodetectors of choice for the EE crystals are vacuum phototriodes [169].

A preshower detector covers the front of the EE at pseudorapidities in the interval $1.653 < |\eta| < 2.6$. The goal of the preshower detector is to distinguish photons from boosted $\pi^0 \rightarrow \gamma\gamma$ decays, which might otherwise be misidentified as high energy photons. Additionally, the preshower improves the positional resolution for electrons and photons, and helps to distinguish electrons from minimum ionizing particles. The preshower detector is made up of two layers, each a lead radiator followed by silicon strip sensors. The radiators initiate electromagnetic showers, whose transverse profile and

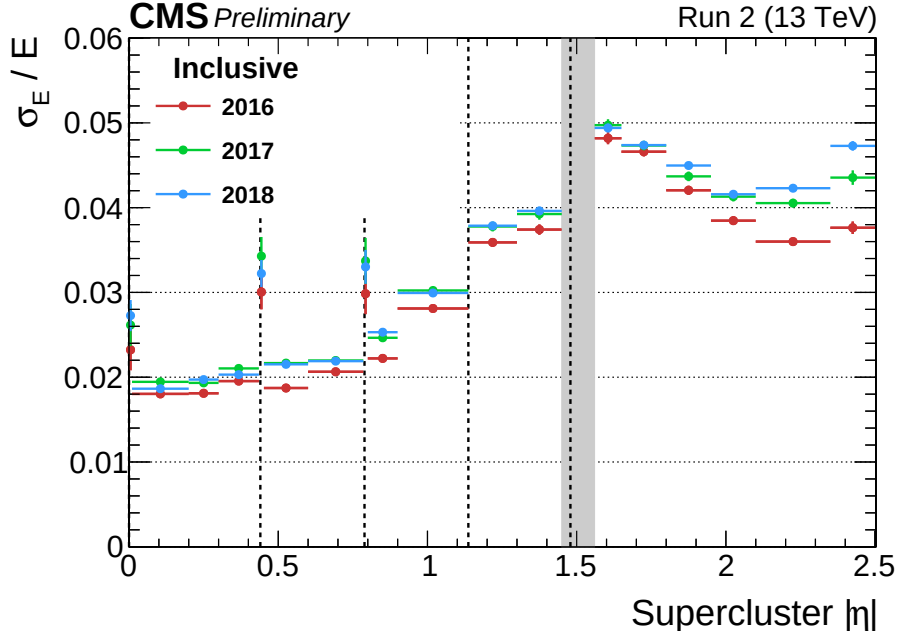


Figure 4.6: Relative energy resolution of the ECAL for electrons from $Z \rightarrow ee$ decays in unfolded bins of $|\eta|$ for each of the Run 2 data-taking years. The figure includes both electrons that emitted low and high amounts of bremsstrahlung [180].

energy deposits are measured by the silicon sensors [169].

The large amount of ionizing radiation experienced by the ECAL during LHC operation results in the production of color centers in the crystals. The color centers can absorb a fraction of the scintillation light and result in a wavelength-dependent loss of scintillation light transmission throughout data taking. The effect is corrected for by measuring the optical transparency of the crystals with laser pulses [169, 178, 179].

The relative energy resolution of the ECAL can be measured in $Z \rightarrow ee$ decays. An unbinned maximum likelihood fit to the reconstructed Z boson mass distribution using a Breit-Wigner distribution convoluted with a Gaussian as the signal model, gives the ECAL resolution as the width of the Gaussian. Such measurements are done in unfolded $|\eta|$ bins as shown in Fig. 4.6. For electrons from Z decays the relative energy resolution varies between 2% in the barrel and up to 5% in the endcaps. The ECAL resolution remained relatively stable during Run 2, regardless of the large LHC luminosity in 2017 and 2018 and the resulting reduction in crystal transparency [180]. The outstanding energy resolution of the ECAL enables a precise measurement of the H boson mass in $H \rightarrow \gamma\gamma$ decays, as demonstrated in [181].

4.6 Hadron calorimeter

The Hadron Calorimeter (HCAL) is designed to stop high momentum hadrons and measure their energy. This is essential for the reconstruction of hadronic jets and to make the detector hermetic. The latter enables the reconstruction of $\vec{p}_T^{\text{miss}}$, crucial for gauging the momentum carried away by weakly interacting particles [169, 182]. The HCAL consists of four major sections, shown in Fig. 4.7. Most of the HCAL is located inside the coil of the superconducting solenoid, around the ECAL.

The HCAL barrel (HB) and HCAL endcaps (HE) are sampling calorimeters, respectively covering the pseudorapidity intervals $|\eta| < 1.3$ and $1.3 < |\eta| < 3.0$. Brass absorber plates are used to initiate hadronic showers: Incoming hadrons interact with the nuclei in the material, producing more hadrons, which continue to interact with the material, as well as photons and electrons which trigger electromagnetic showers. The absorber plates are intertwined with plastic scintillators, equipped with silicon

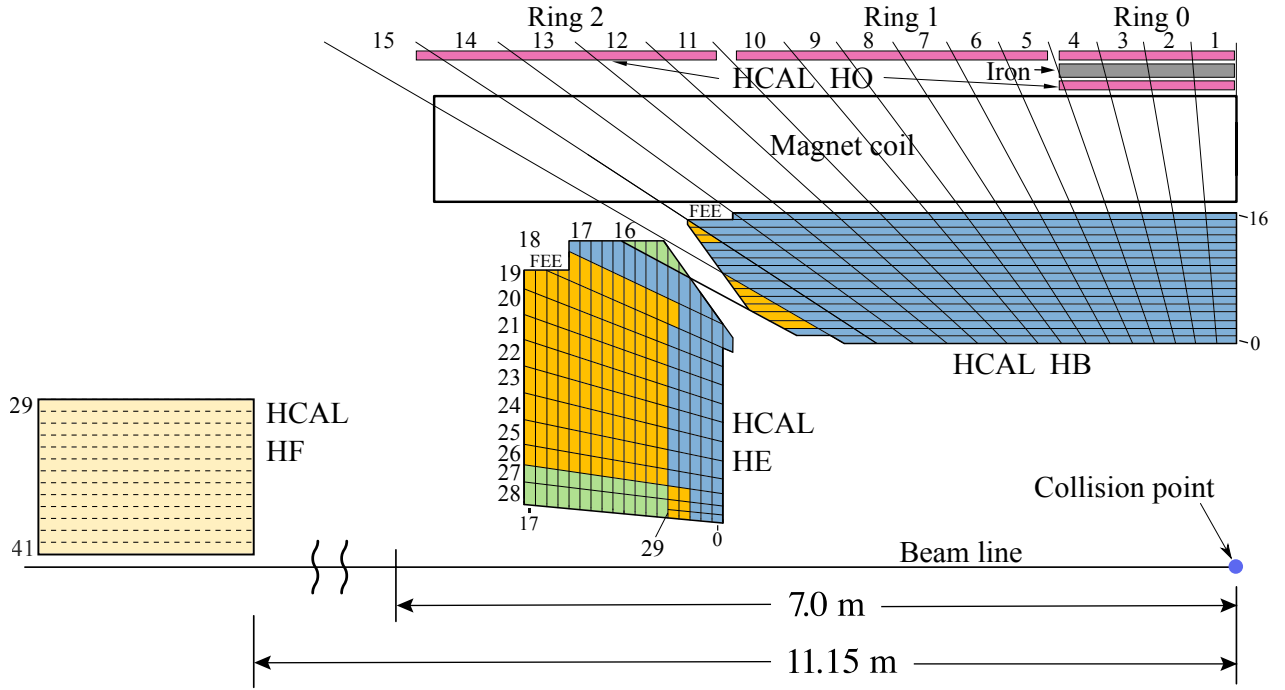


Figure 4.7: Longitudinal cross section of one-fourth of each of the HCAL subsections showing the tower segmentation. Different colors represent the grouping of the scintillator layers for readout [183].

photomultipliers [174], which sample the hadronic showers.

The hadronic shower samples can be used to determine the total energy of the shower. Brass is the material of choice for the absorber plates because of its lack of ferromagnetic qualities and its short hadronic interaction length (λ_I) of 16.42 cm. The total interaction length of the absorber material in HB is limited by the amount of space between the ECAL and the coil of the solenoid magnet. This increases in the longitudinal direction, going from $5.82 \lambda_I$ at $|\eta| = 0$ up to $10.6 \lambda_I$ at $|\eta| = 1.3$. The ECAL adds another $1.1 \lambda_I$ of material. In the endcaps the total calorimeter length, including the contribution from the ECAL, is about 10 interaction lengths. Both HB and HE are divided into towers with a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ at $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi \approx 0.17 \times 0.17$ at higher $|\eta|$. This segmentation is substantially coarser than that of the ECAL.

The stopping power of the HB and EB combination in the central region is insufficient to fully contain hadronic showers from high energy jets. To capture the escaping tails, the outer HCAL (HO) is installed. Scintillators are added beyond the magnet coil, which serves a double purpose as an absorber. Around $|\eta| = 0$ where the total absorber length is lowest, an additional 19.5 cm thick piece of iron is placed on top of the scintillators with a second layer of scintillators behind it. After including the HO, the total calorimeter has a minimum depth of $11.8 \lambda_I$, except at the boundary between barrel and endcaps.

Forward coverage up to $|\eta| < 5$ is provided by the forward HCAL (HF). This region experiences extremely high radiation levels, in particular at the largest $|\eta|$ values. The primary design consideration for HF is the ability to survive this hostile environment. Steel absorber plates, with quartz fibers running through them in parallel to the beam, are used to make a Cherenkov calorimeter. The electromagnetic component of the hadronic showers emits Cherenkov light in the quartz fiber, a small fraction of which is captured [169, 174, 179, 183].

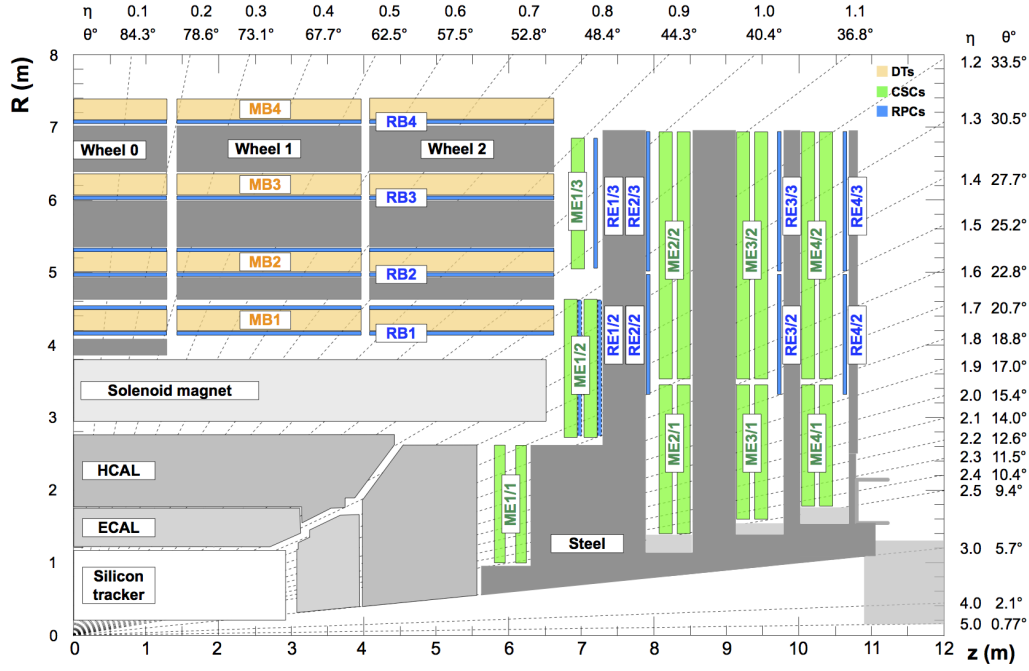


Figure 4.8: Longitudinal cross section of one-fourth of the CMS detector, with the different parts of the muon system highlighted: Drift tubes in yellow, resistive plate chambers in blue, and cathode strip chambers in green [184].

4.7 Muon system

Muon is the middle name of CMS, a highlight of the importance of muons to the CMS physics program. It is relatively straightforward to identify muons with high efficiency and precision, and to measure their momentum with good resolution. This makes them a tool of choice to tag all kinds of interesting physics, ranging from the H boson, to top quarks, to hypothetical supersymmetric particles, and more. Muons are of paramount importance in the analyses included in this dissertation. The focus of the work is on events with multiple leptons, out of which muons are identified with the highest efficiency.

The muon system is the outer layer of the CMS detector. It is located outside of the solenoid volume and housed in the return yoke of the magnet. Its cylindrical form, with a barrel and two endcaps, is guided by the shape of the solenoid. The sheer amount of material preceding, and interleaved with, the muon system makes punchthrough of hadrons rare. This results in an outstanding capacity for identifying muons. The powerful CMS magnet also leads to an excellent momentum resolution for muons. The muon system is made up of three types of gaseous detectors that cover different parts of the pseudorapidity range, indicated in Fig. 4.8. A vast area is covered by the muon system, much of it relatively inaccessible, so the detectors have to be cost-efficient and long-lasting [169, 174].

The barrel section of the muon system uses drift tube (DT) chambers and covers the pseudorapidity region $|\eta| < 1.2$. A DT is a gas-filled tube containing a charged wire. Upon passage, muons ionize the gas and the resulting electrons drift towards the wire. The position along the wires where a muon hits, together with the distance between the muon and the wires, give two dimensional measurements along muon tracks. The distance is calculated from the speed of electrons in the gas [184, 185]. The electron drift-time in DTs ranges up to 400 ns, much longer than the time between bunch crossings in the LHC. This makes DTs mainly useful in the barrel region where the muon rates and background rates from hadrons are low, and the magnetic field is uniform. The barrel muon system consists of four wheels, each holding four chambers. A chamber consists of three superlayers made up of several layers of DTs with parallel wires. The outermost superlayers of each chamber give measurements in the (r, ϕ) plane, whereas the middle superlayer measures the (r, z) position. The positional resolution of a hit is typically around $250 \mu\text{m}$ in (r, ϕ) and ranges from 250 to $600 \mu\text{m}$ in (r, z) [184].

The region $0.9 < |\eta| < 2.4$ is instrumented with cathode strip chambers (CSC). These consist of positively charged cathode strips running radially outwards, and approximately orthogonal anode wires, both contained in a gas volume. Muons knock electrons out of gas atoms, and the resulting electrons flow to the anode wire, creating an electron avalanche. This results in measurements of muon positions in the (r, z) plane. The positively charged ions flock to the cathode strips, giving a position measurement in the (r, ϕ) plane. The CSCs have a short drift path, meaning fast response times, and they can be finely segmented. This makes them ideal at high pseudorapidities where the muon and background rates are large. They are also able to cope with the nonuniform magnetic field in this region. Each endcap is instrumented with four stations made up of trapezoidal CSC chambers. Each chamber is made up of six layers, all giving two dimensional measurements of the muon position [169, 174]. The CSCs mainly measure positions in the (r, ϕ) plane precisely. This is the coordinate in the magnet bending plane, used to determine the muon p_T . The positional resolution increases from about 50 to 150 μm with r , because of the increasing distance between cathode strips [184].

To be able to precisely assign muons to the correct bunch crossing at the high LHC luminosities, a dedicated triggering detector made up of resistive plate chambers (RPC) is installed. A large voltage is applied over two parallel plates made up of a high resistivity plastic, with a narrow gas gap in between. Upon passage a muon ionizes the gas and causes an electron avalanche, which is able to penetrate the plates to reach a set of detecting strips [186]. The RPCs used in CMS consist of two such gas-gaps with readout strips between them. An RPC has a time resolution of about 1 ns, much smaller than the time between bunch crossings. Their spatial resolution, on the other hand, is worse than that of DTs and CSCs, typically of the order of 1 cm [184]. Both the barrel and part of the endcaps are instrumented with RPCs in addition to the DTs and CSC. The coverage of the RPC system extends up to $|\eta| < 1.6$ [169, 174].

4.8 Trigger system

The LHC collides proton bunches with a rate of about $4 \cdot 10^7$ Hz, massively exceeding the amount of collisions that can be stored and processed for further analysis. However, only a small fraction of proton collisions are of interest to the CMS physics program. Interesting events are selected by a two-level trigger system. Firstly, the Level-1 trigger (L1 trigger) makes fast decisions to reduce the event rate to 10^5 Hz. The L1 trigger is implemented on custom programmable electronics. Events accepted by the L1 trigger are read out and passed to the High Level Trigger (HLT), software running on a farm of commercially available processors. During Run 2, the output rate of the HLT was typically around 10^3 Hz [187, 188].

The L1 trigger uses data read out with coarse granularity from the calorimeters and the muon system to quickly identify collision events that might come from a process of interest. Meanwhile, the full resolution data is kept in a memory pipeline. The physics targeted by CMS ranges from measuring properties of the H boson, to searching for new physics, studying QCD and much more. A number of L1 trigger trigger algorithms, each called a “seed”, target different physical scenarios. Production of massive particles often result in one or more objects of high p_T and low $|\eta|$. For this reason the L1 trigger selects events containing muons, photons, electrons, and hadronic jets passing criteria on their p_T , and $|\eta|$, as well as events with a large scalar sum of particle transverse momenta or p_T^{miss} . About 75% of the events selected by L1 trigger seeds are required to have one high- p_T object, or two such objects of the same kind. Seeds for muons and electrons are designed to efficiently select those coming from the decays of W and Z bosons, or heavier hypothetical particles. These are of paramount importance to the work presented in this dissertation. Other seeds require a combination of different kinds of objects such as a lepton and a jet, or hadrons coming from a particular decay. Some seeds are “prescaled”, meaning that only one in N events passing the criteria is accepted, with N the prescale factor [188]. Triggers with prescaled seeds are used in this work for measuring the lepton fake rate, as explained in Chapter 6.

The HLT has access to the readouts of all subdetectors in full resolution, including the tracking

detector. It runs a version of the CMS reconstruction algorithms, which are introduced in Chapter 5, modified for computational speed. The HLT uses a set of “paths”, each of which is an ordered set of processing steps that reconstruct and select particles. The subsequent steps increase in computational complexity and events failing the selection at any step are discarded. Initial selections are based on information from the calorimeters and the muon system, while later steps perform the computationally demanding task of track reconstruction (Section 5.1.1). After passing all requirements specified in an HLT path, an event is permanently stored and processed for physics analysis [169, 187]. The HLT paths most important to the work in this dissertation are those selecting one or multiple electrons and muons.

4.9 Detector Simulation

Simulating pp collision events by themselves is not enough to model the event yields one is expected to observe in data collected with the CMS detector, or to design effective analyses of said data. A detailed simulation of the entire detector, and the interactions between particles and the detector is necessary. A GEANT4 [189] model of the CMS detector is used for these purposes. The GEANT4 software package can be used to build large geometric models made up of different materials, predict the interactions particles undergo while traversing the detector material, model the effects of external electromagnetic fields, and simulate the detector readout.

Collision events are first simulated with the procedures and software packages introduced in Section 3.5. The generated particles are then passed through the GEANT4 simulation of the CMS detector, resulting in readout signals similar to those obtained for real collision data. The simulated signals are then processed with the same algorithms and software used on real collision data. The algorithms used to reconstruct collision events are the subject of Chapter 5. A GEANT4 simulation of the CMS detector is a part of all simulated collision events used throughout this dissertation. The only exception are the simulated signal events used in Chapter 9, which use an alternative simulation of the CMS detector optimized for computational speed in favor of maximal accuracy. More details are given in Section 9.2.4.

Chapter 5

Event reconstruction

*“Gentlemen, we have run out of money.
It’s time to start thinking.”*

Ernest Rutherford

To reconstruct what happened in a pp collision event, the measurements from all the CMS sub-detectors have to be interpreted in terms of the particles that induced them. To this end CMS uses a particle-flow (PF) algorithm [190] which combines the information from all the subdetectors to reconstruct the individual particles that interacted with the detector during a collision event. Details on the CMS implementation of the PF paradigm are given in Section 5.1. This includes the reconstruction of muons and electrons, central to this dissertation. Identifying those electrons and muons that originate from the decay of electroweak bosons is of crucial importance in this work, but is not covered in this relatively general chapter. This chapter exclusively summarizes existing methods and includes no original work of my own. Details on the identification of leptons and the contributions I made to the corresponding techniques are instead given in Section 6.1.

Quarks and gluons produced in LHC collisions, with the exception of the top quark, develop a parton shower and hadronize (Section 2.2.3). This results in collimated jets of particles. To infer the properties of the initial parton, particles reconstructed by the PF algorithm are clustered into jets. The clustering procedure is described in Section 5.2. The energy scale and resolution of the resulting jets is corrected for the nonuniformity and nonlinearity of the detector response. The correction procedure is based on both simulations and pp collision data and is introduced in Section 5.3. The identification of jets originating from b and c quarks is the topic of Section 5.4. Jet flavor identification is an important tool for rejecting background events in the analyses of chapters 7 and 9, while it is used to tag signal events in Chapter 8. The reconstruction and identification of hadronically decaying τ leptons (τ_h) is described in Section 5.5, and is used in Chapter 9.

5.1 Particle flow reconstruction

An overview of the signals induced by different particle types in the CMS detector is given in Fig. 5.1. Muons, electrons, and charged hadrons produce tracks in the tracking detector. Electrons and photons are absorbed by the ECAL while forming an electromagnetic shower. Charged and neutral hadrons begin showering in the ECAL, but are typically only absorbed in the HCAL. Muons tend to penetrate the calorimeters, depositing a small amount of energy by ionization along the way, and produce tracks in the muon system. Correlating the measurements made by all subdetectors to reconstruct individual particles leads to an excellent event description, superior to what can be achieved when reconstructing distinct types of particles by using different parts of the detector exclusively. In typical LHC collision events over 1000 particles are produced, requiring a highly segmented detector for the application of a PF algorithm. This condition is satisfied by the CMS detector [182].

A brief overview of the reconstruction of “PF elements” from the signals in individual subdetectors

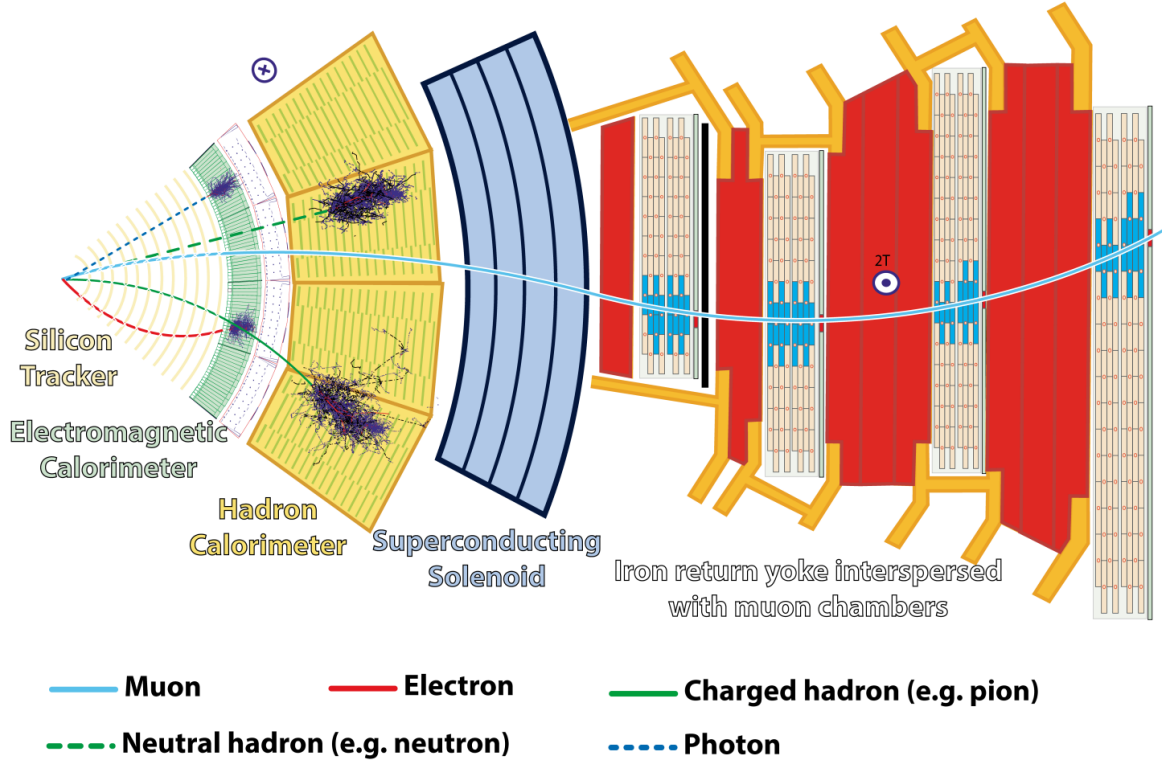


Figure 5.1: Transverse slice of the CMS detector showing how different types of particles interact with each subdetector [191].

is given in sections 5.1.1 and 5.1.2. The combination of the different PF elements into a set of reconstructed particles is discussed in Section 5.1.3.

5.1.1 Track and vertex reconstruction

The inner tracking detector and the muon system sample a succession of positions along a particle trajectory. From these measurements we aim to reconstruct the path of the particle through the detector. Many properties of a particle can be inferred from its trajectory, such as its p_T , charge, and direction. Track reconstruction commences by determining the positions of individual hits, some of which are used to form track seeds. Starting from these seeds, a track finding procedure assigns hits to tracks. When all hits in a track have been found a fit is performed to determine the parameters of the track. The reconstructed tracks in an event can be combined to identify vertices, and from these the beam spot is determined over many events. These subsequent event reconstruction stages are explained in the following subsections. Most of the discussion about track finding and fitting applies to both the inner tracking detector and the muon system.

5.1.1.1 Hit reconstruction

The first step in the track reconstruction process is the clustering of signals induced by particles that traverse the tracking detector into a set of “hits”. A single particle typically deposits a charge in several pixels when traversing a single layer of the pixel detector. Adjacent pixels with significant charge deposits are aggregated into single hits. Subsequently, the position of each hit and the uncertainty therein are reconstructed, where the position of the hit is where the particle crossed the detector. During the final track fit, hit positions are determined from fits to templates derived in simulation. The charge distribution in a pixel cluster is compared to a number of templates for particles crossing the central pixel at different angles and in different positions. Several χ^2 fits determine which template matches the data best, thus establishing the hit position. This procedure can properly account for

radiation damage in the tracker by generating new templates after some period of operation [177, 192]. During track seeding and track finding a faster procedure is used: The difference between the charges deposited in the outermost cells of the pixel cluster yields an estimate of the hit position [177].

Signals in the strip tracker are clustered if several neighboring strips measure a significant charge. The hit position is determined as a charge-weighted average of the strip positions in the cluster. Several corrections are applied, in particular to account for Lorentz drift in the magnetic field [177].

The hit resolution in the pixel detector can be determined by comparing measured and expected hit positions in particle tracks that are refitted without the hit under study. A resolution significantly smaller than the size of individual pixels is found. In the strip detector multiple scattering tends to amount to a larger effect than the hit resolution because the strips are further from interaction point. The resolution measurement for strip detector hits is carried out with the same procedure as for pixel hits, and accounts for the extrapolation uncertainty by comparing the residual differences for overlapping modules. Also in this case, the hit resolution is found to be smaller than the strip pitch [177].

5.1.1.2 Track finding and fitting

Once all hits are reconstructed they are used to estimate the trajectories of charged particles that traversed the detector. Track finding is a pattern recognition procedure in which hits are classified into groups originating from a common particle. After hits are associated to tracks, the tracks are fit again with the goal of estimating their properties at the point of origin with a precision as high as possible. The properties of tracks at their origin are the crucial piece of information for physics analyses that the tracking detector provides. Tracks in the CMS tracking detector are approximately helical because of the near uniform magnetic field inside the solenoid volume. Helical tracks can be specified by five parameters: The three coordinates of their point of origin, their p_T , and θ .

Kalman filters [193] are the basic tool used for both track finding and fitting [194–196]. A Kalman filter is an iterative fitting procedure that aims to predict the underlying state $\mathbf{s}_k$ of a physical system at different steps k in an independent variable. In the particular case of track reconstruction this means estimating the parameters of the particle trajectory at each detector plane. At each step a measurement is performed with a stochastic noise term from the measurement uncertainty that is assumed to be independent between measurements. The state $\mathbf{s}_k$ evolves dynamically between measurements and is itself also subject to stochastic noise. For a track the evolution consists of ionization energy loss and multiple scattering. For each measurement a Kalman filter estimates a prior by propagating the physical state at the previous measurement to the current one with an underlying model. The new measurement is used to update the prior and its covariance, and one obtains a refined estimate of the physical state at the current step. For track reconstruction this propagator step consists of extrapolating the particle trajectory from one measurement plane to the next. This procedure is repeated until all measurements belonging to the system are processed. Here, until all hits induced by the particle traversing the detector are processed. Fig. 5.2 illustrates the basic steps in the application of a Kalman filter to particle tracking. A more detailed description of Kalman filters can be found in [197].

The track reconstruction procedure of CMS starts by forming track seeds from a small number of hits and possibly a constraint on the origin of the track. The track seeds provide coarse initial estimates of the track parameters, and they are iteratively extended by adding hits from subsequent detector layers with the Kalman filter technique. During this stage a simple analytical propagator function is used to extrapolate the tracks from one detector layer to the next. At each detector layer only the few best track candidates, based on the track χ^2 goodness-of-fit and the number of missing hits, are retained to avoid a combinatorial explosion. Subsequently, the procedure is extended to the next detector layer. This iteration continues until all track candidates satisfying certain selection criteria are found. When two track candidates share a large amount of common hits only the best one in terms of χ^2 goodness-of-fit is kept [177].

The final step is the refitting of each track candidate. First a Kalman filter fit to a few of the

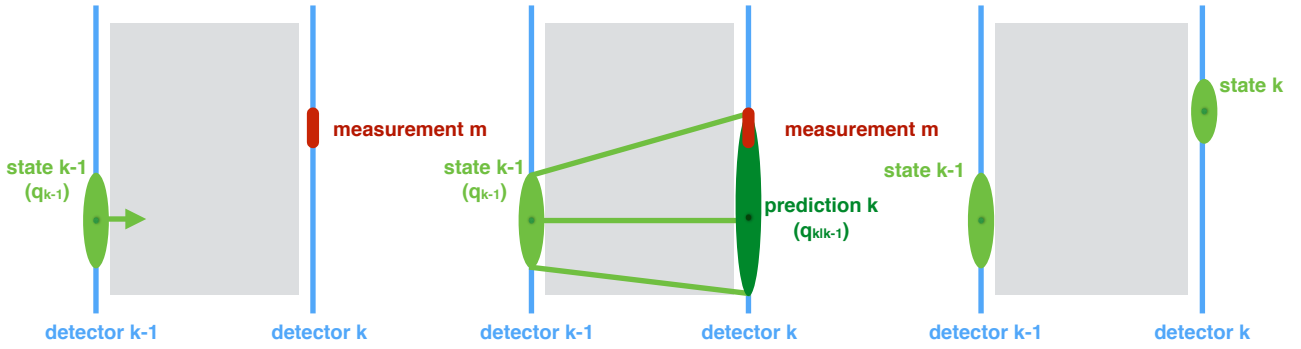


Figure 5.2: Elementary steps in a Kalman filter. One starts with a tracked state at detector layer $k - 1$ and a measurement at detector layer k (left). Given a propagation model, the state at $k - 1$ is extrapolated to detector layer k , with an associated uncertainty estimate (middle). The extrapolated state and the measurement at layer k are combined to make an estimate of the state at detector layer k (right) [198].

innermost hits of each track is done. The resulting track state and an upscaled covariance matrix are used to initialize a second Kalman filter that processes all hits in the track. The result of this fit is once again used as an initialization, also with an increased covariance, to start a third fit that processes the hits in reverse order. The final parameters of the track at each point are given by a weighted average between the results of the inward and outward going fits, yielding an optimal result at each point. The fitting step uses a highly detailed propagator function that numerically extrapolates the track, taking into account material interactions and the nonuniformity of the magnetic field. The positions of individual pixel hits are also refit during this stage with the increased knowledge about the tracks. After being fit, each track is checked for spurious hits by examining the compatibility of hits with their corresponding track. If any hits are removed from a track, it is fit again and the check is repeated, a procedure which continues until no more hits are removed [177, 199].

To limit the number of fake tracks while simultaneously retaining a high reconstruction efficiency for genuine tracks, CMS uses an iterative track reconstruction procedure. Stringent conditions are initially applied during track seeding, finding, and fitting. Hits associated to the reconstructed tracks are masked, and the entire procedure is repeated several times with gradually loosening quality criteria. At first, high- p_T “prompt” tracks, coming from the primary pp interaction vertex, are reconstructed. Later iterations focus on lower p_T tracks, and on particles displaced from the primary interaction vertex. Without this iterative algorithm very stringent selection criteria would have to be applied because the combinatorial background in track finding increases exponentially as the selection is loosened [177, 182]. Applying such stringent requirements would limit the track reconstruction efficiency considerably.

5.1.1.3 Electron tracking

Electrons radiate much of their energy when traversing the tracking detector, and the distribution of the energy loss is highly nongaussian. The Kalman filter approach used for general track finding and fitting assumes Gaussian noise terms and does not perform as well for electrons. Tracks that might originate from electrons are identified and refit with the Gaussian sum filter (GSF) algorithm [200]. In GSF the nongaussian energy loss of electrons, described by the Bethe–Heitler distribution [201], is approximated as a sum of several Gaussian components. A potential electron track is represented as a set of helical track elements, each corresponding to one of the Gaussian energy loss functions. The trajectory components all carry a weight that represents the probability this component describes the true electron trajectory. At every tracking detector layer the energy of the track is decreased and

its uncertainty increased following one of the Gaussians making up the total energy loss function. The track information is updated with the Kalman filter approach. This process is repeated for each energy loss component individually. So at every detector layer a trajectory is split into several trajectory elements, each of which is further propagated in the same way. To avoid an exponentially increasing number of trajectory components, the ones with small probability weights are merged or removed at each detector layer.

The initial identification of electron tracks that are to be fit with the GSF algorithm happens in two ways. The first is to build electron “superclusters” in the ECAL that consist of several clusters (Section 5.1.2) with similar η but spread out in ϕ to account for the bending of electrons in the magnetic field. The location and energy deposited in such a supercluster, together with the knowledge of the beam spot (Section 5.1.1.6), can be used to identify tracks compatible with the supercluster. The second method finds ECAL clusters compatible with tracks extrapolated from the tracking detector to the ECAL. These tracks must pass several quality requirements to be considered as electron seeds. To do the extrapolation, a GSF fit with up to five Gaussian energy loss components is used to improve the track finding for potential electron candidates. The track collections obtained using both approaches described above are merged into a single collection of electron track candidates. After electron track candidates and their corresponding hits have been found, a GSF fit with up to twelve energy loss components is carried out to determine the final track parameters [177, 182, 202, 203].

5.1.1.4 Muon tracking

Muon track reconstruction combines the information from the inner tracking detector and the muon system. Three types of muon candidates are defined:

- **standalone muon:** Track segments are built from hits in the CSC or DT detectors, and used to seed track finding in the muon system. After finding all DT, RPC, and CSC hits corresponding to the muon track, it is fit again. This results in the final standalone muon track.
- **global muon:** Standalone muon tracks are combined with compatible tracks in the inner tracking detector. When a suitable combination is found, the entire track is refit, including all hits in the tracking detector and the muon system. This results in a global muon track. Because of the large lever arm in the muon system, global muon tracks can improve the momentum resolution of tracking detector tracks for p_T values in excess of 200 GeV, despite the much coarser granularity of the muon system.
- **tracker muon:** When an extrapolated tracking detector track matches track segments in the muon system, the tracking detector track is considered a tracker muon track. Muon system track segments are short straight-line tracks through the multiple layers of a CSC or DT detector chamber.

Muons that penetrate several muon detector planes tend to be reconstructed as global muons. Low p_T muons might fail this selection due to scattering on the steel return yoke of the magnet. These are usually recovered as tracker muons. Muons without a track in the inner tracking detector can still be reconstructed as standalone muons. These tracks usually have a substantially worse momentum resolution and cosmic ray muons produce background tracks [177, 182, 184].

5.1.1.5 Vertex reconstruction

Once all tracks in an event are reconstructed they are combined to find the positions of the primary pp interaction vertices. A primary vertex is a vertex corresponding to an interaction between protons from the LHC beams. When multiple pp interactions happen in a single bunch crossing (Section 3.4), it is important to be able to associate the observed final state particles with the different vertices. Secondary vertices originate from the decay of particles that have a sufficient lifetime to travel a

macroscopic distance from their origin before decaying. The distinction between primary vertices and secondary vertices is important. This will become clear in Section 5.4.

Primary vertex reconstruction happens in three steps: Selecting good tracks, clustering the tracks coming from the same vertex, and fitting the positions of the vertices.

Good quality tracks are selected based on the number of hits they have in the pixel and strip tracking detectors. Additionally, they are required to have a small transverse impact parameter relative to the center of the beam spot. The reconstruction of the beam spot is the subject of Section 5.1.1.6.

Primary pp interaction vertices are identified by clustering the selected tracks by their z coordinate at their point of closest approach to the beam center. The clustering is done with a deterministic annealing algorithm [204]. Such an algorithm aims to find a global optimum in a problem with many degrees of freedom by modeling it like a physical system cooling off to an energy minimum through a series of temperature reductions. The vertex positions at “temperature” T are obtained by minimizing a quantity reminiscent of free energy in statistical mechanics:

$$F = -T \sum_i^{\text{tracks}} \log \sum_k^{\text{vertices}} \rho_k \exp \left[-\frac{(z_i^{\text{track}} - z_k^{\text{vertex}})^2}{T \sigma_{i,z}^2} \right], \quad (5.1)$$

where z_i^{track} are the z coordinates of tracks at their closest approach to the beam spot, $\sigma_{i,z}$ the corresponding uncertainty, and z_k^{vertex} the vertex z positions. Vertices are assigned a weight ρ_k representing the number of overlapping vertices at the same position. After minimizing this equation, a track i is assigned to a vertex k with probability:

$$p_{ik} = \frac{\rho_k \exp \left[-\frac{(z_i^{\text{track}} - z_k^{\text{vertex}})^2}{T \sigma_{i,z}^2} \right]}{\sum_{k'} \rho_{k'} \exp \left[-\frac{(z_i^{\text{track}} - z_{k'}^{\text{vertex}})^2}{T \sigma_{i,z}^2} \right]}, \quad (5.2)$$

in which the denominator is a normalization factor. Minimizing Equation 5.1 effectively corresponds to minimizing the log-likelihood of the assignment of tracks to vertices.

The algorithm starts with a single vertex and a very high temperature T . Because of the smearing of the resolution $\sigma_{i,z}$ by $\sqrt{T}$ tracks are equally likely to originate from any vertex at the early stages of the algorithm. The temperature is gradually decreased and at each step the free energy is minimized with respect to z_k^{vertex} and ρ_k to determine the vertices and their positions. When the minimization reaches a saddle point and can no longer continue, stale vertices are replaced by two nearby vertices before further decreasing the temperature.

The algorithm ends up assigning tracks to vertices and finds the number of vertices and their positions. The temperature decreases down to a value of $T_{\min}$ that is fine tuned to result in the correct identification of as many vertices as possible, while not incorrectly splitting up single physical vertices. After reaching $T_{\min}$, the temperature continues to be lowered to improve the assignment p_{ik} of tracks to vertices, but without further splitting of vertices. During this step a term $\exp(-\mu^2/T)$ is added to the inner summation of Equation 5.1 to reject outliers by cutting off the assignment probabilities of Equation 5.2. To further reduce the remaining outliers, vertices with less than two tracks not compatible with any other vertex are removed. Tracks connected to removed outliers are assigned to other vertices by carrying out the minimization of F from Equation 5.1 once again.

After the vertices have been identified and tracks have been assigned probabilities to originate from different vertices, vertex positions and covariance matrices are determined with an adaptive vertex fit [205]. Each track is assigned a weight during this fit to represent the likelihood that it originates from the vertex being fit.

The probability of having two interesting hard scattering interactions in a single bunch crossing is minuscule. So a single vertex, the leading primary interaction vertex, is chosen in analyses from which the objects of interest are supposed to originate. The leading primary vertex is chosen to be the one with the highest p_T^2 sum over the tracks and $\vec{p}_T^{\text{miss}}$ associated to this vertex [177]. In later chapters

primary vertex will refer to the leading primary vertex and not to any of the pileup primary vertices, unless otherwise specified.

5.1.1.6 Beam spot reconstruction

The beam spot is the profile of the luminous region of the detector where the LHC beams collide. It is an average over many events, and the center of the beam spot is used to anticipate the primary interaction region before vertex reconstruction, and to provide additional constraints during track and vertex reconstruction. Knowledge of the beam spot is also useful to estimate the primary interaction region in low multiplicity collision events in which vertices can not be reconstructed with high accuracy.

The center of the beam spot is determined in two ways. In one method a likelihood fit to the three dimensional vertex positions over many events is used to determine the shape and location of the beam spot. In the second method the transverse impact parameters of tracks are parametrized as a function of their longitudinal impact parameters and azimuthal angles. A correlation between the transverse impact parameter and ϕ appears if the beam spot is displaced relative to its expected position, and a fit of this correlation determines the transverse location of the beam spot [206]. The second method is more precise but only yields the transverse location of the beamspot. Therefore both methods are combined.

Aside from the aforementioned likelihood fit to the vertex positions, the transverse beamspot size can be determined in a second way. When an interaction is displaced with respect to the center of the beam spot, all particle trajectories from this interaction will have a nonzero impact parameter relative to the beam spot. The correlation in these shifts provides the second means of determining the transverse width of the beam spot, and the result is combined with that of the first method [177].

5.1.2 Calorimeter clustering

The calorimeters are essential in the reconstruction of electrically neutral particles, namely photons and neutral hadrons. Calorimeter information also improves the energy estimation of charged hadrons for which the track parameters are not accurately determined. This tends to be the case in the dense core of high p_T jets. The identification of electrons, with all their associated bremsstrahlung photons, also relies on calorimeter input.

The PF algorithm groups energy deposits in the calorimeters into clusters. Calorimeter cells are clustered separately in each subdetector: the ECAL, the preshower detector, and the barrel and two endcaps of the HCAL. The clustering procedure starts by identifying a number of seeds, cells with an energy deposit over a given threshold that is also larger than the deposits in neighboring cells. Subsequently, topological clusters are grown by aggregating adjacent cells into the cluster if they have energy deposits greater than double the expected noise level.

A single topological cluster can contain multiple seeds. Clusters corresponding to individual particles inside one topological cluster are reconstructed by postulating that N particles deposited energy with a Gaussian profile, where N is the number of seeds present in the topological cluster. The parameters of the Gaussians are their position in the (η, ϕ) plane and their amplitude. The width of each Gaussian is a fixed value that depends the specific calorimeter in which the topological cluster is found. An iterative algorithm first computes the fraction of the energy in each cell coming from each of the Gaussian deposit profiles, while the parameters of the Gaussians are frozen. In a second step a maximum likelihood fit determines the parameters of the Gaussians. At the first step the algorithm initializes the fit with the energies and positions of the seeds inside the cluster. Thereafter the above steps are iteratively applied until convergence. The final energy amplitudes and positions of the Gaussians are the estimates of the energies deposited by particles in the calorimeter and the direction from which they entered the calorimeter [182].

It is important that the energies of calorimeter clusters are well calibrated, in particular for the reconstruction and identification of neutral particles whose clusters can overlap with those of charged particles. The ECAL energy deposits are calibrated with simulated samples of photons so that the

expected calibrated energy matches the true photon energy. A different parametrization is used in barrel and endcaps, to account for the loss of energy to the preshower detector in the endcaps. The corrections tend to be larger for low energy deposits because they are more sensitive to the energy thresholds used in the clustering algorithm [182]. In a similar manner, HCAL clusters are calibrated with simulated samples of neutral hadrons. The true energy of neutral hadrons is fit with a sum of ECAL and HCAL energies with additional calibration functions preceding each term. A constant term is added to the fit to account for threshold effects. The ECAL calibration found for hadrons is substantially different from that of photons and electrons [182].

5.1.3 Information combination

When a particle traverses the CMS detector it is expected to give rise to PF elements in several subdetectors. A charged hadron might for instance have a track, and clusters in both ECAL and HCAL. A linking algorithm combines the PF elements from all subdetectors that are expected to originate from the same particle. For each element, the nearest neighbors in the (η, ϕ) plane are considered for linking [207]. For each link between elements a link distance is defined that quantifies the quality of the link. All elements that are linked together form a particle flow block.

The first step in the linking algorithm is to extrapolate tracks from their last hit in the tracking detector to the expected depth of the maximum of an electron shower in the ECAL, and to one interaction length into the HCAL. If the extrapolated track falls within the area of a calorimeter cluster, which is enlarged by one cell to account for several uncertainty sources, the track and cluster are linked. If several HCAL clusters are connected to the same track, or several tracks to the same ECAL cluster, only the links with the smallest distance are retained.

Bremsstrahlung deposits are linked to GSF electron tracks by extrapolating track tangents at each layer of the tracking detector to the ECAL. If an ECAL cluster overlaps with an extrapolated tangent, a link is created. Bremsstrahlung photons that converted to electron-positron pairs are identified by a conversion finder [208], and the corresponding tracks are linked to a GSF track if their reconstructed direction matches one of the tangents.

Links between calorimeter clusters are made by determining whether the cluster in the more granular detector (preshower or ECAL) falls within the (η, ϕ) surface of a cluster in the subsequent detector (ECAL or HCAL). Once again, if several overlapping links are present, only those with the smallest distance are kept.

Charged particle tracks can be linked together through a secondary vertex to reconstruct nuclear material interactions in the tracking detector. All tracks corresponding to such a secondary vertex are linked together.

After forming a particle flow block, the particles in the block are reconstructed. First muon candidates are identified and reconstructed, and the corresponding elements are removed from the block. Only isolated global muons are considered initially before moving on to other muon candidates. Muons inside jets are required to either be global and pass stringent identification criteria, or to have a high quality standalone fit [184].

The algorithm proceeds to identify and reconstruct electron candidates and isolated photons in the particle flow block. Electron candidates are seeded from GSF tracks, while photon candidates are seeded from large ECAL deposits. Linked HCAL energy deposits for photon and electron candidates are required to be small compared to those in the ECAL. All ECAL clusters with matching track tangents from bremsstrahlung, as well as conversion tracks, are added to the electron and photon candidates. The ECAL energy is corrected for potentially missing energy deposits. For reconstructed photons the direction and energy are those of the final aggregated ECAL supercluster. For electrons the direction is taken to be that of the GSF track, and the momentum is obtained from a combination of the ECAL energy measurement and the track momentum. All tracks and deposits from electron candidates and isolated photon candidates are removed from the PF block at this stage.

Subsequently, nonisolated photons, charged, and neutral hadrons are simultaneously identified. Any remaining ECAL and HCAL clusters that are not linked to a track are identified as neutral

particles. Inside the tracker coverage, precedence is given to photons for assignment of ECAL clusters because photons carry about three times more energy than neutral hadrons in a typical jet, and neutral hadrons deposit only a small fraction of their energy in the ECAL. Outside of the tracking detector coverage charged and neutral hadrons can not be distinguished so this reasoning no longer applies. In this region all ECAL clusters linked to a HCAL cluster are assumed to be from hadrons.

The momenta and directions of charged hadrons are taken directly from tracks linked to calorimeter deposits. To translate the measured momentum of a charged hadron into an energy, the mass of charged pions is assumed. The sum of the track momenta in a particle flow block is compared to the calibrated calorimeter energy. If there is leftover energy in the calorimeters, the ECAL energy is attributed to a photon, and if an energy discrepancy still remains, it is considered to be from a neutral hadron. On the other hand, if the sum of track momenta exceeds the total calorimeter energy, a search for muons is performed with relaxed requirements. Muons leave very little energy in the calorimeters, so one missed in an earlier reconstruction iteration might explain the excess.

The final part of the PF algorithm consists of post-processing steps that aim to eliminate artificial tails in the distribution of p_T^{miss} . Many new physics searches, such as that presented in Chapter 7 and to an even greater extent that of Chapter 9, aim to identify new physics by looking for a large p_T^{miss} or associated quantities. Consequently, it is of paramount importance that no spurious p_T^{miss} is introduced in events during reconstruction. Muons from cosmic rays are identified by finding those highly displaced from primary pp interaction vertices. Such muons are removed from the event if the p_T^{miss} lowers by at least a factor two with their removal. Muons from jet fragmentation that are displaced with respect to their primary vertex tend to have a direction correlated with that of p_T^{miss} . Such muons are not affected because their removal would increase rather than decrease the p_T^{miss} value. If wrong tracking detector and muon system tracks are associated with each other, the muon momentum assignment might be off significantly. If choosing only the momentum of either the tracking detector track or muon system track reduces the p_T^{miss} by a factor two, this momentum is assigned to the corresponding muon. Charged hadrons punching through to the muon system can also cause large p_T^{miss} if they are identified as muons. Muon reconstruction does not account for the calorimeter deposits of charged hadrons, and spurious neutral hadrons get added to the event. If a muon and a neutral hadron are aligned, have large momentum, and their replacement by a charged hadron halves the p_T^{miss} , the replacement is made.

5.2 Jet clustering

Quarks and gluons produced in the initial pp collision shower and hadronize, producing jets. To infer the properties of the particles produced in the hard scattering, PF candidates are clustered into jets. The jet clustering algorithm of choice in recent CMS analyses is the anti- k_T algorithm [209]. In this algorithm distance measures d_{ij} between any two PF candidates i and j , and d_{iB} between i and the beam spot are introduced. They are defined as:

$$d_{ij} = \min \left(p_{Ti}^{-2}, p_{Tj}^{-2} \right) \cdot \frac{\Delta_{ij}^2}{R}, \quad (5.3)$$

$$d_{iB} = p_{Ti}^{-2}, \quad (5.4)$$

with $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, and y_i and ϕ_j respectively the rapidity and azimuthal angle of PF candidate i . The parameter R is a constant set to 0.4 in the analyses covered in this dissertation. The clustering algorithm starts by finding the smallest d_{ij} or d_{iB} in the event. If d_{ij} has the smallest value, objects i and j are combined into a single object by summing their Lorentz vectors. If d_{iB} is smaller than any d_{ij} , object i is called a “jet” and masked during the remainder of the clustering procedure. The algorithm continues until all objects are part of jets, which might consist of a single

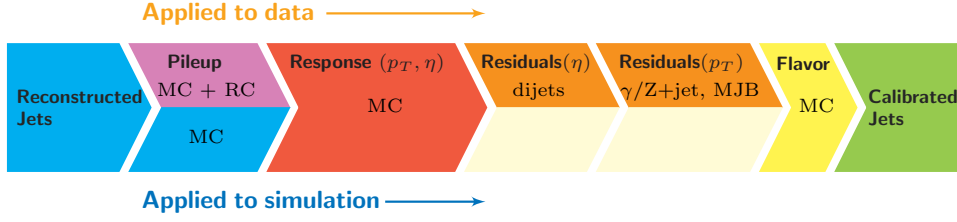


Figure 5.3: Jet energy correction stages in data and simulation. Corrections labeled with MC are based on simulated studies, while RC refers to a random cone measurement, and MJB stands for the multijet balance method. Flavor corrections are not applied to jets used in this dissertation [213].

particle¹.

The name anti- k_T refers to the negative power of p_T , sometimes k_T in the literature, in Equation 5.3. Because of the negative power the distance measure d_{ij} between two objects tends to be very small when a hard particle is involved, while it is large between soft particles. During clustering soft particles therefore tend to be associated to hard particles first before clustering among themselves. This results in an infrared and collinear safe algorithm with jet boundaries that can flexibly include hard radiation, while simultaneously being robust to soft radiation. The content of jets clustered with the anti- k_T algorithm is usually less affected by pileup than the content of jets formed by other algorithms [209].

In all work presented in this dissertation the implementation of the anti- k_T algorithm from the FASTJET package [210] is used. A naive implementation of anti- k_T would have a worst case asymptotic run time of $\mathcal{O}(N^3)$, with N the number of particles in the event². Identifying the smallest d_{ij} among all particles takes $\mathcal{O}(N^2)$ time and this has to be repeated N times. This is solved in FASTJET by observing that the geometrical and p_T -dependent aspects of the problem can be factorized, resulting in a much faster asymptotic run time of $\mathcal{O}(N \log N)$ [212].

5.3 Jet energy scale correction

Jets can contain large numbers of particles of different kinds, with small spatial separations and with overlapping detector signals. In addition, particles from multiple primary pp interaction vertices might be mixed inside a single jet, while we are typically only interested in a single primary vertex. A detailed understanding of the jet energy scale and resolution is crucial to many physics analyses.

The reconstructed momentum of a jet should ideally be that of the vector sum of the particles from the primary pp interaction clustered into it. To achieve this, several successive corrections are applied to jets in fixed order [213, 214]. First the contributions from pileup are subtracted. Secondly, the jet response is corrected based on simulation. Lastly, data events are corrected for residual differences with simulation in the jet response. Further corrections depending on the jet flavor are available, but not used in the analyses covered in this dissertation. The consecutive jet energy corrections and the data sets they leverage are summarized in Fig. 5.3.

5.3.1 Pileup correction

Reconstructed jets can contain contributions from pileup interactions in the current as well the previous bunch crossings. The former is called in-time pileup while the latter is referred to as out-of-time pileup and stems from the nonzero signal decay time in the calorimeters. The in-time pileup contribution from charged particles can be reduced by removing PF candidates that are unambiguously associated with pileup vertices from the jet clustering procedure. This procedure tends to remove about half

¹In terms of the jet clustering algorithm, any lepton or photon isolated from other activity in the event is a jet. In subsequent parts of the dissertation the term jet will however only refer to hadronic jets unless mentioned otherwise.

²A definition of big O notation can be found in [211].

of the total pileup momentum contained in a jet. The out-of-time pileup contribution is determined by simultaneously fitting the in-time and out-of-time pulse shapes to the calorimeter signals. The estimated out-of-time contribution is subtracted.

The remaining contributions from pileup are evaluated by studying the p_T difference in simulated jets reconstructed with and without an overlay of pileup interactions. The p_T difference is parametrized as a function of the uncorrected p_T , η , and area of the jet, as well as the diffuse pileup energy density ρ [215, 216]. A measurement of ρ is given by the median of the distribution of energies in a grid of (η, ϕ) cells in the event. For each jet the pileup contribution, predicted by the parametric equation fit to simulation, is subtracted.

The average pileup offset for jets in data can be measured using the “random cone” method in zero bias data. This is data taken by randomly sampling events without applying any trigger requirement. The random-cone method consists of randomly placing cones in the event, mapping the (η, ϕ) space of the detector. Each cone is taken to be a jet containing the particles in the cone. The average p_T of these jets is a measure of the average pileup contribution to reconstructed jets. This measurement is compared to simulated events to derive scale factors for the parameters in the fitted function used to predict the pileup contribution to jets. The random cone procedure assumes that the contribution from detector noise is negligible to that of pileup.

5.3.2 Response correction

In simulated events one has access to the particles produced by the physics generator, without any detector effects (Section 3.5). All generated stable particles, excluding neutrinos, are clustered with the same algorithm used on the reconstructed PF candidates, giving rise to “particle-level jets”. The particle response, defined as the ratio of mean reconstructed p_T to particle-level p_T of jets, is determined as a function of jet p_T and η in simulated multijet events. The inverse particle response is applied as a p_T correction to all reconstructed jets after they have been corrected for pileup [213].

After the above correction scheme the reconstructed jet p_T and the particle-level p_T agree well in simulated events. In data however, small residual corrections are needed to account for differences with simulation. The p_T balance in events is exploited to derive these corrections. If an event contains a jet for which the response differs significantly from unity, the event has a p_T imbalance. Two methods are used to study this. The p_T -balance method compares the p_T of the jet under consideration to that of a reference jet. The $\vec{p}_T^{\text{miss}}$ projection fraction method [217] considers all hadronic activity recoiling against a reference jet.

First a data sample of dijet events is used to determine the residual jet p_T response relative to barrel jets, which act as the reference object. The ratio of the response between simulation and data is evaluated in bins of η , parametrized with a logarithmic p_T dependence. This ratio is applied as a correction to data. These corrections remove the η dependence in the data to simulation response difference, but do not fix differences in the absolute jet p_T scale. The $\vec{p}_T^{\text{miss}}$ projection fraction method is used to derive the nominal correction because it is less dependent on the modeling of initial state radiation (ISR) and final state radiation (FSR).

The absolute jet p_T scale is determined by evaluating the p_T balance in events containing muons, electrons, or photons. These particles are measured with a relatively higher precision than jets. Data samples of $Z \rightarrow \mu\mu$, $Z \rightarrow ee$, and $\gamma + \text{jets}$ events are used. In $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$ events, the peak of the dilepton invariant mass distribution can be used to validate the lepton energy scale since the Z boson mass is known. An additional data sample of multijet events, in which a high p_T jet is balanced by a recoiling system of jets, is used to derive corrections at high p_T .

A global fit to the muon, electron, photon, and multijet data sets determines the final jet p_T scale. Two parameters are calculated in the fit: The absolute p_T scale, and a parameter that governs the p_T dependence of the response. The p_T dependence is assumed to come from a constant shift in the HCAL response. Such a shift leads to a p_T -dependent jet response as follows: In low p_T jets, PF reconstruction mainly relies on tracks for the momentum determination. In high p_T jets, on the other hand, pixel clusters begin to merge in the jet core, reducing the track reconstruction efficiency.

Additionally, some tracks become too straight for a reliable momentum determination. So higher jet p_T values lead to an increased reliance on the HCAL for the momentum measurement. The final fit uses both the p_T -balance and p_T^{miss} projection fraction methods because they have different biases and uncertainties, leading to a smaller combined error. After the fit all jets in data are corrected for the obtained response [213].

The effect of the different stages of the jet energy scale correction procedure is shown in Fig. 5.4. Before the corrections the particle-level jet response is large, particularly at low p_T , and highly variable as a function of pileup. After the whole correction procedure the response is close to unity across the range of jet p_T values and for different levels of pileup.

5.3.3 Missing transverse momentum

The jet energy corrections are propagated to p_T^{miss} to ensure it is estimated as reliably as possible [218, 219]. The corrected p_T^{miss} is:

$$\vec{p}_{T,\text{corrected}}^{\text{miss}} = \vec{p}_{T,\text{uncorrected}}^{\text{miss}} + \sum_{\text{jets}} \vec{p}_{T,\text{uncorrected}} - \sum_{\text{jets}} \vec{p}_{T,\text{corrected}} - \sum_{\text{jets}} \vec{O}^{\text{Random Cone}}, \quad (5.5)$$

where the vectorial difference between the sum of the uncorrected and fully corrected jets is added to p_T^{miss} . The average pileup contribution $\vec{O}^{\text{Random Cone}}$, determined in a random cone measurement, is kept under all jets. This ensures the pileup offset does not bias the p_T^{miss} measurement. The corrected p_T^{miss} from Equation 5.5 is often called type-I p_T^{miss} , and is what is used in all physics analyses presented in this dissertation. In the next chapters, whenever p_T^{miss} is mentioned, it is implicitly taken to refer to the type-I corrected p_T^{miss} .

5.4 Heavy flavor jet identification

Jets stemming from the hadronization of b or c quarks, respectively b jets and c jets, can be identified by exploiting the properties of the b or c hadrons inside the jet. The identification of b jets in particular provides an important handle for distinguishing signal and background events in the analyses presented in chapters 7 to 9. In this work algorithms for identifying b jets are also used indirectly, as a tool for distinguishing leptons from decays of electroweak bosons or new heavy particles from leptons that come from semileptonic b or c quark decays. This is the case in chapters 8 and 9. The identification of b jets is also called b-tagging, and jets passing a selected threshold of an identification algorithm are called b-tagged.

Because of the small CKM matrix coupling between the third quark generation and the others [29], b hadrons have a relatively long lifetime of about 1.5 ps. Depending on their momentum, such hadrons can travel up to 1 cm through the CMS detector before decaying. Because of the long hadron lifetime, the tracks of its decay products originate from a secondary vertex, displaced with respect to the primary vertex. This is illustrated in Fig. 5.5. The high mass of b quarks leads to a harder fragmentation compared to lighter quarks and massless gluons. In other words, the decay products tend to have a relatively high momentum perpendicular to the jet axis. About 20% of b hadron decay chains contain a muon or an electron, which can be used as an additional handle for identifying b jets [220].

Techniques for identifying b jets are based on simulated events. In order to design and optimize them, a reliable method for determining the true flavor of simulated jets is needed. Flavor assignment uses “ghost association” [215]. When reconstructed PF candidates are clustered into jets, the hadrons originally simulated by the physics generator, before accounting for detector effects, are added to the clustering. The magnitudes of the momenta for the generated hadrons are shrunk to become negligible. In this way they do not cause shifts in the reconstructed jet momentum. Jets containing at least one generated b hadron are considered b jets, those containing a c hadron and no b hadrons are considered c jets. All other jets are classified as light-flavor jets.

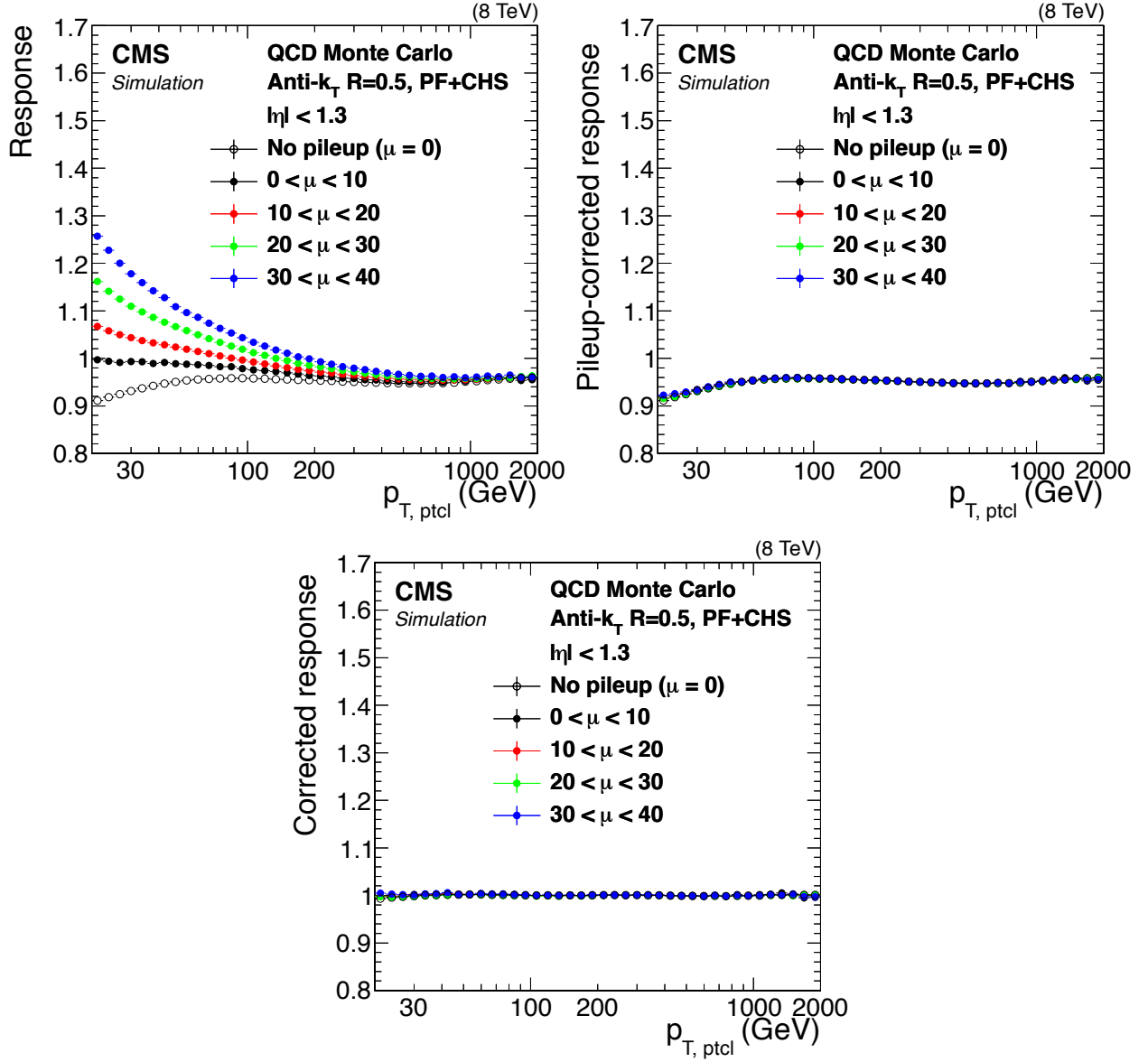


Figure 5.4: Average ratio of the measured jet p_T to the particle-level jet p_T as a function of the particle-level jet p_T , without any jet energy scale corrections (top left), after applying only pileup corrections (top right), and after the application of the full jet energy scale correction procedure (bottom). The distributions are shown for QCD multijet events in simulated 8 TeV pp collisions [213].

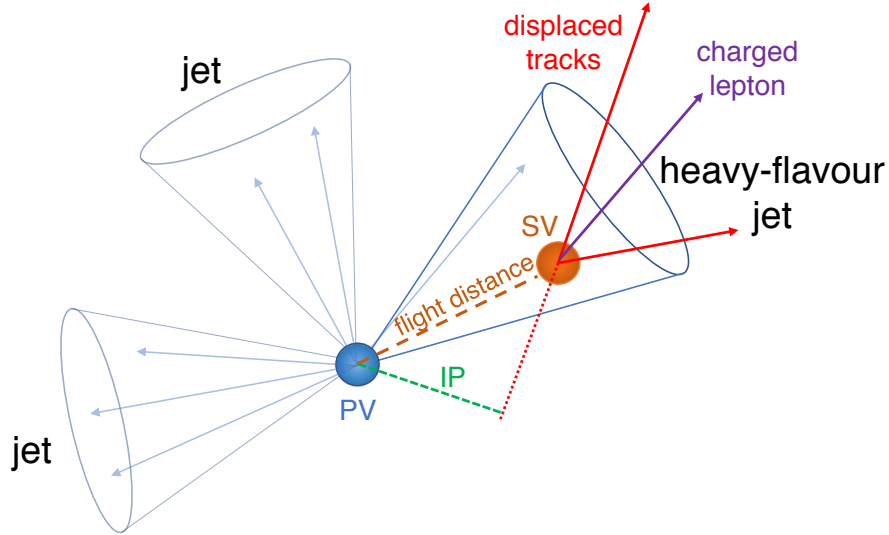


Figure 5.5: Illustration of a heavy flavor jet in which a b or c hadron travels a significant distance before decaying, resulting in a secondary vertex displaced with respect to the primary vertex. This secondary vertex tends to have a significant impact parameter and might contain a lepton [220].

Machine learning algorithms which learn to classify jets according to their assigned flavor are trained in simulated events. The main input variables to these algorithms are the properties of the tracks clustered into the jet, and quantities derived from reconstructed secondary vertices associated to the jets. To be used as a direct input to the flavor classification algorithms, tracks have to pass a set of requirements aimed at rejecting those coming from pileup, material interactions, and long-lived Λ and K_S^0 hadrons. Secondary vertex clustering for flavor classification in Run 2 analyses makes use of the inclusive vertex finding (IVF) algorithm [221] as opposed to the adaptive vertex reconstruction (AVR) [222] used for b -tagging during Run 1 [223]. The main difference between the two approaches is that IVF uses all reconstructed tracks, whereas AVR only uses tracks that pass the aforementioned selection criteria. The more inclusive approach of IVF leads to a more efficient reconstruction of secondary vertices [220].

The b -tagging algorithms used in CMS have evolved considerably during Run 2, reflecting rapid developments in the field of deep learning [224, 225]. The CSVv2 algorithm is an extension of the Run 1 CSV algorithm [223]. Events are categorized according to the presence of a secondary vertex and/or a set of tracks with a significant impact parameter. In each category two fully-connected neural networks with a single hidden layer are trained to discriminate b jets from c jets and light jets. The neural network outputs in the different event categories are combined according to a likelihood ratio based on the number of jets of each flavor in simulated $t\bar{t}$ events. The final discriminant is a linear combination of the outputs of the discriminants against c jets and light jets with relative weights of 1:3, reflecting the relative prevalence of each background source in $t\bar{t}$ events. The algorithm outperforms the old CSV algorithm because IVF is used to reconstruct the secondary vertex, and the number of input variables to the neural networks has been increased. The full list of input variables is given in [220]. The CSVv2 algorithm is used to veto b jets in Chapter 7. The newer algorithms explained below were not available at the time the analysis presented there was carried out.

The more recent DEEPCSV algorithm is a fully-connected neural network with four hidden layers that treats the flavor identification problem as a multiclass classification task [226]. It predicts the probability for a jet to belong to each of five classes: The jet contains a single b hadron, two or more b hadrons, a single c hadron and no b hadrons, two or more c hadrons and no b hadrons, or falls into none of these categories. The probability that a jet contains any b hadron is given by the sum of the output probabilities for the first two classes. The neural network is trained inclusively on a mixture of simulated $t\bar{t}$ and multijet events. It outperforms the CSVv2 algorithm, and is used for

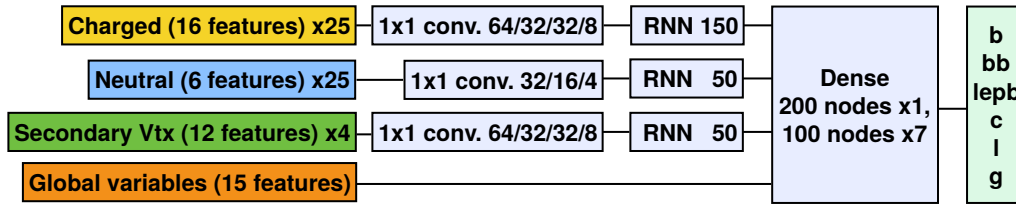


Figure 5.6: Schematic of the neural network architecture of the DEEPCSV algorithm [229]. Several abbreviations are used: Vtx stands for vertex, conv. for a one dimensional convolutional layer, and RNN for recurrent neural network layer, all of which are LSTM layers in this particular case.

identifying b jets in the analyses of Chapters 8 and 9. It is also used as one of the inputs to the lepton identification algorithm used in Chapter 8.

Jets are fundamentally variable-sized sets of particles. This structure lends itself to deep learning representations that are more statistically efficient at learning global properties from local structure than fully-connected neural networks are. This can be achieved by sharing weights across operations, a property satisfied by convolutional and recurrent neural network layers. Such layers are used as building blocks for the DEEPCSV algorithm, the latest b -tagging algorithm deployed in CMS at the time of writing. Collections of charged and neutral jet constituents, and secondary vertices, are separately processed by a chain of one dimensional convolutional neural network layers [227] followed by a long short-term memory (LSTM) [228] layer. These neural network layers learn to identify physical properties useful for solving the given task, without a need for human input feature engineering. The outputs of the three LSTM layers are combined with the human designed input features also used in the DEEPCSV algorithm. All of this is fed to a fully-connected neural network with six hidden layers to predict the probability a jet belongs to one of six classes. The classes are similar to those used in DEEPCSV, but an additional class containing jets with a b hadron decaying to a soft lepton is added, and the categories for one and multiple c hadrons are merged. Light quark and gluon jets are also considered to be separate classes in DEEPCSV. The final b -tagging output is given by the sum of all three output categories involving b hadrons. The architecture of the DEEPCSV algorithm is illustrated in Fig. 5.6.

The efficiency for correctly identifying b jets is typically around 5% higher for DEEPCSV than it is for DEEPCSV at a given misidentification rate for light and c jets [229, 230]. The performance of both algorithms in simulated $t\bar{t}$ events is shown in Fig. 5.7. The DEEPCSV algorithm was not yet supported when the analysis of Chapter 8 was performed, and would in all likelihood have improved the results achieved there. For the search discussed in Chapter 9 DEEPCSV was used because it has a smaller associated uncertainty in the misidentification rate. The DEEPCSV algorithm is however used as an input to the lepton identification algorithm used in Chapter 9.

All b jet identification algorithms introduced above are exclusively trained on simulated events where the true jet flavor is known. They have to be corrected for differences between data and simulation. Correction factors for the misidentification rates and b -tagging efficiencies of the different algorithms are measured in a combination of several data sets. The measurements are done as a function of jet p_T . The misidentification rate of light flavor jets is measured in multijet events by exploiting the fact that tracks in such jets have an almost symmetric impact parameter distribution. Tracks in b or c jets, on the other hand, have predominantly positive impact parameter values. A highly pure sample of light jet data can be obtained, even at high b -tag scores, by using only tracks with a negative impact parameter. This data sample can be used to measure the misidentification rate for various working points of the different b -tagging algorithms [220, 223].

The misidentification rate for c jets is extracted from a combined measurement in data samples enriched in $W + c$ production and $t\bar{t}$ events with a single leptonic t quark decay. The $W + c$ sample is obtained by noting that the W boson and c quark tend to have an opposite charge, while in the dominant $W + q\bar{q}$ background there are an equivalent amount of events where the W boson and the

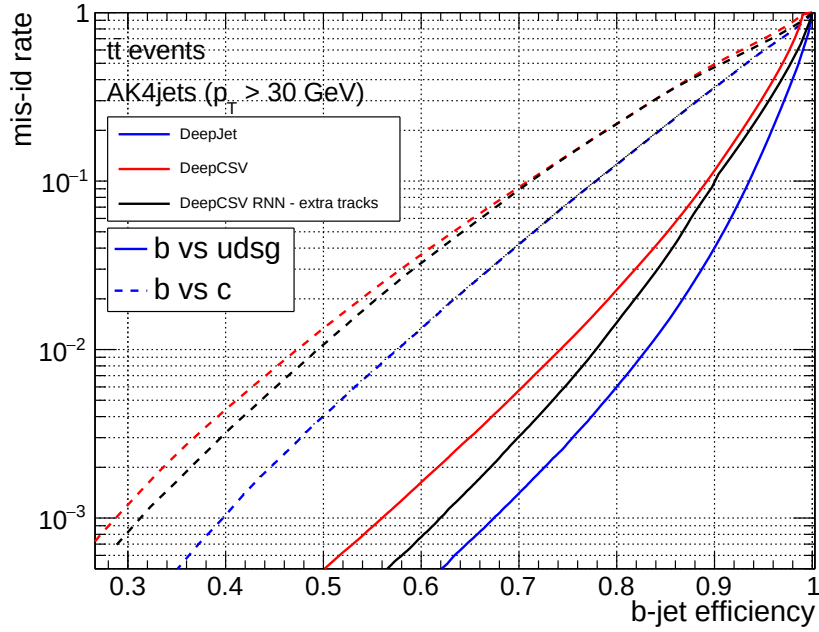


Figure 5.7: Misidentification rates of different jet types as b jets versus the efficiency of identifying genuine b jets. The values are drawn for jets with p_T values greater than 30 GeV in simulated $t\bar{t}$ events [229]. The performance curves of the DEEPCSV and DEEPJET algorithms are drawn in red and blue respectively. The black line shows a refined version of the DEEPCSV algorithm that is not used in this dissertation.

$q\bar{q}$ pair have opposite and equal charges. Subtracting the same charge event yields from the opposite charge event yields results in a relatively pure sample of $W + c$ production [231]. In single lepton $t\bar{t}$ events there is an hadronically decaying W boson, which decays to a c quark in 50% of cases. This sample can be further enriched, or depleted, in c quarks by using the fact that in $t\bar{t}$ events down-type jets from a W boson decay tend to be less boosted than up-type jets from such a decay [220].

The efficiency to identify genuine b jets with the b-tagging algorithms is evaluated in multijet events in which at least one of the jets contains a muon. A muon inside a jet is a telltale sign of a b hadron decay. A multitude of complimentary methods, explained in [220], are used for performing the efficiency measurement in this data set. The efficiencies are also measured using an arsenal of varying methods in both single lepton and dilepton $t\bar{t}$ events, where two b jets should be present [220]. The results of all of these measurements are combined to obtain the final data to simulation scale factors for the b-tagging efficiency.

The measured correction factors for the misidentification rates tend to be larger than unity while they are typically smaller than unity for the efficiency of tagging b jets. This reflects the fact that the b-tagging algorithms are machine learning algorithms trained in simulated events. Discrepancies between what they learned in simulation and the conditions in real data lead to a degradation in performance [220]. The scale factors tend to deviate further from unity for more advanced algorithms, with DEEPJET having larger corrections and associated uncertainties than the other algorithms [230]. This is the reason DEEPJET was not chosen over DEEPCSV for the vetoing of b jets in Chapter 9.

5.5 Reconstruction and identification of hadronic τ lepton decays

Unlike electrons and muons, τ leptons, with their mass of 1.7 GeV, are heavy enough to decay to hadrons. About two thirds of τ leptons decay to hadrons and the remaining third to muons or electrons. All decay modes include a ν_τ in the final state. An overview of the τ lepton decay modes

Decay mode	Resonance	Branching fraction (%)
Leptonic decays		35.2 total
$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$		17.4
Hadronic decays		64.8 total
$\tau^- \rightarrow h^- \nu_\tau$		11.5
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	$\rho(770)$	25.9
$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5
$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other hadronic decays		3.3

Table 5.1: Overview of τ lepton decay modes with their branching fractions. The branching fractions of τ antileptons are almost identical, but with inverted charges. The branching fractions are taken from [29], and the table is modified from [232].

with their branching fractions is given in Table 5.1. In particular models for new physics, and in some SM processes such as H boson decays, the expected final state predominantly includes τ leptons instead of other lepton flavors. This is the case in several of the signal hypotheses studied in Chapter 9. To increase the sensitivity to such processes, hadronically decaying τ leptons (τ_h) are reconstructed. Because of the relatively short τ lepton lifetime, leptonic τ decays are almost indistinguishable from muon or electron production without an intermediate τ lepton. For this reason these are simply considered to be muons or electrons, and no further attempt is made to infer whether they come from a τ lepton decay [232].

The reconstruction of τ_h candidates at CMS starts from jets and is performed with the hadron-plus-strips algorithm introduced during Run 1 [233, 234], and subsequently improved for Run 2 [232]. Many of the common decay modes of τ leptons include one or more π^0 mesons. A π^0 decays to two photons which can convert into electron-positron pairs by interacting with the tracker material. These electron-positron pairs tend to have a very similar η , but are spread out in ϕ by the magnetic field of CMS. This gives rise to strips in (η, ϕ) with a long axis in the ϕ direction.

Given a jet which is checked to be a τ_h candidate, the reconstruction of π^0 candidates as strips proceeds as follows: The electron or photon (e/γ) of highest p_T in the jet that has not yet been clustered into a strip is taken as a new strip candidate. The second-highest p_T e/γ in the jet is merged into the strip candidate if it falls within a certain $\Delta\eta$ and $\Delta\phi$ from the current strip candidate. The $\Delta\eta$ and $\Delta\phi$ limits are functions of the p_T of the current strip and the p_T of the e/γ under consideration. These functions are determined in simulated single τ lepton events so that 95% of electrons and photons from τ lepton decays would be contained in reconstructed strips. If a new e/γ is added to the strip, its position is recomputed as a p_T -weighted average of the positions of all its e/γ constituents. The procedure is repeated until there are no further e/γ candidates that satisfy the p_T -dependent proximity criteria for addition to the strip. Once the strip is completed, the remaining e/γ of highest p_T is used as a starting point for the reconstruction of the next strip.

Aside from π^0 mesons, the τ_h decay modes include one or more charged hadrons in the final state. If the total charge of the hadrons inside a jet is not ± 1 , the jet is not considered as a possible τ_h candidate. If the jet constituents are compatible with a $\tau \rightarrow h\pi^0\nu_\tau$ decay, the mass of the visible decay products (m_{vis}) must match the expectation for a ρ resonance. In case of $\tau \rightarrow h\pi^0\pi^0$ or $\tau \rightarrow hhh\nu_\tau$ decays m_{vis} must be compatible with the expected mass of an a_1 resonance. The $\tau \rightarrow hhh\pi^0\nu_\tau$ decay mode is not considered. This particular decay is too hard to distinguish from hadronic jets, because of its higher particle multiplicity with respect to the other decay modes. Exact details on the mass constraints used in practice can be found in [232].

Reconstructed τ_h candidates suffer from large misidentification rates by QCD jets. This background is reduced with a gradient boosted forest (BDT) classifier that aims to separate genuine τ_h candidates

from misidentified jets. The classifier is trained using genuine τ_h candidates from simulated $Z/\gamma^{(*)} \rightarrow \tau\tau$ decays, as well as from hypothetical W' , Z' , and heavy supersymmetric H boson decays to τ leptons. Background τ_h candidates are taken from a mix of simulated QCD multijet, W + jets, and $t\bar{t}$ events. The most important input variables to the discriminant are the scalar p_T sum of charged hadrons and photons in a cone around the τ_h candidate, and the pileup correction term to this sum. Other important input variables are the two and three dimensional impact parameters of the leading track in the τ_h candidate and their significance, the τ_h candidate flight length and its significance, and the electron and photon multiplicity in the τ_h candidate. A detailed description of all the input variables is given in [234] and [232]. In Chapter 9 we use the tight working point of the BDT classifier. This working point has an efficiency of about 50% to identify genuine τ_h candidates, with a misidentification rate of 0.2% for QCD jets [232].

Muons can be misidentified as single-prong hadronic τ lepton decays because of the similarity between the detector signals induced by muons and charged hadrons. The rate of misidentification of muons as τ_h candidates can be reduced by imposing requirements on the activity in the muon detector along the direction of the τ_h candidate [232].

Electrons often emit bremsstrahlung photons which can mimic π^0 mesons, one of the τ lepton decay products. This results in a significant misidentification rate of electrons as τ_h candidates. A BDT discriminant is trained to distinguish genuine τ_h candidates from misidentified electrons. The inputs to this discriminant are several properties of the photons clustered into the τ_h candidate [232].

The efficiency for τ_h candidates to pass all aforementioned reconstruction and identification criteria is measured in a number of different data samples. For τ_h candidates with transverse momenta below 60 GeV, $\mu\tau_h$ events enriched in $Z \rightarrow \tau\tau$ decays are used. Efficiencies for τ_h candidates with intermediate p_T values, up to 100 GeV, are measured in $t\bar{t}$ enriched $\mu\tau_h$ events. Events with a single τ_h candidate from high-mass virtual W boson decays are used for efficiency measurements in the high p_T regime. The resulting uncertainties range between 1 to 3% in the selection efficiency of τ_h candidates, and depend on the p_T and η of the candidate, and the selection working point.

Chapter 6

Multilepton final state

*“The ash is of all trees the biggest and best.
Its branches spread out over all the world
and extend across the sky. Three of the
tree’s roots support it and extend very, very
far.”*

Snorri Sturluson, *Edda*

All analyses performed for this dissertation are based on events with multiple leptons. Most of the focus is on light leptons, electrons and muons. Compared to other objects light leptons are easily identified and their properties can be measured with superb resolution. The leptonic decay channels of heavy particles are therefore often the ones that provide the most precise measurements, or have the most potential for discovery. Prominent examples of the power of the leptonic final state are the discovery of the H boson in its decay to four leptons [14, 15], and the discoveries of the W [50, 51] and Z [52, 53] bosons.

Events with three or more leptons are uncommon in the SM because processes that can result in this final state have relatively small production cross sections compounded by small branching fractions to multiple leptons. If an undiscovered process has a significant branching fraction to multilepton events, this final state provides a golden channel for the potential discovery of the process. This point will be clearly made in Chapter 8 with the first observation of single top quark production in association with a Z boson. One disadvantage of multilepton events is that they often involve one or more leptonic W boson decays. In such a decay a neutrino is produced that escapes detection. This typically makes it impossible to completely reconstruct the masses of heavy particles that were produced in the event.

This dissertation discusses three different searches using pp collision events resulting in multiple leptons. Chapter 7 presents a search for hypothetical sterile neutrinos produced in leptonic W boson decays. The sterile neutrino can decay into two more leptons and a neutrino, leading to events with three leptons and p_T^{miss} . Chapter 8 describes a search for SM production of a single top quark in association with a Z boson. Leptonic decays of both the top quark and the Z boson result in events with three leptons, accompanied by several jets. The final analysis of this dissertation unfolds in Chapter 9 and constitutes a search for direct electroweak production of supersymmetric chargino-neutralino pairs. The decays of these supersymmetric particles lead to a rich variety of final states, many of them with multiple leptons.

The primary focus in this dissertation is on events containing electrons and muons. Due to their short lifetime τ leptons decay inside the detector and only their decay products are detected (Section 5.5). Leptonic τ decays are experimentally quasi indistinguishable from other electrons or muons, and are thus also implicitly considered whenever electrons or muons are reconstructed. Hadronic τ decay reconstruction tends to suffer from much larger backgrounds, lower reconstruction and identification efficiencies, and worse measurement resolutions compared to electrons and muons. For this reason the analyses of chapters 7 and 8 only consider events with light leptons. Depending on the sig-

nal model, the supersymmetric particles studied in Chapter 9 might or might not decay predominantly to τ leptons. To cover all possible scenarios, final states containing one or more τ_h candidates are considered, but the bulk of my work concentrated on events with three light leptons. The remainder of this chapter will therefore predominantly focus on events with three light leptons. In the rest of this chapter “leptons” will refer to electrons and muons, unless otherwise specified.

This chapter gives an overview of elements and techniques common to the three analyses discussed in the subsequent chapters. First an overview of light lepton identification techniques is given in Section 6.1, after which the triggers used for online event selection during data taking are introduced in Section 6.2. The most important SM processes in the multilepton final state and how to reduce and predict their yields are examined in Section 6.3. Details on the data-driven prediction of the background from nonprompt leptons are laid out in Section 6.4. Finally, Section 6.5 reviews the common systematic uncertainties affecting all the analyses presented later in this work.

This chapter can be read directly, or used as reference material when reading the chapters covering the concrete analyses done for this dissertation. When reading the following chapters, one could jump back to this one to find explanations on many of the analysis and background estimation techniques. Later chapters contain the necessary cross-references to facilitate such a reading approach.

6.1 Light lepton identification

Electrons and muons are reconstructed with the PF algorithm along with other particles in the event (Section 5.1). Reconstructed electrons and muons originate from a variety of sources. In this work, only a subset of the reconstructed leptons is of interest. The target signals of each analysis performed in this dissertation contain several leptons, all directly stemming from the decay of electroweak bosons or heavy hypothetical particles (sterile neutrinos in Chapter 7 and supersymmetric particles in Chapter 9). In each of these cases the parent particle of the leptons has a terrifically short lifetime, so the leptons hail from the primary pp interaction vertex. Referring to this property, such leptons are collectively labeled “prompt”¹. Light leptons from the decay of a prompt τ lepton are also considered prompt.

Semileptonic decays of b and c quarks result in genuine leptons. The lifetime of the decaying quarks is significant and the resulting leptons originate from a point displaced with respect to the primary pp interaction, possibly a secondary vertex. Hadrons or photon conversions inside jets can also be incorrectly reconstructed as leptons. We group these types of reconstructed leptons, whether they are genuine or fake, under the umbrella “nonprompt”.

The production rate of jets in the LHC is vast compared to that of electroweak bosons, and even more so compared to what is expected for hypothetical new particles. Nonprompt leptons are therefore far more common in LHC collisions than prompt leptons. Effectively distinguishing prompt from nonprompt leptons is of crucial importance for all the results included in this work. Without adequate methods to make this distinction the signals we target drown in an overwhelming background from nonprompt leptons. Chapter 8 provides an unambiguous example of the importance of good lepton identification techniques: The first observation of a previously undiscovered process is made possible by improvements to the lepton identification strategy.

Standard lepton identification techniques, used in many CMS analyses, are explained in Section 6.1.1. More advanced machine learning classifiers, to which the work in this dissertation made significant contributions, are discussed in Section 6.1.2. Techniques for measuring the efficiency of identifying prompt leptons in pp collision data are the topic of Section 6.1.3.

A technical note on photon conversions is necessary before continuing. Leptons from internal and external conversions of prompt photons, i.e. photons directly originating from a hard scattering interaction, are considered prompt during the estimation of the expected event yields. Internal conversions,

¹If a sterile neutrino, subject of Chapter 7, is sufficiently light, its decay products might have a significant lifetime. Leptons from such a decay can no longer be considered prompt. More details on this can be found in the corresponding chapter.

meaning $\gamma^* \rightarrow \ell\ell$ decays, produce leptons that are on an individual basis indistinguishable from those coming from the decay of W, Z, or H bosons. External conversions from interactions of a photon with detector material almost exclusively result in electrons. Such electrons do not share most of the features typical of nonprompt leptons, and the background they induce can not reliably be estimated with the same data-driven techniques used for the prediction of the nonprompt lepton background (Section 6.4). For this reason such electrons are also considered prompt when discussing background events. It must be noted though that techniques that try to select prompt leptons generally also aim to reduce the contribution from photon conversions. The fewer external conversions entering the lepton selection, the better. In most cases specific criteria are applied in the lepton selection to reduce this background source. Note that any lepton that emanates from the conversion of a nonprompt photon is also considered nonprompt.

6.1.1 Standard lepton identification techniques

The rate of particles being incorrectly reconstructed as leptons can be reduced by imposing quality criteria on the reconstructed PF objects. An overview of the corresponding techniques for muons (electrons) is given in Section 6.1.1.1 (6.1.1.2). Genuine leptons that are nonprompt can not be distinguished from prompt leptons by looking at the quality of the lepton reconstruction. Instead, other techniques based on the physical characteristics of nonprompt leptons are used. An overview of these is given in Section 6.1.1.3.

6.1.1.1 Muon identification

An important distinction between real and fake muons is a good match between the tracks in the tracking detector and the muon system. A muon PF candidate passes the “loose muon identification”² defined in [184] if it is either a tracker muon or a global muon (Section 5.1.1.4).

For the more stringent “medium muon identification” requirements, the concept of segment compatibility is introduced. The tracking detector track of a muon candidate is extrapolated to the muon system, and the number of matched track segments (Section 5.1.1.4), and their proximity in position and direction to the extrapolated track, are evaluated. A score between 0 and 1, the segment compatibility, is assigned with 1 indicating a perfect compatibility between the tracking detector track and the segments in the muon system [235]. To be considered medium a muon must pass the loose identification criteria and have a tracking detector track that contains hits in at least 80% of the layers it traversed. Additionally, the muon must either pass a stringent requirement on the segment compatibility, or be a global muon that passes a looser segment compatibility criterion and a number of additional conditions: The global track fit must have a low χ^2 goodness-of-fit, the muon tracks in the tracking detector and muon system must have a matching position, and the tracking detector track must pass a threshold on a kink-finding algorithm. At several points along the track, the kink-finding algorithm attempts to split the track into two separate tracks and then compares them to see if they are compatible. The segment compatibility cuts in the medium muon identification are designed to result in an efficiency of 99.5% for selecting muons from simulated W and Z boson decays [184].

Both the loose and medium muon requirements aim at selecting genuine muons, prompt as well as nonprompt muons from the decay of b or c quarks. The requirements reject most charged hadrons that managed to penetrate into the muon system and are falsely reconstructed as muons. The nominal muon selection used in all analyses presented in this work requires muons to pass the medium muon requirements. A considerable background from genuine nonprompt muons remains after the application of this selection. A more stringent “tight muon identification”, as well as other criteria for low- and high- p_T muons are introduced in [184], but not used in this work.

²Not to be confused with the term loose selection introduced in Section 6.4, which refers to a set of looser lepton requirements compared to those nominally used in an analysis. These loose selection requirements are used to facilitate the data-driven prediction of the background from nonprompt leptons. Whenever a lepton is said to pass a loose selection, we refer to this definition, not the loose muon identification requirement introduced in this section.

6.1.1.2 Electron identification

To reject electron candidates coming from misidentified jet constituents and photon conversions, a number of different variables are used. An overview can be found in [203], and a brief review is given here.

Purely calorimetric observables try to separate genuine electrons from jets or jet constituents with large electromagnetic components. Compared to showers with an hadronic component purely electromagnetic showers are narrower in the ECAL and the fraction of energy they deposit in the HCAL is expected to be much lower.

The properties of the track of an electron candidate can also give important clues about its provenance. Electrons typically lose a significant portion of their energy through bremsstrahlung. An example of track based variable used in electron identification is the difference in momentum of a track at its origin, and when exiting the tracking detector. Large momentum differences might indicate the track is that of a genuine electron. Track related variables are also useful for reducing the background from photon conversions. Photons convert to electron-positron pairs by interacting with the detector material. Conversion tracks therefore tend not to have hits in the innermost pixel detector layers.

Variables that use a combination of a calorimeter deposit and a tracking detector track can also hint at the compatibility of a PF candidate with the electron hypothesis. An example of a commonly used variable of this kind is the difference between the inverse of the energy the candidate deposited in the ECAL and the inverse of the momentum of its track at its origin vertex.

A number of discriminating variables such as those described above are combined into BDT discriminants trained to discriminate prompt from nonprompt electrons in simulated Drell-Yan + jets events. Separate discriminants are trained in different bins of p_T and η ³. Genuine nonprompt electrons from quark decays are relatively rare in Drell-Yan events, so the trained discriminants mainly focus on rejecting jet constituents that are misidentified as electrons. We will refer to this set of discriminants as “electron MVA” in the rest of the dissertation, and always assume the discriminant for the p_T and η bin in which a given electron falls is the one evaluated. Versions of the electron MVA with and without the isolation variables introduced in Section 6.1.1.3 as inputs are available. We always use the version agnostic about isolation since we either apply isolation requirements separately (Chapter 7), or use the electron MVA in conjunction with isolation and other variables to build a higher level discriminant (chapters 8 and 9).

6.1.1.3 General light lepton identification

Aside from quantities specific to electrons and muons introduced in the previous sections, a number of variables used in the identification of prompt leptons are shared between the light lepton flavors. Prompt leptons are expected to come from the primary pp interaction vertex and should have very small impact parameters with respect to this vertex. Genuine nonprompt leptons, on the other hand, often have a significant impact parameter value because of the lifetime of their parent quark. The following three impact parameter variables are used in this work:

- d_{xy} : Impact parameter of the lepton in the (x, y) plane, transverse to the beam direction.

³This is a strategy that almost certainly results in a nonoptimal performance of the electron identification algorithm. It is clear that varying p_T and η will significantly affect the properties of an electron track and ECAL cluster. Using a small number of p_T and η bins will not capture this complexity, and in addition reduce the statistical power of the training data sets seen by the machine learning algorithms since each binned training is agnostic about the others. A better approach to achieve a discriminant which does not depend on the p_T and η spectrum of the particular processes used in the training data set, would be to use p_T and η as input variables during training, but reweight either the signal or background distribution so both are identical. The resulting discriminant could then fully exploit the p_T and η -dependent track and ECAL cluster properties of prompt and nonprompt electrons, without preferring any process-specific kinematic spectrum. This is one of several reasons why the BDT based lepton selection introduced in Section 6.1.2 has a performance vastly superior to that of the electron MVA.

- d_z : Longitudinal impact parameter of the lepton, parallel to the beam direction.
- $d_{3D}/\sigma(d_{3D})$: Significance of the three dimensional impact parameter, defined as the three dimensional impact parameter d_{3D} divided by the uncertainty in its measurement.

In the three analyses of this dissertation we require $d_{xy} < 0.05$ cm and $d_z < 0.1$ cm for all electrons and muons. These thresholds are relatively loose and cut out obvious nonprompt leptons, while being close to 100% efficient for prompt leptons. In Chapter 7 a relatively tight cut $d_{3D}/\sigma(d_{3D}) < 4$ is used, which is efficient for prompt leptons coming directly from the decay of an electroweak boson or a heavy hypothetical particle and rejects many nonprompt leptons. Some leptons from prompt τ decays are lost because of this tight cut. The analyses in chapters 8 and 9 use a looser cut $d_{3D}/\sigma(d_{3D}) < 8$, which rejects nearly no prompt leptons. The variable is subsequently used as an input variable to a second lepton selection stage based on a machine learning classifier.

Whether they are genuine or not, nonprompt leptons originate from jets. This means they tend to be surrounded by a host of other activity in the detector. Prompt leptons, on the other hand, tend to have little or no other activity in their vicinity. This makes the isolation of leptons a powerful tool for identifying prompt leptons. The isolation of a lepton is defined as the scalar p_T sum of all PF candidates in a cone around the lepton direction:

$$I_{\text{abs}} = \sum_{\text{charged}} p_T^{\text{charged}} + \max \left[0, \sum_{\text{neutral hadrons}} p_T^{\text{neutral hadron}} + \sum_{\gamma} p_T^{\gamma} - p_T^{\text{neutral pileup}} \right], \quad (6.1)$$

where “charged” refers to all charged particles and γ to all photons, and the summations run over all objects inside a given cone. Only charged particles coming from the primary pp interaction vertex are considered in the summation. The typical cone size used for this variable is $\Delta R < 0.3$ for electrons and $\Delta R < 0.4$ for muons. We deviate from the CMS norm in Chapter 7 and use a cone size of 0.3 for both muons and electrons. The term $p_T^{\text{neutral pileup}}$ is an estimate of the total pileup p_T coming from neutral particles. There are two ways in which this term is typically estimated in CMS analyses. One method derives effective areas in simulated events, which when multiplied with the offset energy density ρ in the event (Section 5.3) correctly estimate the neutral pileup contribution in the isolation cone on average. The other method, commonly known as $\Delta\beta$, takes the neutral pileup p_T sum in the isolation cone as half of the summed p_T of charged pileup particles. This is motivated by the fact that two thirds of the momentum of a typical jet is carried by charged hadrons and the remaining third by photons and neutral hadrons [216]. Conventionally, effective area corrections are used in electron identification and $\Delta\beta$ corrections in muon identification. In this work we however consistently use effective area corrections for both electrons and muons because the resulting identification performance is found to be marginally better.

Prompt leptons with a higher p_T tend to have more activity in a fixed cone around them because the other particles in the event typically have a significant transverse boost. High- p_T electrons also radiate more bremsstrahlung than soft electrons. In practice relative isolation, simply defined as the isolation divided by the lepton p_T :

$$I_{\text{rel}} = \frac{I_{\text{abs}}}{p_{T,\text{lepton}}}, \quad (6.2)$$

is therefore the identification variable of choice. Most CMS analyses that aim to select prompt leptons apply a requirement on I_{rel} with a fixed isolation cone size. This approach is taken in Chapter 7. In the newer analyses of chapters 8 and 9 no direct cuts on I_{rel} with a fixed cone size are used, but it serves as an input variable to the final lepton selection discriminant explained in Section 6.1.2.

Relative isolation with a p_T -dependent cone size has more recently found use. The variable $I_{\text{rel}}^{\text{mini}}$ is defined as the relative isolation within a cone of size:

$$\Delta R(p_T^{\text{lepton}}) = \frac{10 \text{ GeV}}{\min \left[\max \left(p_T^{\text{lepton}}, 50 \text{ GeV} \right), 200 \text{ GeV} \right]} \quad (6.3)$$

The cone shrinks with rising lepton p_T to take into account the increased collimation of hadron decay products under transverse boosts. In this way $I_{\text{rel}}^{\text{mini}}$ manages to contain the other jet decay products inside the isolation cone of nonprompt leptons, while being very efficient for prompt leptons because the probability for accidental activity inside the small isolation cone is low [236, 237]. Chapters 7 and 8 use a very loose cut $I_{\text{rel}}^{\text{mini}} < 0.4$ in their baseline lepton selection. This cut has an efficiency of over 99% for prompt leptons from W and Z boson decays. In addition $I_{\text{rel}}^{\text{mini}}$ is also used here as one of the inputs to the machine learning algorithms used for the final lepton selection decision.

6.1.2 Machine learning based lepton identification

The standard approach to lepton selection in CMS is to place a number of independent cuts on several of the variables described in Section 6.1.1. Large improvements to this strategy can be made by exploiting the properties of the jet containing the lepton, and by combining all variables that distinguish prompt from nonprompt leptons in a machine learning algorithm.

The anti- k_T jet clustering algorithm (Section 5.2) assigns every lepton to a jet. If a lepton is completely isolated from other activity the jet will only contain the lepton itself. Several properties of this jet are found to be useful for identifying prompt leptons:

- p_T^{ratio} : Ratio of the p_T of a lepton to the p_T of the jet that contains this lepton. The jet energy corrections (Section 5.3) are removed from the contribution of the lepton to the jet momentum vector when computing this variable to make the lepton p_T treatment consistent. This variable tends to be large for prompt leptons which are either alone in the jet or form the dominant contribution. Nonprompt leptons usually have a lot of activity around them, resulting in a small p_T^{ratio} value. The idea behind this variable is similar to that of I_{rel} , but p_T^{ratio} is more robust because a dynamic cone size is implicitly used for computing the isolation. The variable-sized jet returned by the clustering algorithm is used instead of a fixed cone. Taking into account jet energy corrections also improves the performance of the variable for lepton selection compared to standard I_{rel} .
- p_T^{rel} : The momentum of the jet after removal of the lepton, perpendicular to the lepton direction. This variable is typically large when a jet accidentally overlaps with a lepton, whereas it is expected to be small for leptons originating from an hadronic decay inside the jet because such a lepton tends to be aligned with the rest of the jet. This variable can help to recover prompt leptons that are not isolated due to their accidental proximity to other particles.
- Probability that the jet holding the lepton contains one or more b hadrons, as predicted by a b-tagging algorithm (Section 5.4). Nonprompt leptons from a semileptonic b quark decay, tend to be associated to jets with a high b-tagging score. In Chapter 8 the DEEPCSV algorithm is used (DEEPCSV_{closest}), while in Chapter 9 the more advanced DEEPJET algorithm (DEEPJET_{closest}) is used.
- Track multiplicity of the jet containing the lepton. Prompt leptons tend to have fewer tracks inside their containing jet than nonprompt leptons. This variable is of relatively minor importance compared to those introduced above, but can nonetheless result in small improvements in the final lepton selection.

Machine learning discriminants that combine these jet related variables, together with many of the lepton identification variables introduced in earlier sections previously were first used for the search for $t\bar{t}H$ in the multilepton final state [238]. In this search the minute $t\bar{t}H$ signal had to be separated from a

large nonprompt lepton background. Standard lepton identification methods proved insufficient, and BDT classifiers were trained to identify prompt leptons. These classifiers combine the following input variables: p_T , η , d_{xy} , d_z , $d_{3D}/\sigma(d_{3D})$, the neutral and charged parts of $I_{\text{rel}}^{\text{mini}}$ separately, p_T^{rel} , p_T^{ratio} , the CSVv2 value and the track multiplicity of jet containing the lepton. The addition of p_T and η to the list of input variables ensures that the BDT classifiers can learn to optimally distinguish prompt from nonprompt leptons at every p_T and η value. Largely because of these BDT classifiers, the multilepton channel proved to be the most important one for the observation of the $t\bar{t}H$ process [239].

In Chapter 8 an improved version of the BDT based lepton selection is introduced which I specifically designed for being able to observe the tZq process. The CSVv2 of the containing jet is replaced by the newer DEEPCSV discriminant and I_{rel} with a fixed cone size of 0.4 is added to the list of input variables. In addition, the size of the training data sets is increased by a large margin. This results in significant improvements in the performance of the classifiers. A more careful definition of the training data set for electrons proved of particular importance and results in a large efficiency increase of up to 20% per selected electron compared to the classifiers trained for the $t\bar{t}H$ observation. The performance of the classifiers is shown in Fig. 8.3 in comparison to a standard I_{rel} based lepton identification. A vast performance increase is seen, with a reduction in the background rate by a factor 8 (2) for muons (electrons) while being more efficient at selecting prompt leptons by 8 (12)%. More details on the training and design of this discriminant are given in Chapter 8.

This new lepton identification algorithm proved to be the crucial ingredient in observing the tZq process, and vastly reduced the uncertainties in the analysis compared to any earlier work. Because of its potency in the tZq search, the lepton identification developed for the work in this dissertation was also used in the CMS search for $H \rightarrow \mu\mu$ in the full Run 2 data set [240]. The power of the lepton selection algorithm at rejecting nonprompt leptons was one among many factors contributing to this analysis seeing the first evidence of this rare H boson decay mode.

The last and newest search in this dissertation, presented in Chapter 9, uses a version of the $t\bar{t}H$ lepton classifiers updated for the $t\bar{t}H$ measurement in the full Run 2 data set. In this case the DEEPCSV algorithm is used for computing the b-tagging score of the containing jet. Otherwise the algorithm is similar to that used originally for the $t\bar{t}H$ observation. This analysis depends less crucially on the distinction between prompt and nonprompt leptons, so the definition of custom lepton classifiers would give only tiny improvements to the analysis. This was considered not worth the effort.

6.1.3 Lepton identification efficiency measurement

The identification of leptons relies on many inputs that are relatively complex to model in simulation. Predictions of the contents of the jet containing a lepton and the activity inside a cone around the lepton rely on an accurate description of the underlying event, pileup conditions, detector noise, and many other factors. Yields from processes involving only prompt leptons are computed from simulation in each analysis, and these need to be corrected for potential differences in the prompt lepton selection efficiency between data and simulation. The background contribution from nonprompt leptons is predicted from data (Section 6.4) and does not directly rely on the modeling of the lepton selection in simulation.

The efficiency of selecting prompt leptons is measured with the tag-and-probe technique [184, 202]. A data sample enriched in Drell-Yan production is defined by selecting events with two electrons or two muons with an invariant mass around m_Z . One lepton, the tag, is required to pass a set of stringent selection criteria that makes it unlikely to be a fake. The tag lepton must also pass a trigger path that selects one lepton. In this way we can measure the efficiency of selecting the second lepton, the probe, without any bias from the trigger selection. The set of selection criteria for which the efficiency is to be measured is applied to the probe lepton. The number of events with prompt probe leptons that pass and fail the selection are counted in different bins of p_T and η and give the efficiency of the selection criteria in data.

The number of prompt and nonprompt leptons in each bin of p_T and η , and for events in which the probe passes and fails the selection, are determined from fits to the dilepton invariant mass distribution.

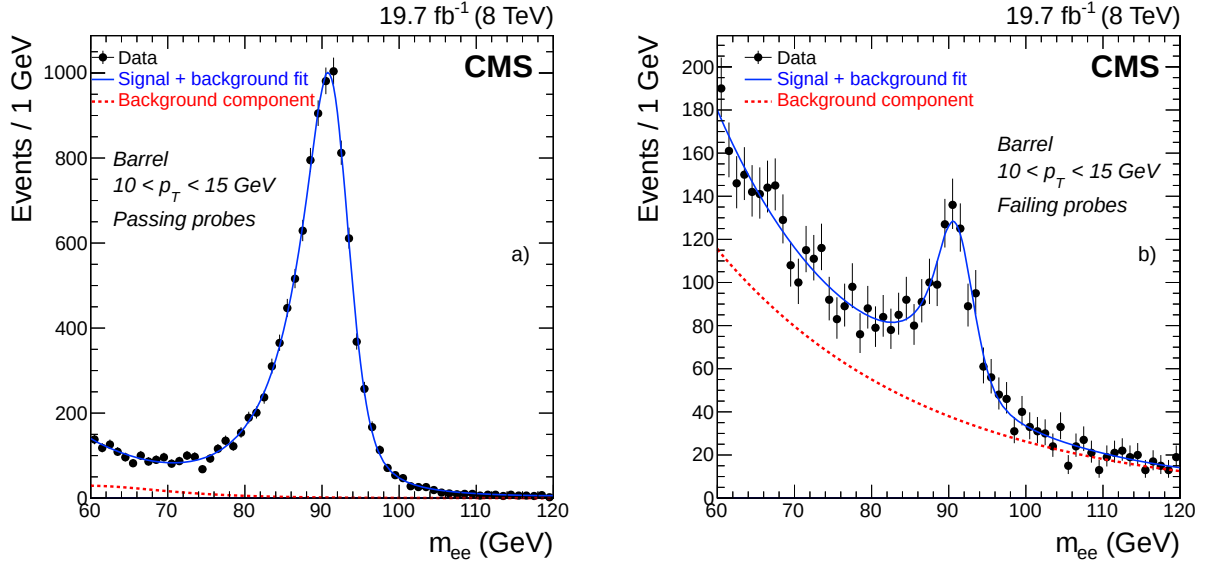


Figure 6.1: Example of fits to the dilepton invariant mass distribution used to determine the number of prompt leptons that pass and fail particular selection criteria. Events from 8 TeV pp collision data in which probe electrons pass a set of standard impact parameter and I_{rel} requirements are shown on the left, while events in which the probes fail the selection are shown on the right [202].

The mass distributions are made up of a peak centered around m_Z , and a smoothly falling background distribution. A fit with two components, one for the Z boson mass peak and one for the background contribution yields the number of signal and background events. The signal events are those that contain prompt leptons. The fit functions are templates based on simulated events. An example of such fits is shown in Fig. 6.1 for electrons with p_T values between 10 and 15 GeV in 8 TeV pp collision data.

Several sources of systematic uncertainty affect the measurement of lepton selection efficiencies with the tag-and-probe procedure. The selection of the tag lepton is varied to be looser and tighter, and the effect on the measured probe selection efficiencies is taken to be a systematic uncertainty. The simulation based fit templates for the signal and background invariant mass shape are separately replaced with empirical functions. The differences between the nominal efficiency measurements and those obtained with the alternative fit functions are considered systematic uncertainties. The generator used to determine the fit templates in simulation is also varied. In the past the selected mass range was also varied, but this uncertainty is no longer considered in modern analyses because it can artificially increase the background to the efficiency measurement.

In recent analyses measurements of the lepton selection efficiency in Drell-Yan enriched data have become so precise, of the order of 0.1% in many p_T and η bins, that variations in the lepton selection efficiency from the extrapolation to the analysis signal region have to be considered. If the underlying event density and number of objects selected in the analysis increases, the chance for accidental overlap between prompt leptons and other activity in the event increases. As a consequence the efficiency of selecting prompt leptons decreases. A simple example of this is that the efficiency of selecting a given muon or electron in $e\mu$ events is typically lower than in ee and $\mu\mu$ events because $e\mu$ events originate mainly from $t\bar{t}$ production and tend to contain several jets. Each analysis presented in the following chapters includes a brief discussion of the extrapolation uncertainties that are being considered. These differ depending on the event selection criteria of the analysis.

6.2 Trigger selection

Section 4.8 introduced the CMS trigger system which filters out interesting collision events during data taking for permanent storage. Because of the importance of leptons to many analyses, more than half of the L1 trigger budget is assigned to events with one or more muons or electrons ⁴ [188]. Analyses typically use a data set selected with trigger algorithms matching the final state of interest as closely as possible. In our case this means events with several leptons.

The three analyses in chapters 7 to 9 follow the same strategy when it comes to triggering: Events are retained for analysis if they pass any unprescaled single, two, or three lepton trigger. Single lepton triggers tend to have relatively high p_T requirements and stringent identification criteria on the leptons to limit the rate of events. Both the p_T and identification requirements are loosened in dilepton triggers because these suffer from lower backgrounds, and the Z boson production cross section is markedly lower than that of the W boson (See Fig. 6.2). Three lepton triggers tend to have only very loose p_T and identification requirements because of the rarity of the targeted event topology. The idea behind combining all these trigger algorithms is that it leads to a large redundancy in selecting events with three or more leptons. For a three lepton event to be selected it is enough if any of the three leptons pass one of the single lepton triggers, any pair of two leptons passes a dilepton trigger, or the three leptons pass a three lepton trigger together.

It is not necessarily guaranteed that an event selected offline could pass a trigger with equivalent selection criteria. This is due to the fact that the reconstruction algorithms used in the trigger system are modified for speed in favor of precision, and do not yield exactly the same results as those used during offline reconstruction. In order to accurately predict expected event yields with simulation, the trigger algorithms must also be applied to simulated events. The remaining differences between data and simulation must then be accounted for by comparing the efficiency with which selected events pass a set of trigger paths.

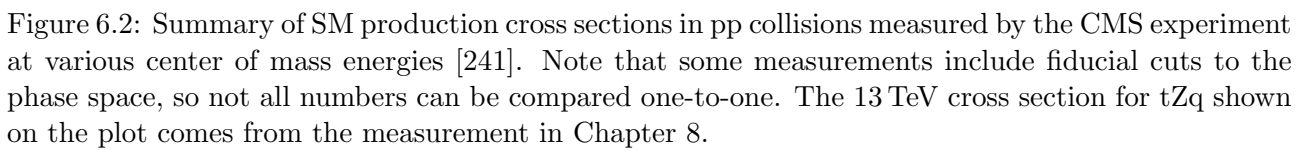
In each analysis the trigger efficiency is measured in an unbiased data set obtained by selecting events that pass a trigger requirement on the p_T^{miss} or hadronic activity in event. Imposing such requirements does not bias the efficiency measurement of leptonic triggers. The lepton selection criteria of the analysis are applied on this data set, and the probability that selected events pass the set of leptonic trigger requirements is evaluated. In events with three or more light leptons the trigger redundancy leads to an efficiency that is effectively 100% across most of the selected kinematic spectrum. Exceptions to this are events with three low p_T leptons in Chapter 7, and events with two or fewer light leptons in Chapter 9. In the former case an additional per lepton efficiency for each trigger path, measured with the tag-and-probe technique, is used to gain more insight in the global trigger efficiency. More details are given in the corresponding chapter.

In what follows whenever plots or reconstructed quantities are shown it will be assumed that the events have passed the set of trigger requirements mentioned here. Specific trigger efficiencies and their corresponding uncertainties for each analysis are given in the corresponding chapters.

6.3 Standard model processes in the multilepton final state

This section gives an overview of SM processes contributing to the multilepton final state. Most of these processes are backgrounds to the analyses performed in this work. Several quantities that can be used to reduce the yields of specific background processes are introduced. The contribution of each process to the analyses is estimated from simulation, if possible with input from a control region in data. The exception to this is the background from nonprompt leptons, the prediction of which is separately discussed in Section 6.4. A summary of SM cross sections measured by the CMS experiment is given in Fig. 6.2. This gives an idea of the relative rates and importance of different processes.

⁴Trigger seeds targeting electrons also simultaneously target photons since they can not be efficiently distinguished without information from the tracking detector.



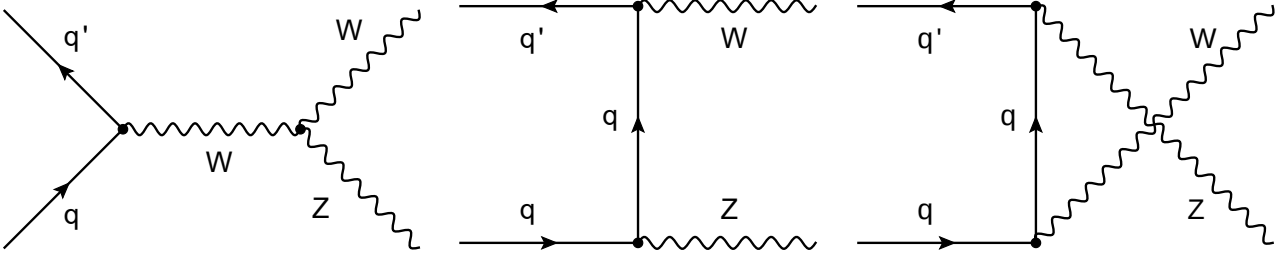


Figure 6.3: Leading order Feynman diagrams for WZ production [242].

6.3.1 WZ production

Associated production of a W and a Z boson (WZ) results in events with three prompt leptons and $p_{\text{T}}^{\vec{\text{miss}}}$ when both the W and the Z boson decay leptonically. The leading order Feynman diagrams for WZ production are shown in Fig. 6.3. The product of the WZ production cross section and branching fraction to three leptons is large compared to that of other processes in the three lepton final state. After applying criteria for selecting prompt leptons this process is therefore the dominant source of data events with three leptons. The process is characterized by the presence of an opposite-sign and same-flavor (OSSF) lepton pair from the Z boson decay, and the bulk of WZ events contain none or just a few jets, and no b jets.

The WZ process forms an irreducible background in the analyses of chapters 7 and 9 which target multilepton events without additional jets. The tZq measurement in Chapter 8 targets events with several jets, at least one of them b-tagged. This last requirement severely reduces the WZ background contribution. Nonetheless it remains the largest background in the most sensitive event category of this search.

Two kinematic quantities are particularly useful for reducing the background yield from WZ production. The reconstructed invariant mass of the two leptons from the Z boson decay ($M_{\ell\ell}$) peaks strongly around m_Z . In $e\mu\mu$ and $\mu\mu e$ events leptons can be unambiguously assigned to the Z boson decay. In $\mu\mu\mu$ and eee events we assign the OSSF lepton pair forming the invariant mass closest to m_Z to the Z boson decay. The remaining lepton is assumed to come from the decay of the W boson. The neutrino from the W boson decay can not be directly detected, but leads to a sizable $p_{\text{T}}^{\vec{\text{miss}}}$. This can be combined with the lepton from the W boson decay to reconstruct the transverse mass of the W boson candidate with the following equation:

$$M_{\text{T}} = \sqrt{2p_{\text{T}}^{\ell}p_{\text{T}}^{\vec{\text{miss}}} [1 - \cos(\Delta\phi)]}, \quad (6.4)$$

with $\Delta\phi$ the azimuthal separation between the lepton and $p_{\text{T}}^{\vec{\text{miss}}}$. Physically, M_{T} represents the invariant mass of the transverse projection of the lepton and $p_{\text{T}}^{\vec{\text{miss}}}$ in the ultrarelativistic limit. The distribution of M_{T} peaks around m_W and has a broad distribution towards lower values because the longitudinal component of the W boson decay products is not accounted for. Beyond m_W , the M_{T} distribution falls off sharply, as can be seen in Fig. 6.4. All three of the analyses presented in this work use M_{T} as a discriminating variable against the background from WZ production.

The tail of the M_{T} distribution in multilepton WZ events is of particular importance in searches for heavy hypothetical particles, such as those of chapters 7 and 9. Production of massive new particles is expected to result in high- M_{T} events. Events from WZ production with M_{T} values exceeding m_W mainly come from the ambiguity in assigning leptons to the W and Z boson decays in eee and $\mu\mu\mu$ events. The lepton from the W boson decay can accidentally form an invariant with one of the Z boson decay products that is closer to m_Z than the invariant mass of the two leptons from the actual Z boson decay. One of the Z boson decay products is then wrongly assigned to the W boson decay. If this happens, M_{T} no longer has a boundary at m_W and can take on large values. To a lesser extent, $p_{\text{T}}^{\vec{\text{miss}}}$ mismeasurements and the W boson width can also lead to M_{T} values larger than m_W . The former is the dominant contribution to the tail of the M_{T} distribution in $e\mu\mu$ and $\mu\mu e$ events. These

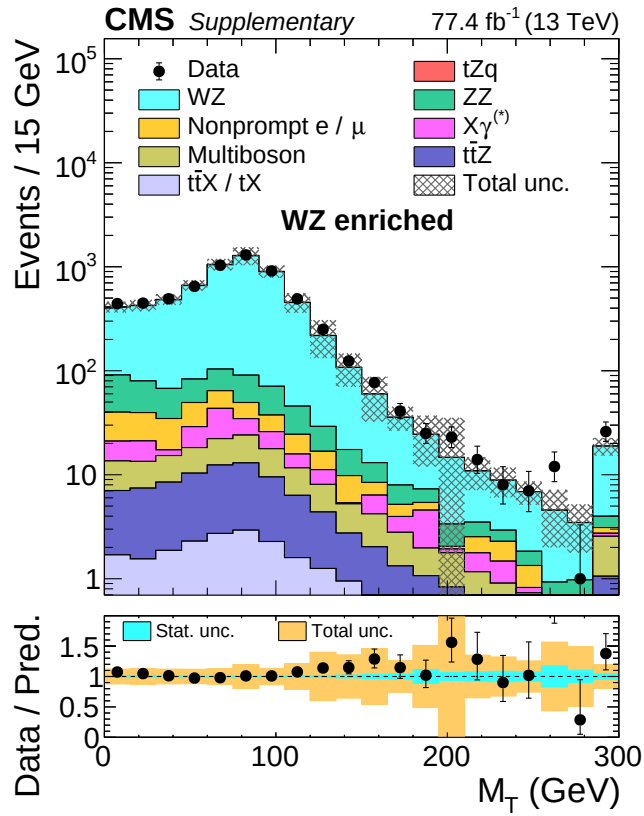


Figure 6.4: Observed and expected number of events in a WZ enriched data sample as a function of M_T . The shown data sample is the WZ enriched region of the tZq search presented in Chapter 8. A detailed description of the corresponding event selection can be found in that chapter.

properties of the M_T tail in WZ events were identified by me in the context of [5], a previous iteration of the search presented in Chapter 9. Chapter 9 will go into more detail on this subject as it is the most dependent on an accurate description of high- M_T WZ events.

Throughout this work the normalization of the WZ event yield is implicitly obtained from data by including a data region enriched in WZ events in the final fit of each analysis. The exact definition of this control region varies depending on the constraints of the analysis, and is discussed separately in each analysis chapter.

6.3.2 ZZ production

Simultaneous production of two Z bosons (ZZ) with subsequent leptonic decays of both bosons yields events with four prompt leptons. It is the largest source of four prompt lepton events, and the main background in the analysis categories of Chapter 9 that target signals in the four lepton final state. When one of the leptons fails to pass the reconstruction or identification criteria, the process can enter the three lepton data set. Light lepton identification is efficient and the ZZ production cross section is not as high as that of WZ production. The contribution of ZZ production to the three lepton final state therefore tends to be a great deal smaller than that of WZ production. Nonetheless, it remains a significant background source in three lepton events.

No neutrinos are produced in leptonic Z boson decays, so multilepton ZZ events tend to have low p_T^{miss} values. Usually, they also have few jets and no b jets. Similar to what is done for WZ, each of the three analyses uses a data set with a high ZZ purity to constrain the production cross section of this process. More details on the exact definition of these data samples are given when discussing each analysis.

6.3.3 Photon conversions

Photon conversions can add additional leptons to an event. In external conversions real photons interact with the detector material to produce an electron-positron pair. Virtual photons, on the other hand, promptly decay to a dilepton pair, an internal conversion. In the analyses presented in this dissertation all nominally selected leptons are required have a p_T larger than 10 GeV, muons in Chapter 7 being the sole exception with a 5 GeV threshold. Conversions in which both leptons pass such a p_T requirement are rare. The conversions that do result in leptons that pass the selection tend to be “asymmetric conversions” in which one of the leptons carries the bulk of the momentum of the converted photon and passes the lepton selection alone.

In this work we always estimate the background from prompt photon conversions from simulation. Leptons from the conversion of nonprompt photons are considered nonprompt, and the corresponding background yields are estimated with the data-driven techniques of Section 6.4. In what follows we will only consider conversions of prompt photons when discussing the conversion background.

The main contribution from conversions in three lepton events comes from Z boson production in association with a real or virtual photon ($Z\gamma^{(*)}$). The Z boson decays to two leptons, and an asymmetric photon conversion can result in an additional selected lepton. In many $Z\gamma^{(*)}$ events the photon comes from FSR of one of the leptons, resulting in an $\ell\ell\gamma$ system with the mass of a Z boson. If the conversion is asymmetric, this in turn results in a three lepton system of invariant mass ($M_{3\ell}$) close to m_Z . This feature is clearly visible in Fig. 7.5. One of the Feynman diagrams of $Z\gamma^{(*)}$ production with an FSR photon is drawn in Fig. 6.5. A data sample enriched in $Z\gamma^{(*)}$ events can thus be obtained by selecting events with three leptons and requiring $M_{3\ell}$ to be compatible with m_Z . Such a control region is used in chapters 7 and 9 to derive the normalization of the $Z\gamma^{(*)}$ event yield in data. In Chapter 8 such a sample is only used to validate the simulated predictions for conversions because the conversion background is of lesser importance in that chapter.

Photon conversions leading to multilepton events can spring from many other process than $Z\gamma^{(*)}$, including $t\bar{t}\gamma^{(*)}$, $W\gamma^{(*)}$, diboson production with an additional photon, and many more. The contribution from $W\gamma^{(*)}$ in three lepton events tends to be low because both leptons from the conversion are

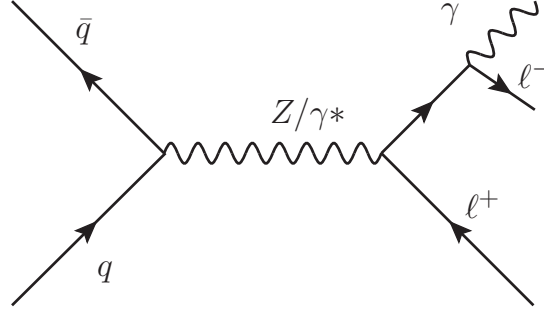


Figure 6.5: A possible Feynman diagram of $Z\gamma^{(*)}$ production in which a photon is radiated by one of the final state leptons. If the photon undergoes an asymmetric conversion, the final state will contain three leptons with an invariant mass very close to m_Z [243]. This three lepton invariant mass peak is prominent in Fig. 7.5.

unlikely to pass the lepton selection. The other listed processes have small cross sections compared to $Z\gamma^{(*)}$, making their yields minute. The combination of all backgrounds containing converted photons is referred to as $X\gamma^{(*)}$ in what follows, but is usually largely made up of $Z\gamma^{(*)}$. Events with a final state consisting of two leptons of the same sign, selected in Chapter 9, are the exception. In this event category the dominant contribution from photon conversions comes from $W\gamma^{(*)}$ production, again with an asymmetric photon conversion.

6.3.4 Nonprompt leptons

Events with one or more nonprompt leptons form an important background source in multilepton data. Most nonprompt light leptons can be rejected by imposing identification requirements (Section 6.1). However, even after the application of such selection criteria the background often remains significant. In events with τ_h candidates the background from nonprompt leptons tends to be far larger than any other background source because of the high misidentification rate of jets as τ_h candidates (Section 5.5). The prediction of the nonprompt lepton background in all the analyses presented is done directly from data with the procedure explained in Section 6.4.

A host of processes can result in events with one or more nonprompt leptons that manage to pass the analysis selection. In analyses targeting the three lepton final state, the majority of these come from $t\bar{t}$ production with two leptonic top quark decays, and Drell-Yan production. Both of these processes have two prompt leptons in the final state and need only one nonprompt lepton to pass the selection. Drell-Yan production has a cross section that is several orders of magnitude larger than that of any process resulting in three or more prompt leptons, so even a small misidentification rate for nonprompt leptons can result in a large background. Although the $t\bar{t}$ production cross section is a lot smaller than that of Drell-Yan production, it is still relatively large, and almost every $t\bar{t}$ event contains at least two b quarks which can decay semileptonically. This results in a nonprompt lepton background from $t\bar{t}$ production that is often even larger than that of Drell-Yan production.

Nonprompt leptons constitute the dominant background in the most sensitive signal regions of Chapter 7, and were the main obstacle preventing the discovery of the tZq process by CMS in the 2016 pp collision data set [244]. The reduction of this background by improving the lepton identification procedure led to the first observation of tZq as discussed in Chapter 8. In Chapter 9 the nonprompt background is significant in most event categories and the dominant source of background events in final states with τ_h candidates.

6.3.5 Top quark associated production

Multiple prompt leptons can originate from the production of top quark pairs or single top quarks in association with one or more electroweak bosons. Associated top quark pair production tends to have

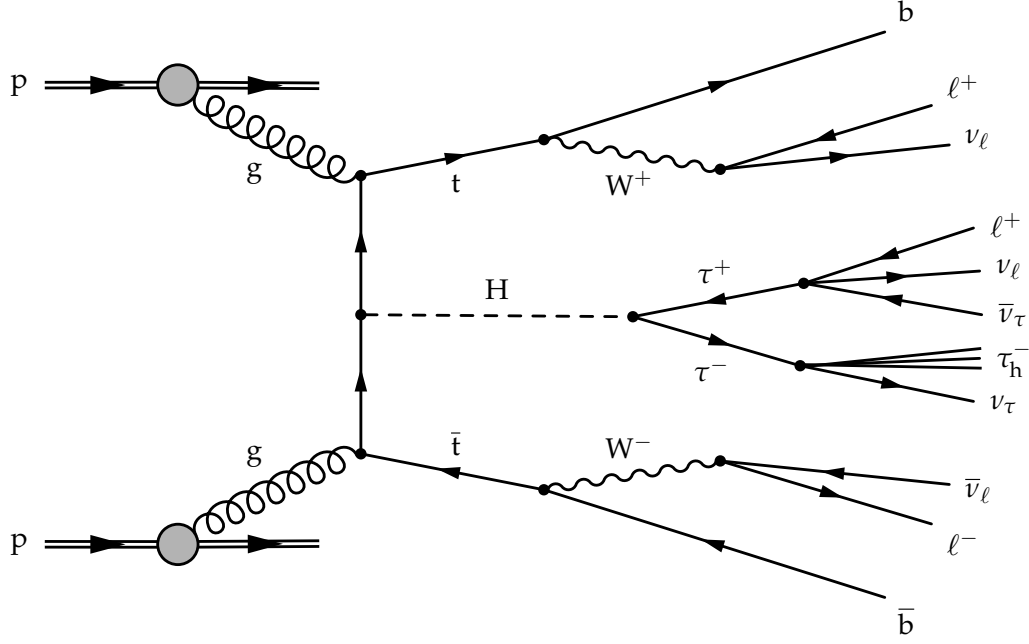


Figure 6.6: Diagram of a $t\bar{t}H$ production event resulting in a final state with three light leptons and a τ_h candidate [238].

both higher cross sections and larger branching fractions to multiple leptons than processes with a single top quark. Among processes with associated production of top quarks and electroweak bosons $t\bar{t}Z$, $t\bar{t}W$, and $t\bar{t}H$ production have the largest three lepton event yields. An example of one of the possible decay modes in $t\bar{t}H$ production that result in multiple leptons is presented in Fig. 6.6.

These processes are characterized by the presence of several jets, including one or more b jets. Signal events are not expected to contain b jets in chapters 7 and 9, so the presence of b jets is vetoed to reduce the background of processes that involve top quarks. Because of the b jet veto and the small production cross sections compared to diboson production, associated top quark production is only a minor background in these analyses. Chapter 8 targets a signal that itself contains a top quark. In this analysis $t\bar{t}Z$ production forms the dominant background in two out of three signal regions.

The tHq , tWZ , and tZq (the subject of Chapter 8) production processes tend to only be tiny sources of multilepton events. Even more rare processes, such as $t\bar{t}t\bar{t}$, or top quark pair production with multiple electroweak bosons are as good as negligible in all analyses discussed in this dissertation. These processes are nonetheless taken into account in the background estimation.

In each analysis the yields of all of the aforementioned processes are estimated from simulated events normalized to the respective theoretical cross sections. In Chapter 8 the $t\bar{t}Z$ cross section is constrained implicitly from data when extracting the tZq cross section.

6.3.6 Multiboson production

Production of three or more electroweak bosons can bring about a large amount of prompt leptons. Examples are WWW , WWZ , and WHZ production with subsequent leptonic decays of the bosons, but many other combinations of electroweak bosons are possible. Multiboson production processes all have extremely small cross sections, and are of limited importance in this work. Without exception these processes are estimated purely on the basis of simulation, using theoretical cross sections to normalize their event yields. Most of these processes have never been precisely measured, so the uncertainties in their predictions are necessarily large. Given the rarity of these processes that presents no problem.

6.4 Nonprompt lepton background prediction

Lepton candidates that originate from jet fragmentation or are outright faked by other particles are collectively classified as nonprompt and are an important background source in any analysis using leptons. Predicting this background accurately from simulation is more difficult than predicting backgrounds with only prompt leptons. For this reason the nonprompt lepton background is predicted directly from data in all the analyses presented in this dissertation. The procedure detailed below aims at predicting the background from the sum of all processes that pass the analysis selection in which at least one of the selected leptons is nonprompt.

The data-driven background prediction relies on the knowledge of the probability that a lepton passing a loosened selection also passes the full lepton selection of the analysis. This probability is called “fake rate”, and the loosened lepton selection will be called “loose”. The nominal selection criteria are designated “tight”. Only events in which all leptons pass the tight selection are considered part of the signal region of an analysis. Events in which one or more leptons fail the tight selection while still passing the loose selection, referred to as sideband or application region events, are used in tandem with the fake rate to obtain a prediction of the nonprompt lepton background. In the background prediction each sideband event is assigned the following weight:

$$w = (-1)^{(N_{\text{loose-not-tight}}+1)} \prod_{\ell}^{\text{loose-not-tight}} \left(\frac{f(\ell)}{1-f(\ell)} \right), \quad (6.5)$$

where the product runs over all leptons that are loose but not tight (loose-not-tight), $N_{\text{loose-not-tight}}$ is the number of such leptons, and $f(\ell)$ is the fake rate which depends on the properties of the lepton ℓ . Weighting sideband events by the factor w gives a prediction of the number of events with one or more nonprompt leptons in which all leptons pass the tight selection. The factor $1-f(\ell)$ in the denominator accounts for the fact that only events in which some leptons fail the tight selection enter the sideband.

The factor $(-1)^{(N_{\text{loose-not-tight}}+1)}$ is negative for even numbers of loose-not-tight leptons to correct for the double counting of events with multiple nonprompt leptons. The necessity of this factor can be seen from the following example: Imagine requiring two tight leptons and predicting the background from nonprompt leptons. First we look at events with one loose-not-tight and one tight lepton. If only one lepton is nonprompt, it has probability $f(\ell)$ to be tight. Applying a weight of $f(\ell)/(1-f(\ell))$ to events with one nonprompt loose-not-tight lepton results in a correct background prediction. If there are two nonprompt leptons, either one of them could have passed the tight selection to result in an event with a tight and a loose-not-tight lepton. Applying the weight $f(\ell)/(1-f(\ell))$ to such events results in an overprediction of the number of events with two tight nonprompt leptons by a factor two. Events with two loose-not-tight leptons weighted by $f(\ell)f(\ell')/[(1-f(\ell))(1-f(\ell'))]$ do give a correct prediction of the background from events with two nonprompt leptons. Events with two loose-not-tight leptons with a negative weight can thus be used to cancel the aforementioned overprediction.

The method presented here applies to both light leptons and τ_h candidates. A successful application of the method to nonprompt light leptons in multilepton events is however far more involved than it is for τ_h candidates. The technical details that follow in this section only apply to light leptons. Events with τ_h candidates are exclusively used in Chapter 9, and the necessary details on the prediction of the nonprompt τ_h candidate background are given there.

The fake rate depends on the properties of the lepton. So measuring the fake rate in data should be done as a function of the variables it depends on most, while still maintaining a sufficient statistical precision for a robust measurement. The detector layout changes significantly as a function of η , so this is an obvious choice for the parametrization of the fake rate. The fake rate of leptons also changes significantly with their p_T , though the situation is a bit more complex than for the η dependence. This is the subject of Section 6.4.1. Multijet data provides a relatively pure and large source of nonprompt leptons in which the fake rate can be measured. Ensuring that the fake rate measured in such events is

applicable to analyses in the multilepton final state requires care. This is the subject of Section 6.4.2. Finally Section 6.4.3 discusses the fake rate measurement in data.

6.4.1 Cone correction

Studies in simulated events reveal that the fake rate depends on the p_T of the parton from which a nonprompt lepton originates ($p_T^{\text{mother parton}}$), more so than on the p_T of the lepton itself. We want to approximate $p_T^{\text{mother parton}}$ based on reconstructed quantities to be able to use it in the parametrization of the fake rate. This is done by correcting the p_T of leptons that are loose-not-tight with a “cone correction” that accounts for the momentum of the other decay products of the original parton. For tight leptons, which will most often be prompt, no momentum correction should be applied since the momentum of the lepton itself is what is of physical interest. The lepton fake rate is parametrized and measured as a function of p_T^{cone} , the cone corrected p_T .

The anti- k_T jet clustering algorithm (Section 5.2) generally clusters the decay products of the mother parton, including the nonprompt lepton, into a jet. A cone correction procedure can thus be defined as replacing the lepton p_T by that of the jet it belongs to. In practice this is done by dividing the lepton p_T by p_T^{ratio} (Section 6.1.2). We obtain the following definition for the cone correction:

$$p_T^{\text{cone}} = \begin{cases} p_T & \text{if lepton passes the tight selection} \\ C \cdot \frac{p_T}{p_T^{\text{ratio}}} & \text{otherwise} \end{cases} \quad (6.6)$$

, where a factor C is introduced to ensure a continuous p_T^{cone} distribution. Tight leptons are not corrected, while the p_T of loose-not-tight leptons almost always increases since the lepton p_T should be smaller than that of the jet containing it. A priori, the distribution of p_T^{cone} therefore has a discontinuity as a function of the variables used in the identification of leptons. The factor C is smaller than unity, and removes this discontinuity.

The analyses of chapters 8 and 9 use a threshold on a BDT classifier to select leptons. I designed a novel way of determining the parameter C , used in these analyses. The average p_T^{cone} with $C = 1$ is studied as a function of the BDT output in simulated events. A discontinuous drop in the average p_T^{cone} at the BDT threshold used in the tight lepton selection is observed. The parameter C is defined as the ratio between the average p_T^{cone} just above and below the BDT selection threshold. A finely binned histogram is used to determine the average p_T^{cone} at both ends of the threshold. As a cross check a polynomial template is fit locally to the average p_T^{cone} as a function of the BDT output on both sides of the threshold. Scaling down the polynomial fit on the lower end of the BDT threshold so it crosses the other one at the threshold results in an equivalent value for C . The value of C that is obtained is 0.95 for both electrons and muons in the tZq search of Chapter 8, while it is only 0.65 for muons in the three and four lepton selection used in Chapter 9. The reason for the smaller factor C in Chapter 9 is a less stringent lepton selection threshold. In general C will be closer to unity the more strict the BDT threshold for lepton selection is. This can be understood by observing that very high BDT scores tend to correspond to very prompt-like leptons with p_T^{ratio} values close to unity.

In Chapter 7, a classical I_{rel} based lepton identification strategy is used instead of BDT classifiers. In this case the cone correction procedure can be simplified by approximating the p_T of the jet that contains the lepton as the sum of the lepton p_T and the p_T of the other objects in the isolation cone around the lepton direction. This is equivalent to scaling up the p_T of the lepton by a factor $1 + I_{\text{rel}}$. This has the advantage that continuity in the average p_T^{cone} can be obtained for free by defining p_T^{cone} as follows:

$$p_T^{\text{cone}} = \begin{cases} p_T & \text{if lepton passes tight selection} \\ p_T \cdot \left[1 + \max(I_{\text{rel}} - I_{\text{rel}}^{\text{tight}}, 0) \right] & \text{otherwise} \end{cases} \quad (6.7)$$

, where $I_{\text{rel}}^{\text{tight}}$ is the isolation threshold used in the tight selection. Defining p_T^{cone} as such corresponds to using a fixed cone size around the nonprompt lepton to aggregate decay products from the mother

parton. This gives a slightly worse approximation of $p_T^{\text{mother parton}}$ than Equation 6.6. The definition is nonetheless adequate for parametrizing the fake rate and achieving a precise prediction of the nonprompt lepton background.

Equation 6.7 is used to define p_T^{cone} in the heavy neutrino search in Chapter 7. The other two, more modern, analyses in chapters 8 and chapter 9 use the generalized definition from Equation 6.6. In all analyses the leptons are cone corrected after selection. This means that for all loose-not-tight leptons their p_T is replaced by p_T^{cone} and their energy is scaled up by p_T^{cone}/p_T . The fake rate is measured as a function of p_T^{cone} , so correcting the p_T of all loose-not-tight leptons ensures a correct prediction of the p_T spectrum of nonprompt leptons passing the tight selection. The cone correction is propagated to the computation of all derived analysis variables, such as invariant masses. In the following sections and chapters whenever lepton p_T is mentioned in the context of an analysis we implicitly refer to p_T^{cone} , unless the discussion is about lepton identification requirements. Note once again that the p_T of tight leptons, which are those that enter the final analysis selection, remains unchanged when applying either Equation 6.6 or 6.7. Only sideband events used to predict the nonprompt lepton background are affected by the cone correction.

6.4.2 Loose selection tuning

Section 6.3.4 discusses that a multitude of processes can result in background events with nonprompt leptons. When selecting three leptons, $t\bar{t}$ and Drell-Yan production make up the overwhelming majority of selected events with nonprompt leptons. The measurement of the fake rate is however performed in a data set enriched in QCD multijet events, following the procedure explained in Section 6.4.3.

Each $t\bar{t}$ event contains at least two b jets. Jets in Drell-Yan events, on the other hand, come from ISR or FSR, so most of these are gluon and light quark jets. Similarly, QCD multijet events have a larger light quark and gluon jet fraction than $t\bar{t}$. The lepton fake rate for light and heavy flavor jets is inherently different, in particular because heavy flavor jets often contain genuine leptons from the decay of a b or c quark. This presents a challenge since a fake rate measured in QCD multijet events is not necessarily applicable to the nonprompt lepton background prediction in the multilepton final state. Meanwhile a correct description of one background source, such as Drell-Yan production, does not guarantee a correct description of the others.

The problem is solved by removing the jet flavor dependence in the fake rate by tuning the requirements that constitute the loose lepton selection. We wish to obtain a loose definition which results in a fake rate that is equivalent for light and heavy flavor jets. If this is achieved, a measurement of the fake rate in multijet events can be used to predict all backgrounds involving nonprompt leptons in a multilepton analysis. At the same time the loose selection should, as the name implies, remain as loose as possible. Otherwise the fake rate will become large, which would lead to a poor statistical precision in the nonprompt lepton background prediction. When the fake rate is small, the sideband region contains an unweighted number of events larger than the background yield predicted with it. It is also important that the tight lepton selection remains unmodified during the fake rate tuning. Otherwise the signal selection efficiency of the analysis might get spoiled. So when tuning the fake rate, additional selection criteria are only added for leptons that fail the tight requirements.

The tuning procedure is based on closure tests in simulated events. First the fake rate is determined in a QCD multijet simulation for a particular loose selection definition. Subsequently, events with three tight leptons, at least one of them nonprompt, are selected in simulated $t\bar{t}$ and Drell-Yan events. For both processes this direct background prediction is compared to that of the fake rate method applied to simulated sideband events, using the fake rate measured in simulated multijet events. The comparison is made as a function of all event variables used in the analysis. Without tuning of the loose selection the direct prediction and that of the fake rate method generally do not agree. The loose selection is adjusted based on the observed discrepancies between the fake rate method and the direct predictions for both $t\bar{t}$ and Drell-Yan production. This procedure is iterated until the predictions of the fake rate method match those of direct simulation to an adequate degree. The tuning of the loose selection is always done separately for each data-taking year because of differences in the detector and

its operational conditions, in instantaneous luminosity, and in the event reconstruction. All of these differences lead to variations in the lepton fake rate.

In the heavy neutrino search in Chapter 7 a classic lepton identification strategy is used which exploits a small amount of variables, limiting the options for tuning the loose selection. For muons only an I_{rel} cut is adjusted, while for electrons cuts on both I_{rel} and the electron MVA are tuned in the loose selection.

The two newer analyses in chapters 8 and 9 use BDT classifiers with a set of thirteen input variables for lepton selection. This results in a complex multivariate fine tuning problem to make the fake rate predictions as good as possible. In practice, cuts on the b-tagging score of the jet containing the lepton and on $p_{\text{T}}^{\text{ratio}}$ are tuned for the loose lepton selection. The former is a particularly powerful tool for regulating the fake rates of nonprompt leptons from heavy- and light flavor jets. For electrons a cut on the electron MVA is simultaneously tuned with those on $p_{\text{T}}^{\text{ratio}}$ and the b-tagging score.

The search presented in Chapter 9 is the most recent analysis in this dissertation, and has the most refined tuning of the loose lepton selection. A p_{T} -dependent cut on the b-tagging score is used to get a precise description of the p_{T} spectrum of nonprompt leptons from the fake rate method, regardless of the origin of the leptons. Simulated closure tests for this analysis that compare the prediction of the fake rate method to direct simulation, following the description given earlier, are shown in Fig. 6.7. The fake rate from a multijet simulation can be seen to yield an excellent prediction of the simulated $t\bar{t}$ background in 2018 data-taking conditions. Closure tests for other data-taking years and for Drell-Yan production are not included here, but satisfactory predictions by the fake rate method are seen in each case. This gives us confidence that measuring the fake rate in multijet data results in an accurate description of the nonprompt lepton background in multilepton events with a well tuned loose lepton selection. Similar closure tests were done for all three analyses presented in this work, but not included for brevity.

6.4.3 Fake rate measurement

After the loose lepton selection is tuned, the fake rate is measured as a function of $p_{\text{T}}^{\text{cone}}$ and $|\eta|$ in a data sample enriched in QCD multijet events. One starts by selecting events with a single loose lepton. At least one jet with a large angular separation from the lepton, characteristic of multijet events, is required to be present. The fraction of multijet events in which the loose lepton also manages to pass the tight selection is a measure of the fake rate for a given loose selection definition. The selected data sample is however contaminated by prompt leptons, chiefly from W boson production, and to a much smaller extent from Drell-Yan and $t\bar{t}$ production. These prompt leptons have to be accounted for in the fake rate measurement to avoid an overestimation of the fake rate induced by their high identification efficiency.

Unprescaled single lepton triggers (Section 4.8) have p_{T} thresholds well above those set in the selection of single leptons in the multilepton analyses presented in this dissertation. These triggers typically also impose identification requirements on the leptons that could spoil a fake rate measurement. For this reason we use data collected with a collection of prescaled single lepton triggers. These prescaled triggers impose only low p_{T} requirements on the leptons and have relaxed or no identification criteria. Triggers are available that select muons with p_{T} values as low as 3 GeV. For electrons the prescaled triggers start at a p_{T} requirement of 8 GeV. In the lowest p_{T} prescaled trigger paths the lepton is required to be associated with a jet. The set of prescaled triggers used in the fake rate measurement contains trigger paths with progressively higher p_{T} thresholds on the leptons. The prescale values decrease with increased p_{T} thresholds because the trigger rate falls off as a function of the lepton p_{T} requirement. To maximize the statistical precision, each $p_{\text{T}}^{\text{cone}}$ bin in the fake rate measurement uses the trigger path with the lowest possible prescale value.

A complication arises here: Triggers place requirements on the p_{T} of the lepton while we want a fake rate measurement binned in $p_{\text{T}}^{\text{cone}}$. For leptons that pass the tight selection p_{T} and $p_{\text{T}}^{\text{cone}}$ are equal, but for those in the denominator of the fake rate measurement $p_{\text{T}}^{\text{cone}}$ is almost always larger than p_{T} . One should ensure that only one prescaled trigger path is used in each $p_{\text{T}}^{\text{cone}}$ bin of the fake

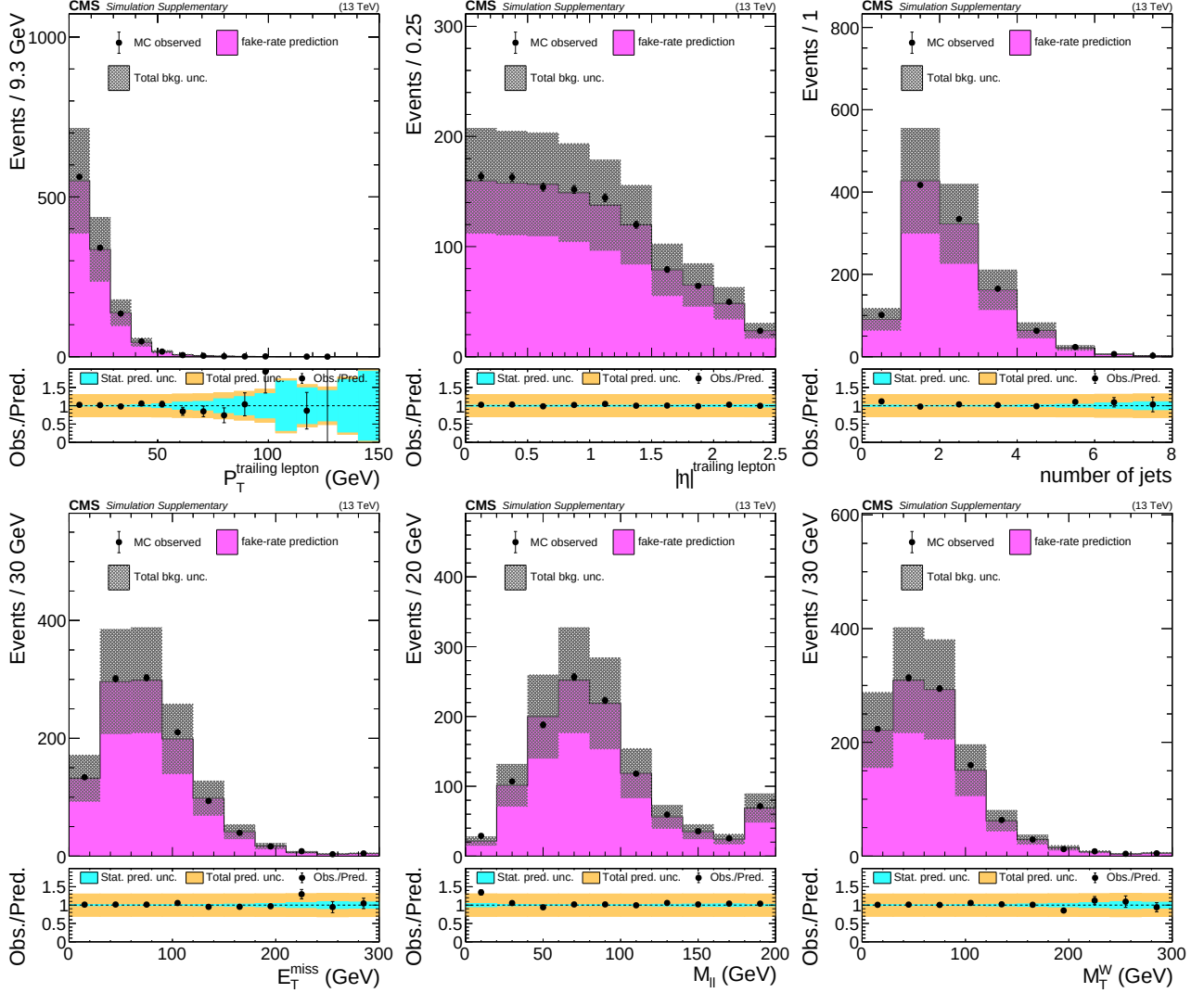


Figure 6.7: Number of events with three leptons, at least one of them nonprompt, passing the tight selection defined in Chapter 9. Direct predictions from simulation (black data points) are compared to the prediction of the fake rate method applied to simulated sideband events as (magenta distribution). Yields are shown as a function of (left to right, top to bottom): The p_T and $|\eta|$ of the lowest p_T lepton in the event, the number of selected jets in the event, p_T^{miss} , the invariant mass of the best $Z \rightarrow \ell\ell$ candidate, and the M_T of the remaining lepton and p_T^{miss} . It is worth noting that the lowest p_T lepton tends to be the nonprompt one. The exhibited events are from a simulation of $t\bar{t}$ production and the applied fake rate is determined in simulated multijet events. Both simulations correspond to 2018 data-taking conditions.

rate measurement. This avoids a bias from having a different admixture of trigger prescale factors in the numerator and denominator of a fake rate measurement bin. The trigger path to be used in a particular bin is selected by finding the minimum p_T^{cone} value that ensures any lepton with this p_T^{cone} value has a p_T over the trigger threshold. Using Equation 6.6, a requirement $p_T > x$ in a trigger path can be translated into $p_T^{\text{cone}} > x \cdot C / p_T^{\text{ratio, min}}$, with $p_T^{\text{ratio, min}}$ a cutoff on p_T^{ratio} imposed in the loose lepton selection. This is the cut tuned for ensuring closure in the fake rate method (Section 6.4.2). During the tuning of the loose selection, care is taken to ensure $p_T^{\text{ratio, min}}$ is large enough that $C / p_T^{\text{ratio, min}}$ does not exceed two. The possibility of very large cone corrections would severely limit the statistical precision in the fake rate measurement because more strongly prescaled triggers would have to be used. Using Equation 6.7 for the cone correction, one obtains a threshold $x / (1 + I_{\text{rel}}^{\text{max}} - I_{\text{rel}}^{\text{tight}})$ on p_T^{cone} given a p_T threshold x , following a similar reasoning to that outlined above. Here $I_{\text{rel}}^{\text{max}}$ is the maximum I_{rel} value a loose lepton is allowed to have.

Once the correct trigger path is chosen for each p_T^{cone} interval, the fake rate is measured in bins of p_T^{cone} and $|\eta|$ as the fraction of loose nonprompt leptons that pass the tight selection. The number of nonprompt leptons is the number of events in the sample of multijet enriched data, after subtracting the prompt lepton background. This subtraction is done in both the numerator and the denominator of the measurement. The prompt lepton contamination is estimated from simulation. To limit the uncertainty in the background subtraction we reduce the prompt lepton contamination with a number of kinematic requirements. Events with a single prompt lepton come mainly from W boson production. They typically have significant p_T^{miss} , and their M_T distribution peaks around m_W . Multijet events with a nonprompt lepton, on the other hand, tend to have low p_T^{miss} and M_T values because there are no high p_T neutrinos in the event, and $\vec{p}_T^{\text{miss}}$ does not balance the transverse lepton momentum vector. The prompt lepton contamination is minimized by requiring p_T^{miss} to be below 20 GeV, eliminating most W boson production events. To avoid a bias in the fake rate from the correlation between M_T and the lepton p_T , we define a new variable to replace M_T :

$$M_T^{\text{fix}} = \sqrt{2 \cdot 35 \text{ GeV} \cdot p_T^{\text{miss}} [1 - \cos(\Delta\phi)]}. \quad (6.8)$$

The M_T^{fix} variable retains most of the discriminating power of M_T because it still depends on p_T^{miss} and its direction relative to the lepton. However, it has no direct dependence on the lepton p_T anymore. In each p_T^{cone} and $|\eta|$ bin of the fake rate measurement, a binned χ^2 fit to the M_T^{fix} distribution determines the number of nonprompt leptons in the numerator and denominator. This fit is performed after the subtraction of the prompt background in each M_T^{fix} bin. Doing a fit as a function of M_T^{fix} results in a smaller uncertainty in the fake rate measurement than simply counting events because lower M_T^{fix} events are generally less polluted by prompt leptons. To illustrate the fake rate measurement, the M_T^{fix} distributions in the numerator and denominator of a single bin of the fake rate measurement are shown in Fig. 6.8.

The normalization of the expected number of prompt lepton events used in the subtraction is constrained from data. For each prescaled trigger path a fit to the distribution of M_T is performed for events with M_T values beyond m_W and $p_T^{\text{miss}} > 40$ GeV. The nonprompt lepton contribution is negligible in such events. The measured normalizations account for both the summed cross sections of the processes producing prompt leptons, as well as the trigger prescale factor. An example of the M_T distributions used to measure the normalization of the number of prompt leptons is shown in Fig. 6.9 for two prescaled trigger paths in 2017 collision data.

The misidentification rate of nonprompt leptons tends to become smaller with increasing p_T . Meanwhile the p_T spectrum of prompt leptons is harder than that of nonprompt leptons. Both of these effects increase the relative contribution of prompt leptons in the fake rate measurement at high p_T^{cone} values. When using a BDT based lepton selection strategy this is aggravated by the superior identification performance, and the situation becomes problematic. For p_T^{cone} values above 45 GeV, the prompt contamination in the numerator of the fake rate measurement is typically so large that the procedure described above no longer provides an accurate measurement of the fake rate. A tiny

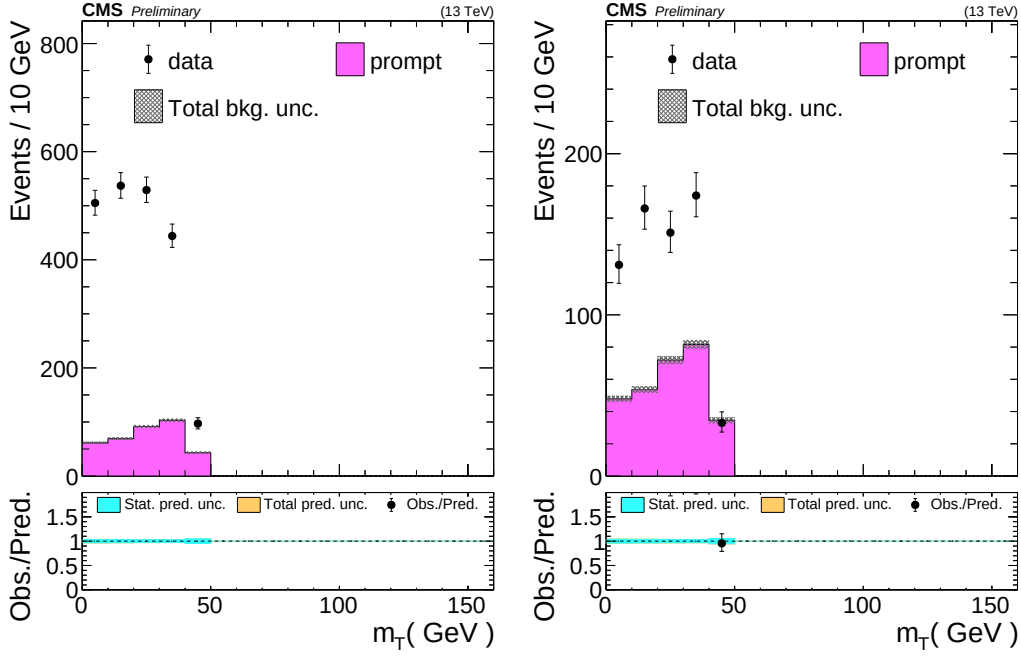


Figure 6.8: Number of events with a single muon satisfying $20 \text{ GeV} \leq p_T^{\text{cone}} < 30 \text{ GeV}$ and $2.1 \leq |\eta| < 2.4$ as a function of M_T^{fix} in 2018 collision data. The left (right) panel shows events with a single loose (tight) muon and acts as the denominator (numerator) of the fake rate measurement in the given p_T^{cone} and $|\eta|$ interval. The number of expected events with a prompt lepton is subtracted in both the numerator and denominator before fitting the M_T^{fix} distributions to determine the event yields. The lepton selection used is that defined in Chapter 9.

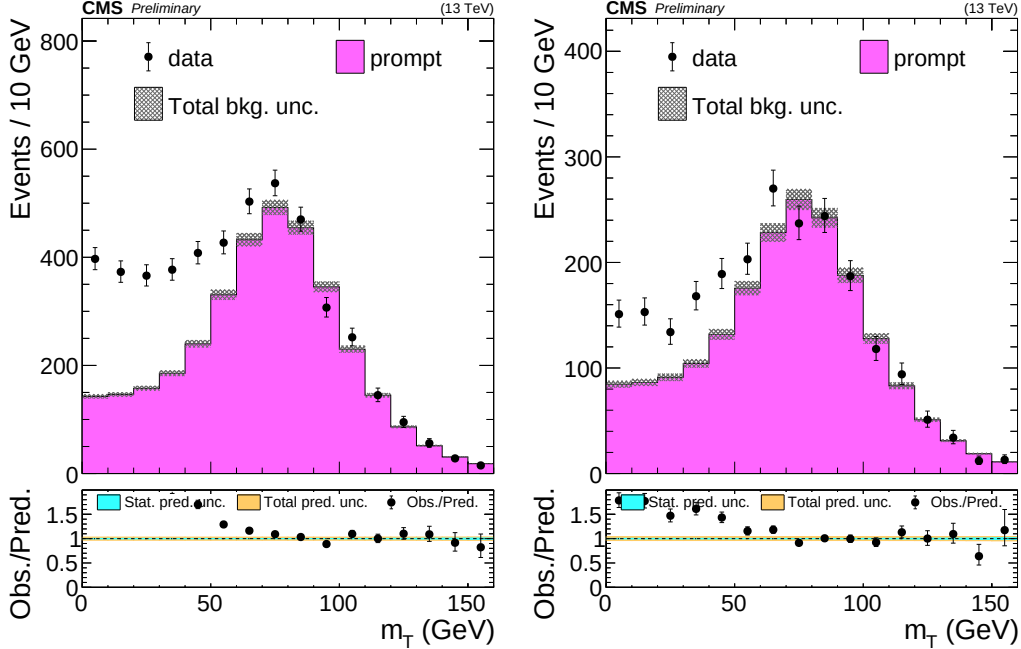


Figure 6.9: Post-fit expected number events with a prompt lepton and the total observed number of events as a function of M_T . The left (right) panel shows events that contain a single muon (electron) passing a prescaled trigger path requiring the presence of a muon with $p_T > 8 \text{ GeV}$ (electron with $p_T > 8 \text{ GeV}$ and a jet with $p_T > 30 \text{ GeV}$) in 2017 collision data. In low M_T bins a sizable number of QCD multijet events with a single nonprompt lepton is present, whereas the prompt lepton contribution is dominant at high M_T values. The normalization of the prompt lepton component is measured by fitting the event yields in the range 90–130 GeV.

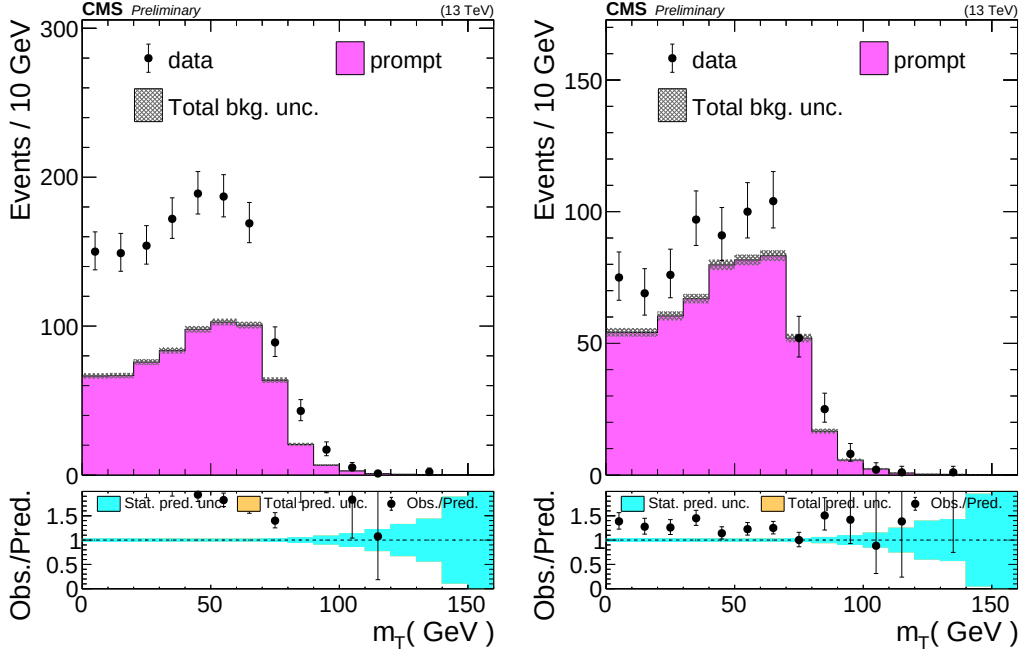


Figure 6.10: Number of events with a single electron satisfying $p_T^{\text{cone}} > 65 \text{ GeV}$ and $0.8 \leq |\eta| < 1.4$ as a function of M_T^{fix} in 2016 collision data. The left (right) panel shows events with a single loose (tight) electron and acts as the denominator (numerator) of the fake rate measurement in the given p_T^{cone} and $|\eta|$ interval. Because of the high electron p_T the number of prompt leptons is large in both the numerator and denominator, and small uncertainties in the estimation of this contamination can have a large impact on the measured fake rate. The lepton selection used is that defined in Chapter 9.

variation in the normalization of the prompt contamination can cause an enormous shift in the fake rate measurement. To give a concrete example, the numerator and denominator of the fake rate measurement for electrons passing the selection of Chapter 9 with $p_T^{\text{cone}} > 65 \text{ GeV}$ and $0.8 \leq |\eta| < 1.4$ in 2016 pp collision data are shown in Fig. 6.10.

In all analyses in this dissertation, the fake rate for p_T^{cone} values below 45 GeV is used in the evaluation of Equation 6.5 for loose-not-tight leptons in the application region that have p_T^{cone} values larger than 45 GeV . This avoids the usage of unreliable fake rate measurements at high p_T^{cone} , but brings along extrapolation uncertainties. These uncertainties are estimated from simulation. The impact of doing this extrapolation is tiny in each of the three analyses, because high- p_T nonprompt leptons that pass the tight selection are rare, as can be seen in Fig. 6.7.

In future analyses the fake rate measurement could be made more reliable at high p_T^{cone} by determining the normalizations of both the prompt and nonprompt contributions in each p_T^{cone} and $|\eta|$ bin from a template fit to M_T^{fix} . The essential difference with the subtraction procedure above is that the latter only uses information about the expected shape of the prompt background, while having no input information on the shape of the expected nonprompt lepton contribution. As a consequence, uncertainties in the multi-differential cross section of prompt lepton production in p_T^{cone} , $|\eta|$, p_T^{miss} , and M_T^{fix} directly affect the fake rate measurement. These effects could be partially offset by also using a shape template for the nonprompt lepton contribution, resulting in a more robust fake rate determination at high p_T . This procedure is applied in analyses carried out later than those presented in this dissertation, examples being [60] and [245]. One could expand upon this even further by simultaneously fitting templates in prompt and nonprompt enriched regions, for instance at low and high p_T^{miss} values. As lepton identification techniques keep improving, such approaches will become necessary.

6.5 Systematic uncertainties

The results of the analyses in this dissertation are evaluated by performing simultaneous binned maximum likelihood fits to the observed event yields in a number of search and control region. These fits follow the prescriptions from [246], and an introduction to such fitting procedures can be found in [247].

Sources of systematic uncertainty are taken into account as nuisance parameters in the fits. Each simulated process is affected by a number of experimental and theoretical uncertainties, and data-driven background estimation procedures have their own associated uncertainties. The experimental uncertainty sources common to all simulated processes are listed in Section 6.5.1, and the common theoretical uncertainty sources in Section 6.5.2.

Aside from those sources listed below, which affect all simulated processes, additional systematic uncertainties might be added to the predictions of specific processes. This is done to cover effects or discrepancies not accounted for by any general uncertainty source. Details of such uncertainty sources are analysis specific, and are given in the corresponding discussions in chapters 7 to 9.

The uncertainty sources described in this section do not directly affect the data-driven estimation of the nonprompt lepton background. A flat 30% normalization uncertainty is assigned to this background prediction in all analyses to account for potential biases in the fake rate method and the applicability of the fine tuned thresholds in real collision data. Additional p_T^{cone} extrapolation uncertainties are introduced in Chapter 9. Sideband events with at least one loose-not-tight lepton, from which the nonprompt background is estimated, however have a contamination from prompt leptons. This contamination is corrected for by using simulated events. All uncertainties that apply to simulated events are propagated in this subtraction. The data-driven nonprompt background prediction is thus also affected, to a smaller extent, by all of the uncertainty sources listed below.

6.5.1 Experimental uncertainties

A number of experimental uncertainties affect all simulated processes. These uncertainty sources are listed in this section. Their effects are evaluated on an event-by-event basis, either as statistical weights or by modifying kinematic quantities used in event selection and binning. Most of these uncertainty sources vary both the post-selection yields and the shapes of the event distributions for simulated processes. Exceptions are the uncertainty in the luminosity measurement in all analyses, and the uncertainty in the trigger efficiency measurement in chapters 8 and 9. These uncertainty sources only affect the normalization of simulated processes. The effects of each of the experimental uncertainty sources is taken to be fully correlated between all processes.

- **Jet energy scale:** Uncertainties in the successive jet energy scale correction steps (Section 5.3) are propagated to all jets. There are a number of sources of uncertainty, listed in [213], which are aggregated into a single global per jet uncertainty. Uncertainties in the jet energy scale also affect p_T^{miss} following Equation 5.5. All jets in an event and p_T^{miss} are varied in a correlated fashion, and the effect is propagated to the computation of all analysis variables that rely on jets and p_T^{miss} .
- **Jet energy resolution:** Similar to the jet energy scale, uncertainties in the jet energy resolution affect all jets in an event, as well as p_T^{miss} . The uncertainty propagation is done in an identical manner as for the jet energy scale variations, but the impact is much smaller. This uncertainty source is only considered in Chapter 9.
- **Unclustered energy:** Energy deposits in the ECAL and HCAL not clustered into reconstructed electrons, photons, or hadrons still contribute to p_T^{miss} [218]. The uncertainty in the energy scale of these unclustered deposits is propagated to p_T^{miss} and its associated quantities.
- **b-tagging efficiency:** The measured b-tagging efficiencies and misidentification rates (Section 5.4) are varied within their uncertainties. These variations are propagated as statistical

weights to each simulated event, both for the selection and veto of b jets. Variations in the tagging efficiency for b jets and the misidentification rate for c jets are considered simultaneously. Uncertainties in the light flavor jet misidentification rate are considered uncorrelated, and treated as a separate uncertainty source [220].

- Lepton identification efficiency:** Several sources of systematic uncertainty, in addition to statistical uncertainties, affect the lepton identification efficiency measurements. Lepton identification efficiencies are measured using the tag-and-probe technique in Drell-Yan production events (Section 6.1.3). In these measurements the number of prompt and nonprompt leptons are determined by fitting a signal and background component to the invariant mass distribution of $Z \rightarrow \ell\ell$ decays. The selection criteria of the tag lepton, and the functional forms used in the fit are varied. The effects of these variations on the efficiency measurement are considered systematic uncertainties. All systematic shifts in the fit are summed in quadrature to obtain a single systematic uncertainty value in each p_T and $|\eta|$ bin of the efficiency measurement. This uncertainty is subsequently propagated to analysis results on an event-by-event basis. Statistical uncertainties in the lepton identification efficiencies are also propagated as systematic uncertainties to the analyses using these efficiencies. Lepton identification efficiencies are measured in Drell-Yan events, but applied to events with a different topology. The lepton identification efficiency, and the corresponding data to simulation scale factors are not necessarily invariant under this extrapolation, necessitating further uncertainty sources. In the tZq search of Chapter 8 it is verified that the uncertainties in the scale factor measurement are large enough to cover extrapolation effects to a region enriched in top quarks. The supersymmetry signal studied in Chapter 9 often contains highly boosted leptons. An additional uncertainty is therefore derived to cover the difference between scale factors at low and high Z boson p_T values. The sterile neutrino search presented in Chapter 7 simply increases the systematic uncertainty in the lepton identification efficiency to a value that is large enough to cover extrapolation effects. More details are given in each of the analysis chapters.
- Trigger Efficiency:** All of the analyses presented here use a combination of single, two, and three lepton triggers (Section 6.2). The efficiency of the entire trigger cocktail is measured in a data set collected with orthogonal trigger paths that require a large amount of p_T^{miss} or hadronic activity. This data set contains a relatively meager amount of multilepton events. The statistical uncertainty in the trigger measurement at low lepton p_T values typically amounts to a few %, and this is propagated as a flat normalization uncertainty to all simulated events in each analysis. These uncertainties are considered uncorrelated between data-taking years because of their predominantly statistical nature. In Chapter 9 the difference between data to simulation scale factors for the trigger efficiencies in different data-taking years is taken as an additional correlated uncertainty source. In Chapter 7 such an uncertainty source is not applicable since only one data-taking year is analyzed. In Chapter 8 the trigger efficiency is effectively 100% across for events passing the analysis selection and the differences in the total trigger efficiency between the data-taking years are negligible.
- Luminosity measurement:** The integrated luminosity of the collected pp collision data set is used to normalize the yields of simulated processes (Equation 3.4). Any uncertainty in the determination of the integrated luminosity affects all signal and background yields predicted from simulation. The total uncertainty in the luminosity measurement of the 2016 [248] and 2018 [249] (2017 [250]) collision data set is 2.5 (2.3)%. This uncertainty is split up into several sources, some correlated between data-taking years and others uncorrelated. Taking into account these correlations, the uncertainty in the integrated luminosity of the full Run 2 pp collision data set is about 1.8%.
- Pileup:** The exact pileup distribution in data depends on the instantaneous luminosity of the LHC during data taking. This distribution was not exactly known a priori when most simulated

events were generated. This is mended by reweighting simulated events so that the distribution of the number of true interactions per bunch crossing matches that in collision data. The number of true interactions, which is not equal to the number of reconstructed vertices, is estimated in data by combining measurements of the instantaneous luminosity with the inelastic pp collision cross section (Equation 3.7). The inelastic pp cross section is varied up and down by its uncertainty of 4.6% [148], and the effect of this variation is propagated to the pileup weights applied to all simulated events.

6.5.2 Theoretical uncertainties

Theoretical uncertainties cause shifts in the predicted cross sections, selection acceptance, and kinematic shapes of simulated events. Shifts in the cross section are treated independently of acceptance and shape variations.

When the cross section of a process is determined from a control region in data, the theoretical uncertainty in the cross section of the process is replaced by a relatively large prior uncertainty. This uncertainty is subsequently constrained from the fit to data. The original input value of the uncertainty has no effect on the analysis unless there is a large mismatch between the observation and expectation in the control region, a situation which did not present itself in this dissertation. This procedure is applied to the normalization of the expected WZ and ZZ production yields in all presented analyses.

The theoretical uncertainty sources considered in the dissertation are given below:

- **QCD scale:** Distributions predicted by event generators depend on the QCD factorization and renormalization scales. The renormalization scale is introduced to remove ultraviolet divergences, while the factorization scale mends infrared divergences from massless particles radiating massless particles. Uncertainties from the choice of these scales are evaluated by simultaneously varying both of them up and down by a factor two. The resulting uncertainties generally become smaller the higher the perturbative order used in the calculation. This is one of the reasons NLO simulation samples are used wherever possible.
- **Proton PDF:** A limited knowledge of the proton PDFs is taken into account by using a set of NNPDF 3.0 (NNPDF 3.1) replicas in simulations of 2016 (2017 and 2018) collisions [251]. A set of 100 alternative PDFs are used, and the analysis results are evaluated with each one of them. The root mean square of all these variations is taken as the total uncertainty, following the recommendation in [251].
- **ISR and FSR scale:** In Chapter 8 the uncertainty in the parton shower simulation is evaluated for two of the most important processes. This is done by separately varying the renormalization scale for initial- and final-state radiation up and down by a factor of two. This results in two additional uncertainty sources. This uncertainty source is less relevant in the analyses of chapters 7 and 9 which do not have jets in the targeted final state.

Chapter 7

Search for heavy neutral leptons

“Neutrino physics is largely an art of learning a great deal by observing nothing.”

Haim Harari

7.1 Introduction

Several of the open questions in the SM can be addressed by postulating the existence of sterile neutrinos (N) that perceive neither electroweak nor strong interactions. Another common name for sterile neutrinos in the literature is heavy neutral leptons. Neutrino oscillations indicate that the SM neutrinos are in fact massive [18, 19, 252]. The current upper limits on neutrino masses are of the order of 1 eV [29, 107], indicating they are many orders of magnitude lighter than any other SM fermion. This raises questions on whether neutrinos also owe their mass to Yukawa couplings with the Higgs field in the way quarks and charged leptons do.

Small neutrino masses can be naturally explained with the seesaw mechanism, which requires the introduction of a heavy sterile neutrino [105, 109–118]. Such models can be extended to provide answers to other fundamental questions. The ν MSM [89, 102–104] explains both dark matter and the baryon asymmetry in the universe by introducing three generations of sterile neutrinos, while retaining the seesaw mechanism to explain the low SM neutrino masses. In this model the lightest sterile neutrino, N_1 , has a mass at the keV scale, and could be what makes up the dark matter observed in the universe [89]. The heavier sterile neutrinos, N_2 and N_3 , could meanwhile provide an explanation for the baryon asymmetry in the universe through neutrino oscillations [253, 254] or leptogenesis [104, 255–257]. These two heavy neutrinos might be accessible to LHC experiments. Giving potential answers to three fundamental open questions in the SM makes the ν MSM and other models including sterile neutrinos highly interesting candidates for searches targeting physics beyond the SM.

If they exist, sterile neutrinos can be produced through their mixing with SM neutrinos. The mixing between a sterile neutrino N and the SM neutrinos is dictated by the mixing matrix elements V_{eN} , $V_{\mu N}$, and $V_{\tau N}$, where $V_{\ell N}$ stands for the mixing matrix element between the SM neutrino of lepton flavor ℓ and N. The production cross section of N depends on the mass, m_N , and is directly proportional to $\sum_{\ell} |V_{\ell N}|^2$.

Searches for sterile neutrinos have been carried out at many experiments operating at a myriad of energy scales. If m_N is below 5 GeV, N can be produced in hadron decays. In this case very small mixing values can be probed in high intensity fixed target experiments [258–261] or b factories [262].

At higher m_N values, N can be produced in $W^{(*)}$ or $Z^{(*)}$ boson decays, where the gauge bosons have to be virtual if N exceeds them in mass. Searches in Z boson decays were carried out at LEP [263–266], independently setting limits on the mixing of N with all SM neutrino flavors for $m_N < 40$ GeV. At

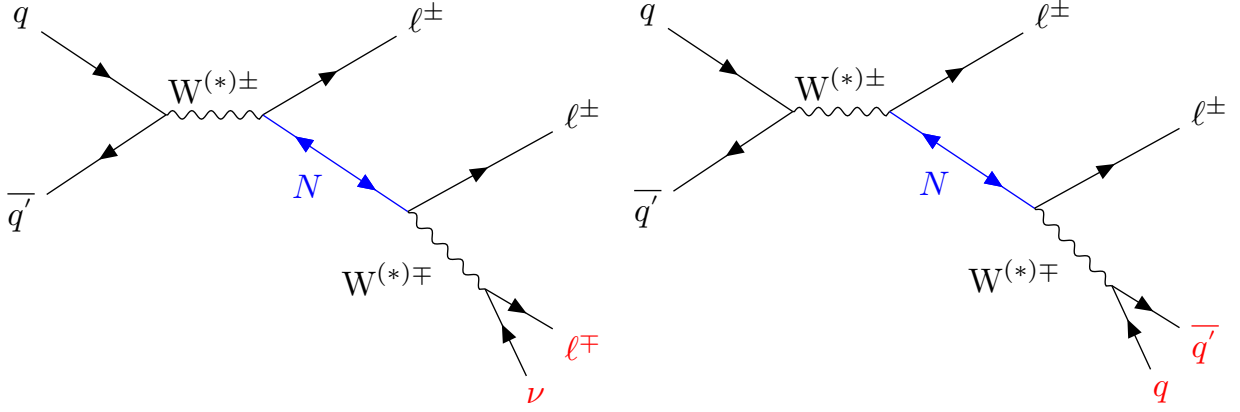


Figure 7.1: Feynman diagrams of N production in a $W^{(*)}$ boson decay. The decay of N goes through its couplings with the SM neutrinos, and results in another lepton and $W^{(*)}$ boson. Depending on m_N either the first or the second W boson is virtual. The final $W^{(*)}$ boson either decays leptonically, resulting in an event with three leptons (left), the subject of this work, or decays hadronically (right).

the LHC, $W^{(*)}$ boson decays are the better probe for N production because of the higher W boson production cross section (see Fig. 6.2), the ability to trigger on the lepton from the $W^{(*)}$ boson decay, and reduced backgrounds when requiring an additional lepton. Before the publication of the work presented here, searches for sterile neutrinos were already carried out by both ATLAS [267–269] and CMS [270–274] using 7 and 8 TeV data. An overview of the constraints on sterile neutrino production prior to the publication of the presented analysis can be found in [275].

Here a search for a heavy sterile neutrino of Majorana nature produced in $W^{(*)}$ boson decays is presented. Shortly after production, the sterile neutrino itself decays to a lepton and another $W^{(*)}$ boson. Depending on the decay mode of the final $W^{(*)}$ boson the final state of the event either consists of three leptons and p_T^{miss} , or two leptons and two jets. Feynman diagrams corresponding to the described production mechanism are drawn in Fig. 7.1. Previous LHC analyses always targeted hadronic decays of the final $W^{(*)}$ boson. This decay mode has the advantage of having a branching fraction that is about a factor two higher than that of the leptonic decay mode. Additionally, hadronic decays allow for a full reconstruction of the N mass, because all of its primary decay products are visible. The reconstructed mass is a powerful tool to discriminate signal and background events, and for the interpretation of a potential excess of events over the SM hypothesis. The hadronic channel however suffers from much larger backgrounds than the leptonic channel, and needs to impose significant p_T thresholds in the jet reconstruction. The jet energy scale (Section 5.3) is only well calibrated down to p_T values of about 20 GeV. This makes it hard to probe the existence of an N of relatively low mass in the hadronic decay channel.

The work exhibited here uses 35.9 fb^{-1} of pp collision data collected with the CMS detector throughout 2016, at a center of mass energy of 13 TeV. It is the first LHC search for N production in the three lepton final state. Leptons can be reconstructed and identified at low p_T , and the search result is the first probe of the region $m_N < 40 \text{ GeV}$ at a hadron collider. Because of the 13 TeV pp collision data set, this search is also the first one to reach sensitivity to N production at m_N values higher than 500 GeV. The search explicitly targets prompt decays of N . If m_N and $|V_{\ell N}|^2$ are sufficiently small, sterile neutrinos get a significant lifetime and travel nonnegligible distances through the detector before decaying. The leptons from the decay of N then become displaced with respect to the primary pp interaction vertex, requiring substantial modifications to the lepton identification and background prediction methods introduced in Chapter 6. The reconstruction of displaced N decays is beyond the scope of this particular search. A more in-depth discussion of the N lifetime is given in Section 7.8 when covering the interpretation of the search results.

After the publication of the sterile neutrino search presented in this dissertation [6], several other

searches for sterile neutrinos have been published. Searches in the same final state were performed by ATLAS [276] and CMS [277] several years after the publication of the presented work. The ATLAS search uses displaced leptons in addition to prompt leptons to gain sensitivity at low m_N values. The CMS search is fully dedicated to long-lived sterile neutrinos. More information about this search is given in the conclusions of this chapter. A CMS result in the dilepton final state using 13 TeV data improved upon the limits of our search at high m_N values, especially for mixing between sterile and muon neutrinos [278]. This improvement is largely driven by the larger branching fraction of the hadronic channel. Recently LHCb carried out a search for W^+ boson decays to sterile neutrinos in dilepton events [279]. The LHCb analysis set exclusion limits for N masses between 5 and 50 GeV that are about an order of magnitude worse than the limits set earlier by the work in this dissertation.

The rest of this chapter is organized as follows: First the concrete signal model used in the search is introduced in Section 7.2. Sections 7.3 and 7.4 respectively discuss the event selection and the strategy for extracting the N production signal from the selected events. The background estimation procedures are laid out in Section 7.5. Systematic uncertainties in the background prediction and the modeling of the signal are covered in Section 7.6. Observed yields in data and their interpretation in terms of N production are presented in sections 7.7 and 7.8 respectively. A summary of the search and an outlook are given in Section 7.10.

7.2 Signal model

In order to design an effective search for sterile neutrinos we need to have a concrete signal model to optimize the search for and we must simulate it. The model is based on that introduced in [280]. In this model the SM is extended with up to three right-handed neutrinos that are singlets under all SM gauge groups. In other words, the new neutrinos are sterile. For this search we consider a single sterile neutrino N , and choose to independently examine the mixing of N with ν_e and ν_μ . Cases in which N mixes with both ν_e and ν_μ were initially studied but later discarded because the interpretation of the results depends on the relative sizes of $V_{\mu N}$ and V_{eN} . Constraints derived on the basis of a model in which N couples to a single SM neutrino flavor are always valid, regardless of the other elements in the mixing matrix. A mixing between N and ν_τ is not studied in this search, though the results themselves do inherently have sensitivity to this mixing because of leptonic τ decays.

The practical implementation of the model is given in the Universal FeynRules Output format [281] in [282, 283]. The model is deployed in the MADGRAPH5_aMC@NLO generator to produce simulated pp collisions that result in N production. Simulated samples are generated at different m_N values, all at NLO precision in perturbative QCD. The simulation includes the charged-current Drell-Yan, gluon fusion, and $W\gamma$ fusion N production mechanisms. The LO diagrams for these N production modes are shown in Fig. 7.2. Charged-current Drell-Yan production of an N is the dominant production mode in most of the m_N mass range the search is sensitive to. At the LHC gluon fusion production is substantially more rare than Drell-Yan production across all m_N values within reach. At future colliders that have a higher center of mass energy than the LHC gluon fusion will become the dominant production mode. The $W\gamma$ fusion production channel only becomes relevant when m_N exceeds about 600 GeV, and is the dominant production mechanism at very high m_N values, upwards of about 1 TeV [284].

The modeling of the parton shower and hadronization in the simulated N production samples are taken care of by PYTHIA 8.212 [285] with the CUETP8M1 underlying event tune [162, 163], similar to what is done for the generation of SM events. As for all simulations in this dissertation at NLO precision in QCD, the FxFx matching scheme is used to remove the double counting of partons between MADGRAPH5_aMC@NLO and PYTHIA. The Drell-Yan and gluon fusion N production modes use the NNPDF 3.0 PDF set [167], also used in the simulation of SM processes under 2016 data-taking conditions. In the case of $W\gamma$ fusion production, the specialized LUXQED_PLUS_PDF4LHC15_NNLO_100 PDF set is employed [286].

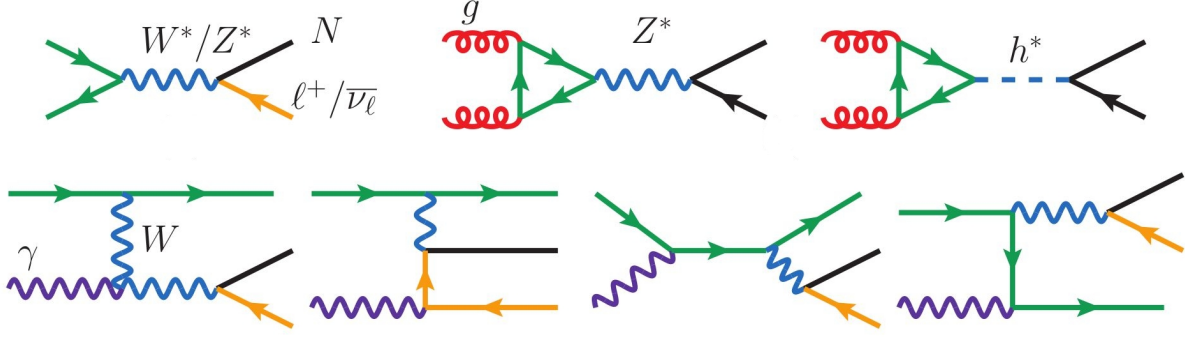


Figure 7.2: Leading order Feynman diagrams for the dominant N production modes at the LHC. The top row shows charged-current Drell-Yan production on the left and gluon fusion production in the middle and on the right. The bottom row shows the Feynman diagrams for $W\gamma$ fusion production. The sterile neutrino is always drawn as a solid black line [284].

7.3 Event and object selection

This section outlines the event selection strategy of the analysis. We target signal events with three leptons, electrons or muons, and little or no hadronic activity. The chosen lepton selection criteria are explained in Section 7.3.1. The selection of jets, which are used to reduce the background rate, is the topic of Section 7.3.2. The selection outlined throughout this section will be referred to as the baseline selection in the rest of the chapter. To be considered for analysis all events must pass the baseline selection, without exception. Some additional selection criteria used to further isolate signal events in particular m_N regions are introduced in Section 7.4.

7.3.1 Lepton selection and triggering

The p_T spectra of the three leptons in signal events are an important consideration in the design of the analysis. The search aims to probe the existence of sterile neutrinos with masses down to just a few GeV, for which one or more of the leptons have compressed p_T spectra. The p_T spectra of the three leptons, ordered by p_T , are shown in Fig. 7.3 and Fig. 7.4 for different m_N hypotheses. These figures show direct predictions by the physics generator, without taking into account detector and reconstruction effects.

If m_N is smaller than m_W , N is almost always produced in the decay of a real W boson. If m_N is many times lower than m_W , the two leptons originating from the N decay are generally very soft, while the third lepton has approximately the same p_T spectrum as a lepton from a $W \rightarrow \ell\nu$ decay, and is comparatively hard (Fig. 7.3 left). To be able to select the two soft leptons from the N decay, p_T thresholds as low as feasible must be used in the selection of leptons. When m_N approaches m_W , the p_T spectrum of the lepton from the W boson decay becomes softer because more of the mass energy of the W boson is needed to produce the sterile neutrino. In this m_N regime, signal events are characterized by three relatively soft leptons with similar p_T spectra (Fig. 7.3 right). When m_N is significantly higher than m_W , the p_T spectra of the leptons from the decay of the sterile neutrino harden. In this case N production requires a highly off-shell W^* boson, resulting in a third lepton that is soft compared to the N decay products (Fig. 7.4). The harder the lepton originating directly from the W^* boson decay is at a given m_N , the more off-shell its parent W^* boson must have been.

To achieve a signal efficiency as high as possible, low p_T leptons must be selected while still retaining a good signal purity. Muons are required to have $p_T > 5$ GeV and fall within the longitudinal acceptance of the muon system, $|\eta| < 2.4$. Electrons, on the other hand, are required to have $p_T > 10$ GeV and be inside the tracking detector acceptance of $|\eta| < 2.5$. Selecting lower p_T electrons and muons is not feasible due to a rapidly increasing background yield with lowering lepton p_T , mainly

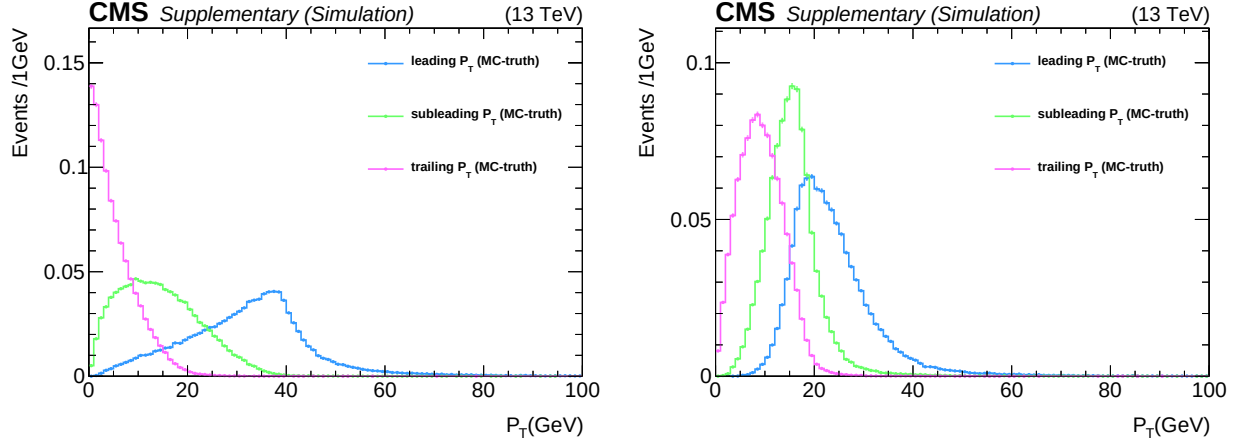


Figure 7.3: Ordered transverse momenta of the three leptons in a $W \rightarrow N\ell \rightarrow \ell\ell\ell\nu$ event with $m_N = 5$ GeV (left) and $m_N = 60$ GeV (right). Direct predictions from the event generator, excluding detector and reconstruction effects, are shown.

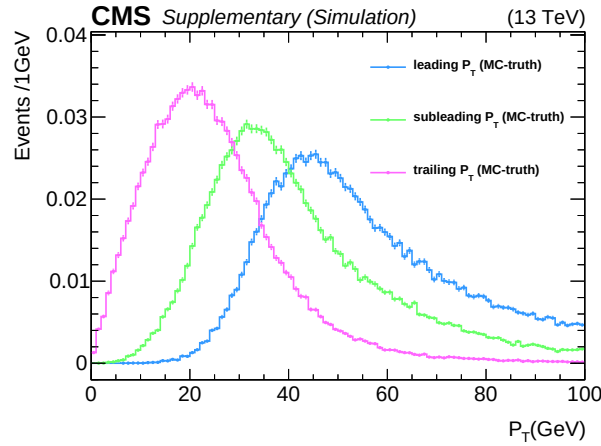


Figure 7.4: Ordered transverse momenta of the three leptons in a $W^* \rightarrow N\ell \rightarrow \ell\ell\ell\nu$ event with $m_N = 130$ GeV. Direct predictions from the event generator, excluding detector and reconstruction effects, are shown.

from nonprompt leptons.

The variables used in the lepton selection, described below, are all introduced in Section 6.1.1. All muons must pass the medium muon identification requirements, a set of quality criteria on the matching between tracks in the muon system and the inner tracking detector. Electrons that originate from photon conversions are vetoed by rejecting those candidates that have missing hits in any tracker layer, as well as those whose track is matched to a secondary conversion vertex [202]. A set of loose quality criteria on the shape and features of the ECAL clusters of all electrons is applied to match the requirements set by the online trigger selection.

The analysis targets promptly decaying sterile neutrinos, meaning those with a relatively high mass or mixing parameter. Selected electrons and muons must be consistent with this hypothesis and are rejected if their transverse (longitudinal) impact parameter with respect to the leading primary vertex exceeds 0.5 (1.0) mm, or if the significance of their three dimensional impact parameter is greater than 4.

A strong discriminant for rejecting nonprompt leptons is I_{rel} . For both electrons and muons a fixed cone size of $\Delta R < 0.3$ around the lepton is used for the I_{rel} computation. The contribution of neutral pileup is corrected for with the effective-area based method. To enter the nominal analysis selection both electrons and muons must pass $I_{\text{rel}} < 0.1$, which rejects most nonprompt leptons. Leptons passing all the above requirements are considered tight following the naming convention of Section 6.4.

The prediction of the nonprompt lepton background that remains after the tight lepton selection is made with the fake rate method (Section 6.4). To apply this method a second lepton selection definition optimized for the nonprompt background prediction, called loose, is defined. To qualify for the loose selection, electrons and muons must have $I_{\text{rel}} < 0.6$. Other than the I_{rel} threshold, loose leptons must also satisfy all the criteria introduced above for tight leptons.

In addition to the requirements specified above, the electron selection uses a multivariate discriminant based on the ECAL shower shape and tracking detector measurements of electron candidates, the electron MVA (Section 6.1.1). Tight electrons are required to pass a p_T -dependent cut on the electron MVA that aims at having a 90% efficiency for prompt electrons while rejecting most hadronic fakes. Loose electrons are required to pass a much less stringent cut on the multivariate discriminant, fine tuned to facilitate the nonprompt lepton background prediction with the fake rate method.

Events are selected for analysis if they have exactly three tight leptons. Events with a fourth loose lepton are vetoed. Three lepton events in which one or more of the leptons fail the tight selection but pass the loose selection are retained for the prediction of the nonprompt lepton background.

Each analysis presented in this dissertation uses a mix of triggers that select either one, two, or three light leptons (Section 6.2). The search for sterile neutrinos is the one that most crucially depends on the entire trigger cocktail due to the compressed p_T spectra of leptons in signal events at low m_N . Single lepton triggers are efficient for the selection of events in which an N with a mass of just a few GeV is produced. Such events usually have one hard lepton as previously explained. Triggers selecting multiple leptons have much less stringent requirements on the leading lepton, which allows them to efficiently select signal events in which m_N is close to m_W . These trigger paths have the drawback of imposing p_T cuts on several or all leptons in the event. Signal events at high m_N values tend to have leptons of sufficient p_T to be able to pass several of the used trigger paths.

Expected signal and background event yields can not be accurately predicted for events in which no trigger path is operating at high efficiency. To ensure that the requirements of at least one trigger path are met in all selected events, a number of additional p_T requirements are imposed on the selected leptons. Given the compressed p_T spectra expected in low-mass N production, these thresholds are kept as low as allowed by the set of trigger paths that was used to collect the data set. The cuts are listed below, and depend on the lepton flavor composition of the event:

- $\mu\mu\mu$: The leading (subleading) muon must have $p_T > 15(10)$ GeV.
- eee : The leading electron is required to have $p_T > 30$ GeV, or the leading and subleading electrons must respectively satisfy $p_T > 19$ GeV and $p_T > 15$ GeV.

Table 7.1: Overview of the lepton p_T thresholds imposed in the analysis and the lepton p_T region in which each used trigger path provides sensitivity. We only specify trigger paths by the lepton flavor combination they target. Different trigger paths that target the same combination of lepton flavors, but with altering p_T and identification criteria, are aggregated together.

Search region	Lepton flavors ℓ^{leading} $\ell^{\text{subleading}}$ ℓ^{trailing}	Selected p_T range (GeV)			Sensitive trigger paths
		p_T^{leading}	$p_T^{\text{subleading}}$	p_T^{trailing}	
Low mass	$ee\mu$	30–55	> 15	> 5	$1e$
		15–30	> 15	> 8	$2e1\mu$
		23–30	10–15	> 8	$1e1\mu$
		25–30	> 15	5–8	$2e$
	$e\mu e, \mu ee$	30–55	> 15	> 10	$1e$ or 1μ
		15–30	> 15	> 15	$2e1\mu$
		23–30	> 10	10–15	$1e1\mu$
	$\mu\mu e$	30–55	> 15	> 10	1μ
		15–30	> 10	> 10	$1e2\mu$
	$\mu e\mu, e\mu\mu$	30–55	> 10	> 10	$1e$ or 1μ
		15–30	> 10	> 9	$1e2\mu$
		23–30	> 10	5–9	$1e1\mu$
High mass	All flavors	> 55	> 15	> 10	All paths

- $ee\mu$: If the trailing lepton is an electron with $p_T < 15$ GeV, the leading lepton must have $p_T > 23$ GeV. If the trailing lepton is a muon with $p_T < 8$ GeV, the leading (subleading) electron is required to satisfy $p_T > 25(15)$ GeV. If the muon is trailing with $p_T > 8$ GeV, either the leading electron must have $p_T > 23$ GeV or the subleading one must have $p_T > 15$ GeV.
- $\mu\mu e$: The leading (subleading) lepton must pass $p_T > 15(10)$ GeV. If the trailing lepton is a muon with $p_T < 9$ GeV, the leading lepton must have $p_T > 23$ GeV.

An overview of the p_T thresholds and which trigger reaches its efficiency plateau at different lepton p_T values is given in Table 7.1. The table makes the distinction between the “high-mass” and “low-mass” selection which depend on the lepton p_T values and are defined in Section 7.10. When targeting signals at high m_N values, the p_T thresholds are increased to reduce the background yield.

The efficiency of a multilepton trigger cocktail is generally measured in an unbiased data set collected by triggering on p_T^{miss} or the hadronic activity in the event (Section 6.2). This method suffers from low event rates and while it can give a global measurement of the average trigger efficiency, it fails to measure the trigger efficiency as a function of lepton p_T values. Given the low p_T requirements in this particular search this is problematic because we can not assume data to simulation differences in the trigger efficiency to be independent of the p_T values of the three leptons. The problem does not present itself in the analyses discussed in later chapters where many of the trigger paths operate in parallel across the selected lepton p_T spectrum.

The tag-and-probe method, introduced in Section 6.1.3 for the measurement of lepton identification efficiencies, offers a solution. It allows for a measurement of the efficiency with which each individual lepton is selected in a given trigger path that selects multiple leptons. A trigger path is split into separate filters that each select a lepton. For all but one of the per lepton filters in a multilepton trigger path, tag leptons are selected that are certain to pass the requirements imposed by the chosen filters. The efficiency of the remaining lepton filter can subsequently be measured. For example, events with three leptons can be selected, and required to pass a trigger path selecting two leptons. The two leptons on which the event is triggered are also guaranteed to pass two lepton filters of a three lepton trigger if correct triggers are chosen. The efficiency of the remaining lepton filter of the three lepton trigger with respect to the third lepton in the event can now be measured. Such measurements were available for all lepton filters of the triggers used in this analysis. Unfortunately

it is not straightforward to translate these per lepton efficiencies into a global efficiency of the trigger cocktail, in which many trigger paths are operating simultaneously. The measured efficiencies can however be used to put a p_T -dependent upper limit on discrepancies between data and simulation in the total trigger efficiency. Data to simulation differences are maximal when only a single trigger path has reached its efficiency plateau. As lepton p_T values rise, the redundancy in the trigger selection increases. This pushes the total efficiency closer to 100%, and generally reduces data to simulation discrepancies. The mismatch between data and simulation in the total trigger efficiency is maximal when all three leptons are relatively soft. Only three lepton trigger paths actively trigger on such events, and the maximum of the data to simulation discrepancy in the total trigger efficiency is the sum of the discrepancies in the efficiency of the three lepton filters. As soon as the leading lepton has a high enough p_T to pass a single lepton trigger, the maximal uncertainty reduces to the difference between data and simulation in the efficiency of this trigger. To cover the effects described above, a flat normalization uncertainty of 5% is assigned to events where the leading lepton p_T is below 30 GeV. This uncertainty corresponds to the sum of the data to simulation efficiency differences for three lepton filters, assuming the differences to be fully correlated. When the p_T of the leading lepton exceeds 30 GeV, the normalization uncertainty is reduced to 2% because the single lepton triggers are active in this regime. These uncertainties are conservative, especially in the region where the 5% uncertainty is applied, because many of the corresponding events will also pass one of the dilepton triggers. No data to simulation scale factors are applied for the trigger efficiency since the lepton p_T -dependent efficiency of the total trigger cocktail can not be measured with sufficient precision.

7.3.2 Jet selection

Jets are reconstructed with the anti- k_T algorithm (Section 5.2) and selected if they have $p_T > 25$ GeV and $|\eta| < 2.4$. Jet momenta are corrected for pileup and the nonuniformity and nonlinearity in the detector response (Section 5.3). Jets likely to originate from anomalous energy deposits in the detector are rejected.

The most important background in this analysis comes from events with one or more nonprompt leptons. One of the largest contributions to this background comes from $t\bar{t}$ production resulting in two prompt leptons and one nonprompt lepton. To reduce this background as much as possible, events with at least one b-tagged jet are vetoed. This is the only use case for jets in this search. Jets are considered b-tagged if they pass the loose working point of the CSVv2 algorithm (Section 5.4). This working point has an efficiency of about 80% for tagging genuine b jets, while it has a substantial misidentification rate of 10% for light-quark and gluon jets, and 40% for c jets [220]. Jets overlapping with a lepton are not removed when applying the b jet veto. Not doing the overlap removal increases the signal to background ratio by further reducing the $t\bar{t}$ background, at the cost of a small drop in signal efficiency.

7.4 Signal extraction

The aim of the search is to provide maximal sensitivity to N production at both low and high m_N values, and when N is produced by mixing with either ν_μ or ν_e . This is achieved by categorizing events into a set of orthogonal search region bins that aim to isolate signal events from the background, and by applying additional selection criteria that reduce certain background sources.

If N is of Majorana nature, its decay can violate lepton number. In the leptonic N decay channel lepton flavor violation can result in events with three leptons, none of them forming an OSSF pair. Events of this type are extremely rare in the SM. This is illustrated in Fig. 7.5 which shows the expected and observed yields of events passing the baseline selection as a function of the invariant mass of the three lepton system ($M_{3\ell}$), separately for events with and without an OSSF lepton pair. The set of events in which an OSSF pair is present contains relatively large background contributions from a host of SM processes. On the other hand, there are orders of magnitude fewer events without

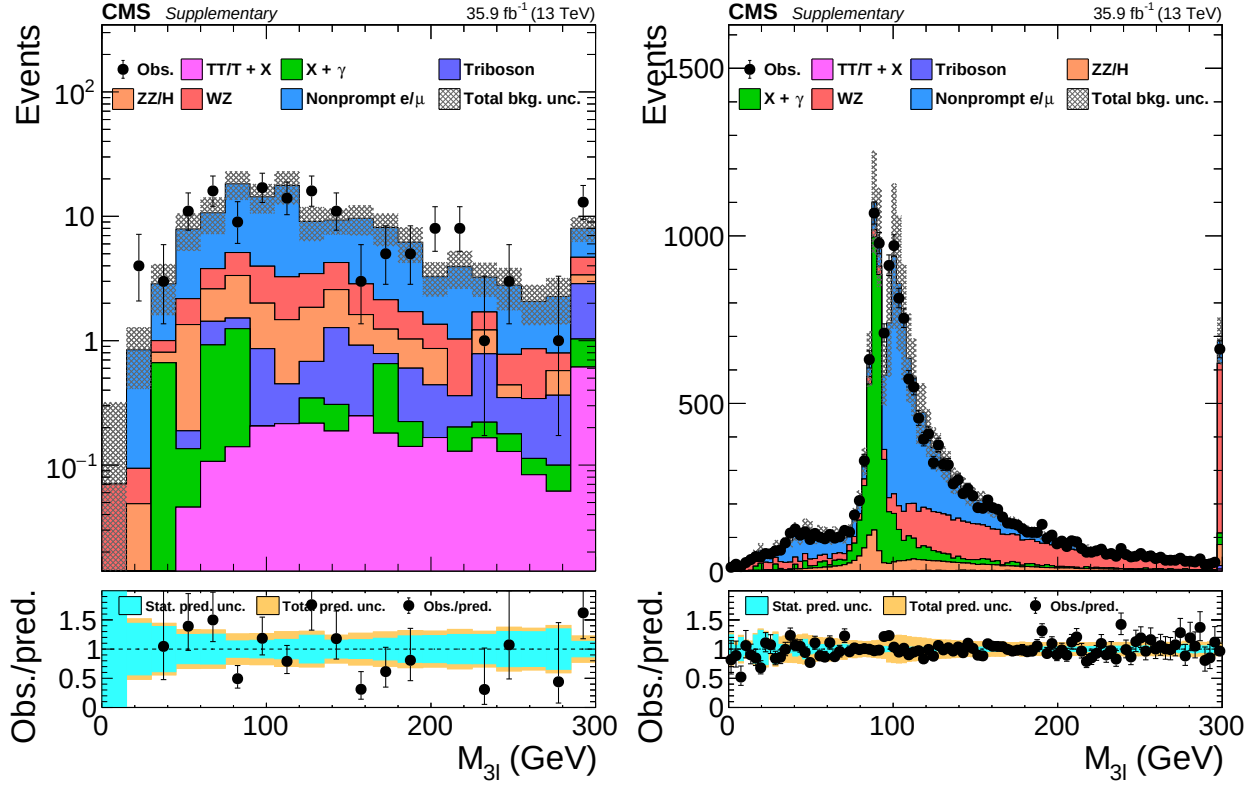


Figure 7.5: Observed and expected event yields after the baseline selection as a function of invariant mass of the three lepton system. Three lepton events without (with) an OSSF lepton pair are shown on the left (right).

an OSSF pair, with the only sizable contributions coming from nonprompt leptons, leptonic τ decays, and lepton sign mismeasurements. We exploit the potential Majorana nature of N and split events into two categories, differing by whether they contain an OSSF lepton pair or not. Events without an OSSF pair provide the bulk of the search sensitivity since we explicitly assume our signal to be a sterile Majorana neutrino.

Low-mass N production results in events with a relatively soft leading lepton compared to SM events, in particular when m_N is close to m_W . This can be seen clearly in Fig. 7.6 where the number of events that pass the baseline selection is shown as a function of the leading lepton p_T . High-mass N production, on the other hand, typically yields a hard leading lepton. For this reason p_T^{leading} is used to split the search into two distinct regions: The low-mass region, which satisfies $p_T^{\text{leading}} < 55$ GeV, and the high-mass region where $p_T^{\text{leading}} > 55$ GeV holds. The low-mass region itself is further subdivided into $p_T^{\text{leading}} < 30$ GeV and $30 < p_T^{\text{leading}} < 55$ GeV bins to separate signals in which m_N is similar to m_W from those at the lowest m_N values.

In the leptonic N decay mode m_N can not be fully reconstructed because the neutrino from the N decay eludes detection. Leptons from the decay of a low-mass N are usually collimated and form a small invariant mass. In this type of event the invariant mass of the opposite sign lepton pair of lowest mass ($M_{2\ell\text{OS}}^{\text{min}}$) is almost always formed from the leptons stemming from the N decay. This makes $M_{2\ell\text{OS}}^{\text{min}}$ a good proxy for m_N , which can be used to discriminate signal and background events and to isolate different m_N hypotheses. When m_N is large, $M_{2\ell\text{OS}}^{\text{min}}$ is no longer exclusively formed from the N decay products, but is still empirically found to be powerful for distinguishing signal and background events. The transverse mass of the lepton not entering the $M_{2\ell\text{OS}}^{\text{min}}$ computation and p_T^{miss} (M_T)¹ is

¹In this analysis M_T is computed using the lepton not included in the opposite sign lepton pair of minimum mass. This differs from the definition used in chapters 8 and 9 where the lepton not included in the best reconstructed Z boson candidate is used.

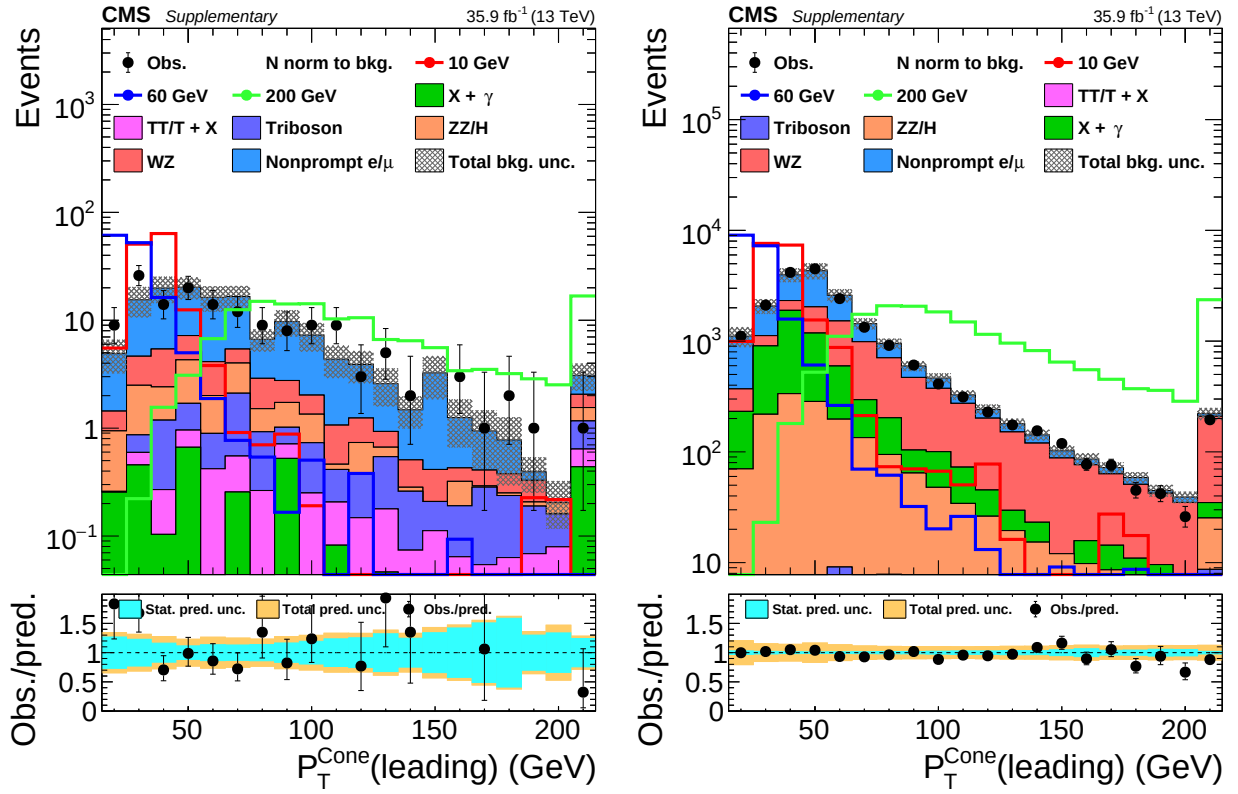


Figure 7.6: Observed and expected event yields after the baseline selection as a function of the p_T of the leading lepton. Three lepton events without (with) an OSSF lepton pair are shown on the left (right). The red, blue, and green lines show the expected distributions for N production at different m_N values, with yields normalized to the total SM background yield.

Table 7.2: Search region definitions in the low-mass region (AX, BX). Events are considered part of the low-mass region if they satisfy $p_T^{\text{leading}} < 55$ GeV. Only events without an OSSF pair are considered in the low-mass region, and the requirements $M_{3\ell} < 80$ GeV and $p_T^{\text{miss}} < 75$ GeV are applied on top of the baseline selection.

p_T^{leading} (GeV)	$M_{2\ell\text{OS}}^{\text{min}}$ (GeV)			
	< 10	10–20	20–30	> 30
< 30	A1	A2	A3	A4
30–55	B1	B2	B3	B4

also found to be a powerful discriminant between N production and SM events. The distributions of $M_{2\ell\text{OS}}^{\text{min}}$ and M_T are shown in Fig. 7.7 for events passing the baseline selection.

A combination of the discriminating variables introduced above is used to define the set of search regions in which selected events are binned for maximum sensitivity to N production. Different variable combinations and binning schemes were explored and the best combination of search region variables, ranked in terms of expected exclusion limits, is used. After selecting suitable variables for event binning, a large amount of bins is defined. These bins are iteratively merged until no bins contain significantly less than one expected background event. Having sufficient expected background events in each search region bin ensures a robust estimation of the background. Meanwhile a finer binning generally results in better search results.

In the low-mass region only events without an OSSF lepton pair are kept. Search regions for events with an OSSF pair are also optimized, but these are found not to add significant sensitivity to the search due to an overwhelming background from Drell-Yan production with an associated nonprompt lepton, and photon conversions. The low-mass region targets N production in decays of real W bosons, so we impose the requirement $M_{3\ell} < 80$ GeV. This is effectively guaranteed to be satisfied by signal events. We additionally require $p_T^{\text{miss}} < 75$ GeV to reduce the background from $t\bar{t}$ production that remains after vetoing the presence of b-tagged jets. The selected events are categorized according to p_T^{leading} and $M_{2\ell\text{OS}}^{\text{min}}$. The search region definitions are listed in Table 7.2.

Production of a very massive sterile neutrino results in energetic events that are relatively straightforward to distinguish from SM events. For this reason high-mass region events with an OSSF lepton pair have a significant signal sensitivity. These events have the advantage that $W^{(*)}$ -mediated N decays are combinatorically three times more likely to result in an OSSF pair than not, despite the Majorana nature of N. In addition, N decays can be mediated by $Z^{(*)}$ bosons, which invariably results in OSSF lepton pairs. Events with an OSSF pair are therefore retained in the high-mass region. Those without an OSSF pair however still provide more sensitivity because of the much lower background rates. Events are subdivided into several $M_{2\ell\text{OS}}^{\text{min}}$ bins to isolate different m_N mass hypotheses. Each $M_{2\ell\text{OS}}^{\text{min}}$ bin is further divided in terms of M_T to separate the signal from the remaining background. Events in which $M_{3\ell}$ is below 100 GeV are treated separately to further segregate certain background sources. Events in which an OSSF lepton pair is present must satisfy the additional criteria $|M_{\ell\ell} - m_Z| > 15$ GeV and $|M_{3\ell} - m_Z| > 15$ GeV, where $M_{\ell\ell}$ stands for the invariant mass of the OSSF pair of closest mass to m_Z . These requirements are introduced to respectively reduce the background contributions from WZ and $Z\gamma^{(*)}$ production. The p_T requirement on the subleading (trailing) lepton is increased to 15 (10) GeV for all high-mass region events. Increased lepton p_T thresholds significantly reduce the background rate from nonprompt leptons, at a minimal cost to the signal efficiency. The definitions of all high-mass search regions for events without (with) an OSSF lepton pair are given in Table 7.3 (7.4).

The results of the analysis are to be interpreted separately in terms of $|V_{eN}|^2$ and $|V_{\mu N}|^2$, as stated in Section 7.2. If V_{eN} is the only nonzero mixing parameter, production of N in a $W^{(*)}$ boson decay is always in association with an electron: $W^{(*)} \rightarrow e\nu_e^* \rightarrow eN$. The subsequent decay of N also goes through its mixing with ν_e , resulting in at least one more electron: $N \rightarrow eW^{(*)}$. The second electron can take on either sign because of the Majorana nature of N. So if V_{eN} is the only nonzero mixing matrix element, signal events always contain at least two electrons. Equivalently, if $V_{\mu N}$ is the sole

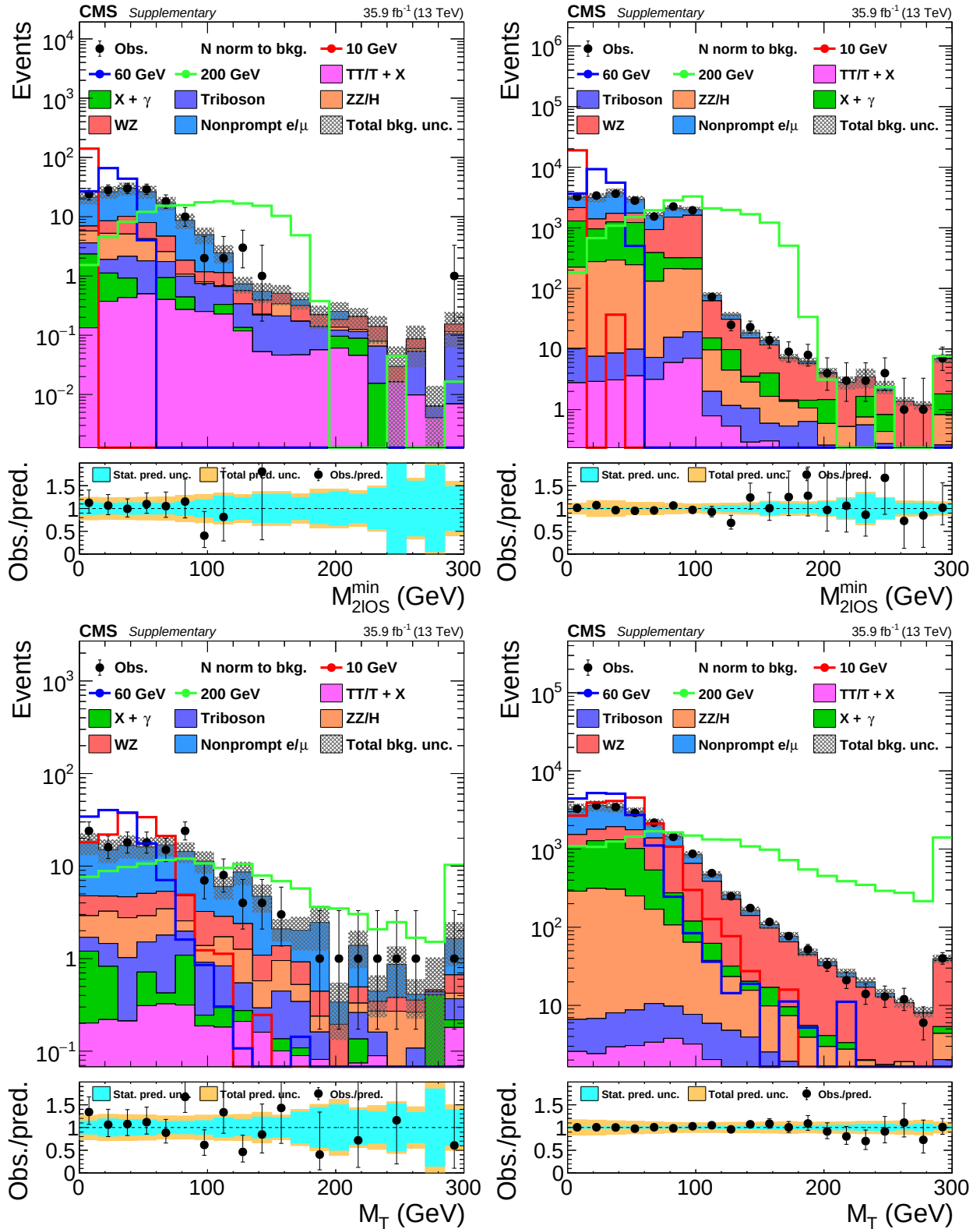


Figure 7.7: Observed and expected event yields after the baseline selection as a function of $M_{2LOS}^{\min}$ (top) and M_T (bottom). Three lepton events without (with) an OSSF lepton pair are shown on the left (right). The red, blue, and green lines show the expected distributions for N production at different m_N values, with yields normalized to the total SM background yield.

Table 7.3: Search regions definitions for events in the high-mass region without an OSSF lepton pair (CX). Events are considered part of the high-mass region if they satisfy $p_T^{\text{leading}} > 55 \text{ GeV}$. The subleading (trailing) lepton is required to have $p_T > 15(10) \text{ GeV}$.

$M_{3\ell}(\text{ GeV})$	$M_{\text{T}}(\text{ GeV})$	$M_{2\ell\text{OS}}^{\text{min}}(\text{ GeV})$		
		< 100	$100\text{--}200$	> 200
0–100	< 100	C1		
	> 100	C2		
> 100	< 100	C3	C7	C9
	100–150	C4	C8	
	150–250	C5		
	> 250	C6		

Table 7.4: Search regions definitions for events in the high-mass region that contain an OSSF lepton pair (DXX). Events are considered part of the high-mass region if they satisfy $p_T^{\text{leading}} > 55 \text{ GeV}$. The subleading (trailing) lepton is required to have $p_T > 15(10) \text{ GeV}$. Events with $|M_{\ell\ell} - m_Z| < 15 \text{ GeV}$ or $|M_{3\ell} - m_Z| < 15 \text{ GeV}$ are rejected.

$M_{3\ell} (\text{ GeV})$	$M_T (\text{ GeV})$	$M_{2\ell\text{OS}}^{\text{min}} (\text{ GeV})$		
		< 100	100–200	> 200
0–100	< 100	D01		
	100–200	D02		
	> 200	D03		
> 100	< 100	D04	D09	D13
	100–200	D05	D10	D14
	200–300	D06	D11	D15
	300–400	D07	D12	D16
	> 400	D08		

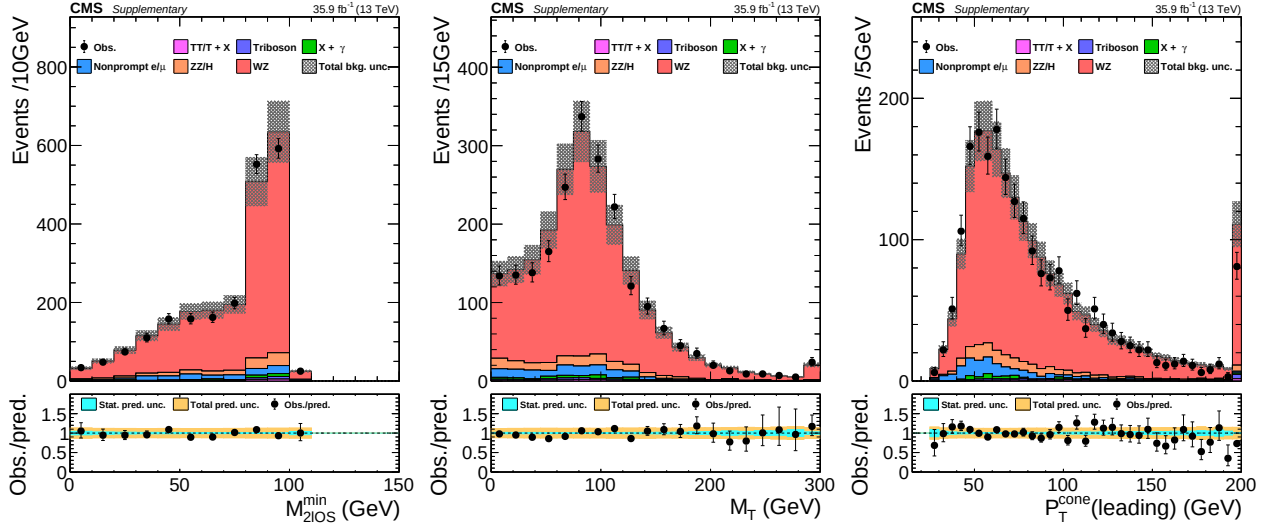


Figure 7.8: Observed and expected event yields in the WZ enriched region as a function of $M_{2\ell OS}^{\min}$ (left), M_T (center), and p_T^{leading} (right).

nonzero mixing matrix element signal events have two or more muons. This signal property is utilized by splitting up events according to whether they contain two or more electrons, or two or more muons. The former set of events is exclusively used to search for a sterile neutrino mixing with ν_e , and the latter to search for a sterile neutrino mixing with ν_μ . The same search region definitions are used in both cases. The split by lepton flavor composition reduces the background to each signal hypothesis by a factor of roughly two, while having no cost in terms of signal efficiency.

7.5 Background estimation

An overview of SM processes that make up the background in multilepton events is given in Section 6.3. The contributions to the event yields in the search regions from all background sources involving prompt leptons are estimated from simulation. The yields of backgrounds with one or more nonprompt leptons are estimated with the fake rate method (Section 6.4).

To constrain the normalization of important background sources with prompt leptons and to validate simulated predictions, three control regions are used in the analysis and defined below. Each control region is enriched in one of WZ, $X\gamma^{(*)}$, and ZZ production. The normalizations of the aforementioned processes are extracted from a simultaneous fit to the three control regions. The signal contribution to the control regions is found to be negligible at $|V_{\ell N}|^2$ values close to the expected exclusion limits, and is not taken into account in this fit ².

Production of WZ and $W\gamma^{(*)}$ is the largest source of high-mass region events with an OSSF pair. The WZ control region consists of events that pass the baseline selection, have an OSSF lepton pair, and satisfy $|M_{\ell\ell} - m_Z| < 15$ GeV. To increase the purity in WZ production, the three leptons, ordered in p_T , must pass $p_T > 25/15/10$ GeV. In addition, p_T^{miss} is required to be larger than 50 GeV. The resulting control region is a highly pure sample of WZ production in which the ratio of the predicted to the observed WZ yield is found to be 1.08 ± 0.09 . This ratio is applied to correct the predicted WZ yield in the search regions, and the corresponding uncertainty is propagated to all results. A good agreement between the observed and predicted event yields in the WZ enriched region is found for all analysis variables. Some of these are shown in Fig. 7.8. No additional process specific uncertainties are assigned to the prediction of WZ production, aside from the general uncertainties listed in Section 6.5.

²A better approach would have been to include the control regions in the fits used to interpret the search region results in terms of each signal hypothesis. This would properly account for signal contamination in the control regions, and is the approach taken in the analyses presented in chapters 8 and 9.

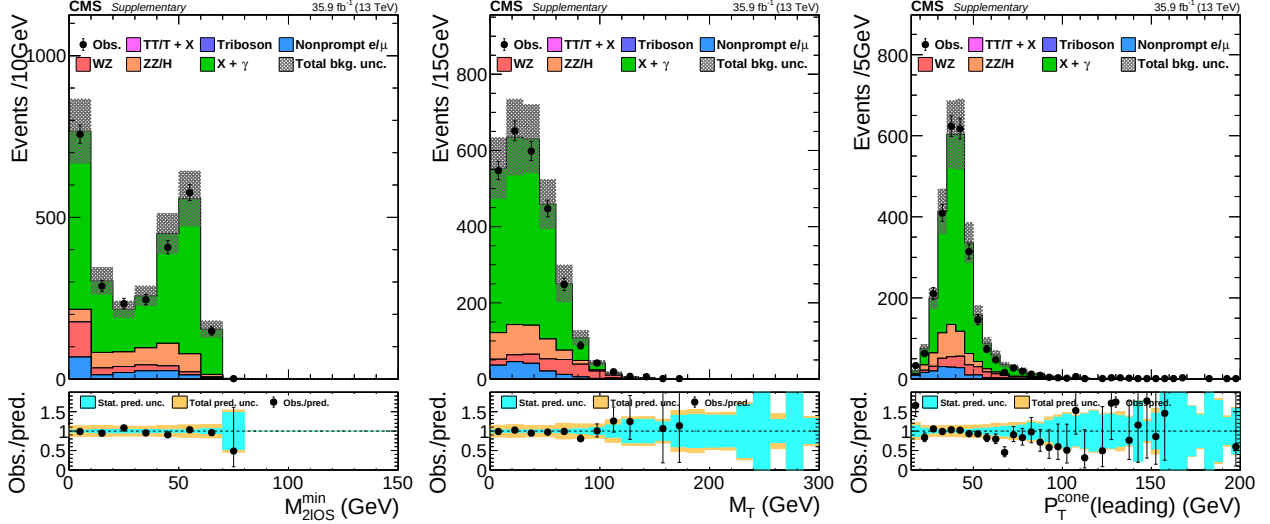


Figure 7.9: Observed and expected event yields in the $X\gamma^{(*)}$ enriched region as a function of $M_{2\ell OS}^{\min}$ (left), M_T (center), and p_T^{leading} (right).

Background events from $X\gamma^{(*)}$ and ZZ production also form a sizable contribution to the high-mass search regions with OSSF lepton pairs. The control region enriched in $X\gamma^{(*)}$ production is defined by selecting three lepton events that pass the baseline selection, have an OSSF pair, and satisfy $|M_{3\ell} - m_Z| < 15 \text{ GeV}$ and $|M_{\ell\ell} - m_Z| > 15 \text{ GeV}$. The former requirement isolates $Z\gamma^{(*)}$ which has a clear $M_{3\ell}$ peak around the Z boson mass as seen in Fig. 7.5. The latter requirement reduces the contamination from WZ production. The ratio of the observed to the predicted $X\gamma^{(*)}$ event yield in the control region is 0.95 ± 0.08 . This ratio is propagated as a correction to the prediction of $X\gamma^{(*)}$ in the rest of the analysis. Distributions of the most important analysis variables are shown in Fig. 7.9. A generally good agreement between the observation and the simulated predictions in the control region is seen, except for some discrepancies at high p_T^{leading} . To cover these discrepancies, the uncertainty in the normalization of the $X\gamma^{(*)}$ simulation is increased to 15%.

The ZZ enriched region is defined by requiring four leptons, forming two exclusive OSSF pairs that both have an invariant mass in a 30 GeV window centered at m_Z . This results in a control region extremely pure in ZZ production. A normalization ratio of 1.03 ± 0.10 for the theoretical ZZ production cross section is measured in the control region. The agreement between data and simulation as a function of important kinematic variables is checked after releasing the mass requirement on one of the lepton pairs. Some distributions are shown in Fig. 7.10. At low M_T , the simulated predictions are within the total experimental and theoretical uncertainties, but at high M_T slight discrepancies are observed. To account for this, an additional 25% uncertainty is applied to predicted ZZ events with $M_T > 75 \text{ GeV}$.

In events without an OSSF lepton pair, nonprompt leptons form the dominant background source. Such events are mainly from Drell-Yan and $t\bar{t}$ production. The nonprompt lepton background is predicted with the fake rate method, and the closure of the method is verified in simulation in the way described in Section 6.4. To ensure that the fake rate is well measured, and the conclusions drawn in simulation can be extrapolated to data, we define two data control regions enriched in nonprompt leptons. A data set is enriched in Drell-Yan production by selecting events that pass the baseline selection, have an OSSF lepton pair with $|M_{\ell\ell} - m_Z| < 15 \text{ GeV}$, a third lepton forming a transverse mass below 30 GeV with p_T^{miss} , and $p_T^{\text{miss}} < 30 \text{ GeV}$. The cuts on p_T^{miss} and the transverse mass remove the bulk of events from WZ production.

A $t\bar{t}$ enriched region is defined by using the baseline selection, but with the requirement on the absence of b jets inverted. If an OSSF lepton pair is present, the contributions from WZ and Drell-Yan production are reduced by requiring $|M_{\ell\ell} - m_Z| > 15 \text{ GeV}$. A requirement of $|M_{3\ell} - m_Z| > 15 \text{ GeV}$ is

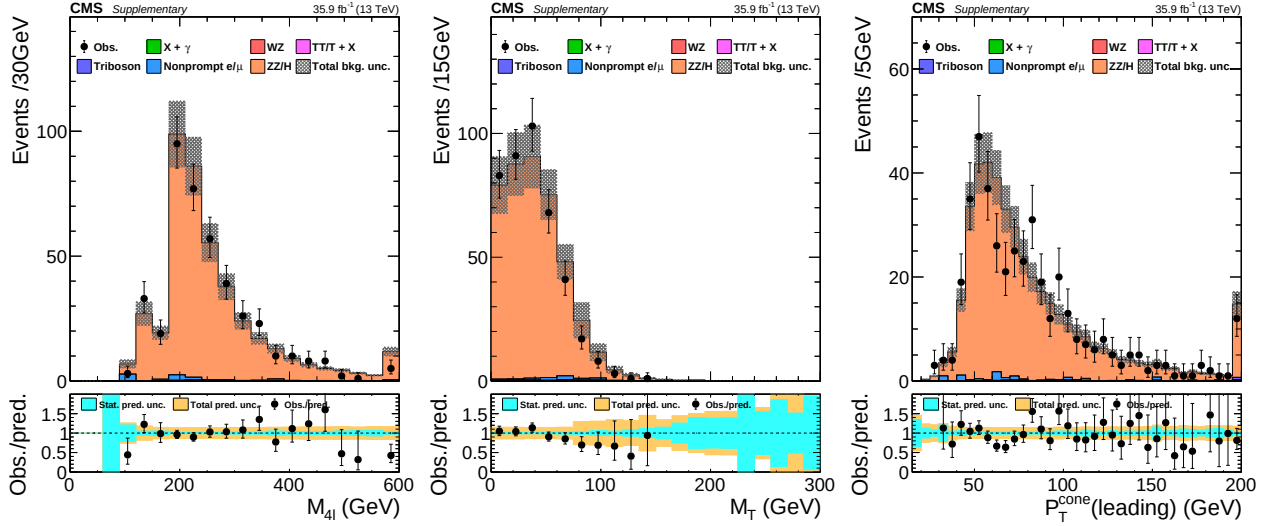


Figure 7.10: Observed and expected events in the ZZ enriched region as a function of $M_{4\ell}$ (left), M_T (center), and p_T^{leading} (right). The small peak at low $M_{4\ell}$ values is due to H boson production with a subsequent $H \rightarrow ZZ^*$ decay.

added to reduce the $X\gamma^{(*)}$ contamination. Some remaining conversions and low-mass meson resonances are removed by vetoing any event with an OSSF lepton pair that has an invariant mass below 12 GeV.

Comparisons between the predicted and observed yields in the Drell-Yan ($t\bar{t}$) enriched region as a function of several search variables are shown in Fig. 7.11 (7.12). The fake rate method can be seen to give accurate predictions of the nonprompt lepton background from both Drell-Yan and $t\bar{t}$ production, matching the observation within the 30% uncertainty assigned to the method.

7.6 Systematic uncertainties

Section 6.5 gives a general overview of experimental and theoretical uncertainties in the predictions of signal and background yields. All of these uncertainties with the exception of those in the jet energy resolution and on the ISR and FSR scale, are applied in this analysis. The uncertainty in the luminosity measurement only affects the total yields of all processes. The size of other general uncertainty sources, on the other hand, varies across search region bins. All experimental and theoretical uncertainties are taken into account for simulated signal and background events, as well as in the subtraction of the prompt contamination in the data-driven estimate of the nonprompt background. An overview of systematic uncertainty sources, and the corresponding sizes of the variations in the search region yields is given in Table 7.5.

The normalization of several simulated background processes is constrained in control regions in data, as discussed in Section 7.5. Theoretical uncertainties in the cross sections of these processes are not considered, instead we use the uncertainties in the normalization measurements obtained in the control regions. Simulated background processes without a dedicated control region are associated production of one or more top quarks with electroweak bosons and production of three or more electroweak bosons. The contribution of these processes to the search regions is tiny, and a 50% normalization uncertainty is assumed for simplicity. This is large enough to cover a potentially large degree of mismodeling, while not significantly affecting the sensitivity of the search.

The uncertainty in the knowledge of the trigger efficiencies is taken to be 5% when p_T^{leading} is below 30 GeV and 2% otherwise, following the reasoning laid out in Section 7.3.1.

Uncertainties from variations in the choice of PDF sets and QCD renormalization and factorization scales affect both the normalization and shapes of each process. For background processes only the shape uncertainties are considered since normalization uncertainties are already accounted for. For

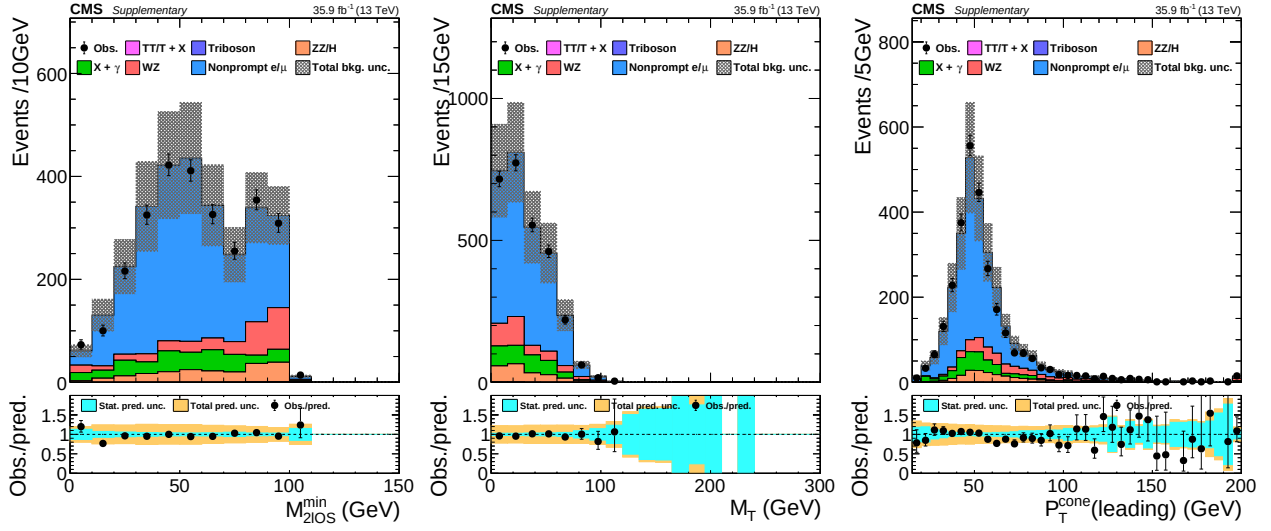


Figure 7.11: Observed and expected event yields in the Drell-Yan enriched region as a function of $M_{2\ell OS}^{\min}$ (left), M_T (center), and p_T^{leading} (right).

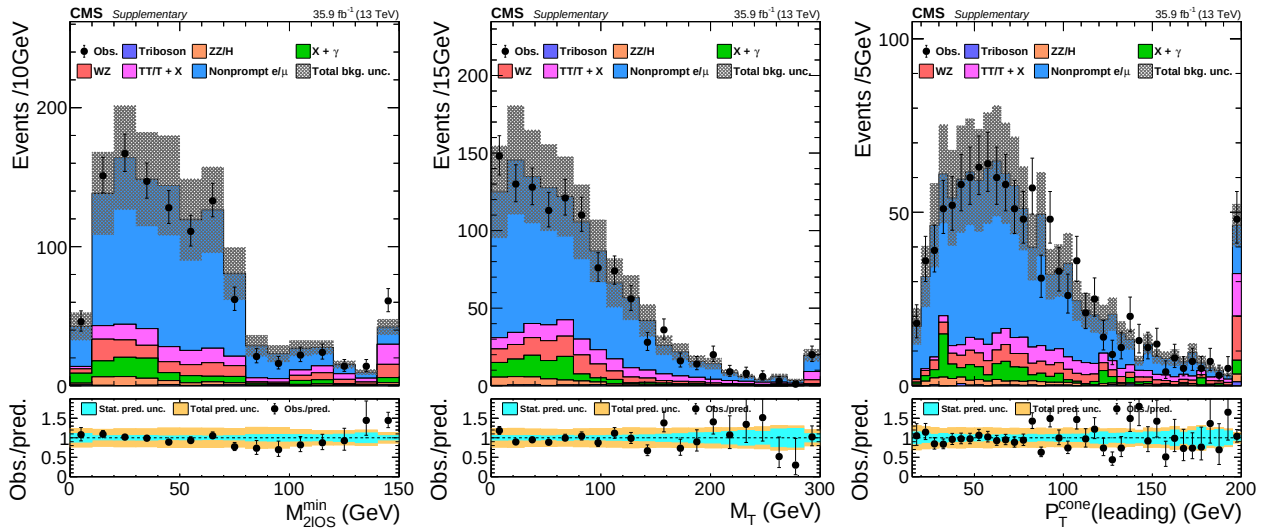


Figure 7.12: Observed and expected event yields in the $t\bar{t}$ enriched region as a function of $M_{2\ell OS}^{\min}$ (left), M_T (center), and p_T^{leading} (right).

Table 7.5: Overview of systematic uncertainties in the search for sterile neutrino production. For each uncertainty source the range of variations induced in the expected yields across the search regions is given. Experimental and theoretical uncertainties that affect all simulated processes are listed first, and subsequently uncertainty sources specific to individual processes are given.

Source	Uncertainty range (%)	Treatment
General experimental uncertainties		
e/ μ identification efficiency	3 (per ℓ)	normalization
Trigger efficiency	2–5	shape
Jet energy scale	0–3	shape
b jet veto	1–5	shape
Pileup	1–5	shape
Integrated luminosity	2.5	normalization
Theoretical uncertainties		
QCD scale	1–15	shape & normalization
PDF	0.1–1	shape & normalization
Process specific uncertainties		
Fake rate method closure	30 (of nonprompt)	normalization
Prompt leptons in fake rate measurement	5–20 (of nonprompt)	shape
$X\gamma^{(*)}$ normalization	15 (of $X\gamma^{(*)}$)	normalization
WZ normalization	8.5 (of WZ)	normalization
ZZ normalization	10 (of ZZ)	normalization
ZZ M_T mismodeling	25 (of ZZ if $M_T > 75$ GeV)	shape
Rare process normalization	50 (of rare SM processes)	normalization

signal events the uncertainties in the normalization are taken into account.

The estimation of the nonprompt lepton background is taken to have a normalization uncertainty of 30%. This covers any non closure observed in the fake rate method. An additional source of uncertainty is the subtraction of the prompt lepton contamination in the fake rate measurement. The production of W and Z bosons forms a background in the single lepton data set in which the fake rate is measured. The normalization of this prompt lepton contamination is determined in events with high p_T^{miss} and M_T , depleted of nonprompt leptons (Section 6.4.3). The uncertainty in the measured normalization is propagated to the fake rate measurement performed at low p_T^{miss} and M_T . This results in an additional uncertainty ranging from 5 to 20% in measured the fake rate, with larger uncertainties at high lepton p_T .

Lepton selection efficiencies are measured in Drell-Yan events using the tag-and-probe technique (Section 6.1.3). Data to simulation scale factors for the lepton efficiency measured in Z boson decays are not necessarily applicable to signal events. Differences in the underlying event structure between signal and $Z \rightarrow \ell\ell$ events can cause minute variations in the I_{rel} values of prompt leptons. This in turn results in slightly different lepton selection efficiencies. To cover these extrapolation effects, a flat 3% uncertainty per lepton is assumed throughout the analysis. This value significantly exceeds the uncertainties in the lepton selection efficiencies measured in Drell-Yan events, which are well below 1% for most lepton p_T and η values. This is a conservative estimate of the per lepton uncertainties³, but the impact on the analysis sensitivity is negligible because the sensitivity is driven by bins with very low event yields in which the dominant background contribution is estimated directly from collision data.

³In the analysis described in Chapter 9 the extrapolation uncertainties are estimated under similar conditions, and they are found to only approach 3% at very high p_T and $|\eta|$ values.

7.7 Search region results

The observed and expected yields in each of the search region bins of the analysis are shown in Fig. 7.13. The figures include both the low- and high-mass regions, and events with two or more electrons as well as those with two or more muons. The systematic uncertainties in the figures are those listed in Section 7.6. The observed yields are compatible with the SM expectation in all bins, and no evidence for N production is seen. The interpretation of the results is the subject of the next section.

7.8 Interpretation

No significant excess of events over the SM expectation is found in the search regions. The observed and expected yields are therefore used to set 95% confidence level upper limits on $|V_{eN}|^2$ and $|V_{\mu N}|^2$ as a function of m_N with the CL_s method [287, 288]. To set an upper limit on $|V_{eN}|^2$ ($|V_{\mu N}|^2$) simultaneous maximum likelihood fits to the search region bins with two or more electrons (muons) are performed. The control regions are not directly included in the fits since the normalization factors of the SM processes studied in these regions are determined a priori.

All systematic uncertainties listed in Section 7.6 are included as nuisance parameters in the likelihood function of the fit. Log-normal probability density functions are used for the nuisance parameters of uncertainty sources that only affect the normalizations of processes. Nuisance parameters that affect both the shapes and normalizations of processes are included in the likelihood by means of the template morphing technique [289], and are assumed to have Gaussian probability density functions. Statistical uncertainties due to the finite size of simulated samples, and the sideband data set used in the data-driven estimation of the nonprompt lepton background, are accommodated in the likelihood function with the Barlow-Beeston method [290]. The computation of the limits themselves uses the asymptotic approximation of the profile likelihood test statistic [291, 292].

For a given m_N , the expected number of signal events in the search regions is computed from a simulation sample produced with an arbitrary $|V_{\ell N}|^2$ value. The described fitting procedure is then used to set an upper limit on the signal strength, the ratio of the expected to observed number of signal events, given the simulated signal model. The N production cross section is directly proportional to $|V_{\ell N}|^2$, so an upper limit on the signal strength of a model with a known $|V_{\ell N}|^2$ value can be directly translated into an upper limit on $|V_{\ell N}|^2$ itself. This procedure is applied to compute exclusion limits for m_N values between 1 GeV, and 1.2 TeV. At sterile neutrino masses below 1 GeV the analysis rapidly loses sensitivity because the N lifetime becomes large, a topic which will be addressed shortly. When m_N exceeds 1.2 TeV, the upper limits the analysis can set on $|V_{\mu N}|^2$ and $|V_{eN}|^2$ exceed the physical maximum of 1, so we cut off the interpretation there. The observed and expected exclusion limits on $|V_{eN}|^2$ ($|V_{\mu N}|^2$) are shown in Fig. 7.14 (7.15).

The decay width of N scales as the fifth power of its mass and is directly proportional to $|V_{\ell N}|^2$ [293, 294]:

$$\Gamma(N) \propto m_N^5 |V_{\ell N}|^2. \quad (7.1)$$

The decay width of a particle is inversely proportional to its mean lifetime, so the following holds:

$$\tau(N) \propto \frac{1}{\Gamma(N)} \quad (7.2)$$

$$\propto m_N^{-5} |V_{\ell N}|^{-2}. \quad (7.3)$$

This implies that the lifetime of sterile neutrinos with a sufficiently small mass and mixing with SM neutrinos is expected to be large. By requiring prompt leptons with small impact parameters in the event selection (Section 7.3.1), the search is explicitly designed to target almost instantaneous sterile neutrino decays. Assuming every N has a zero lifetime we obtain the upper limits represented

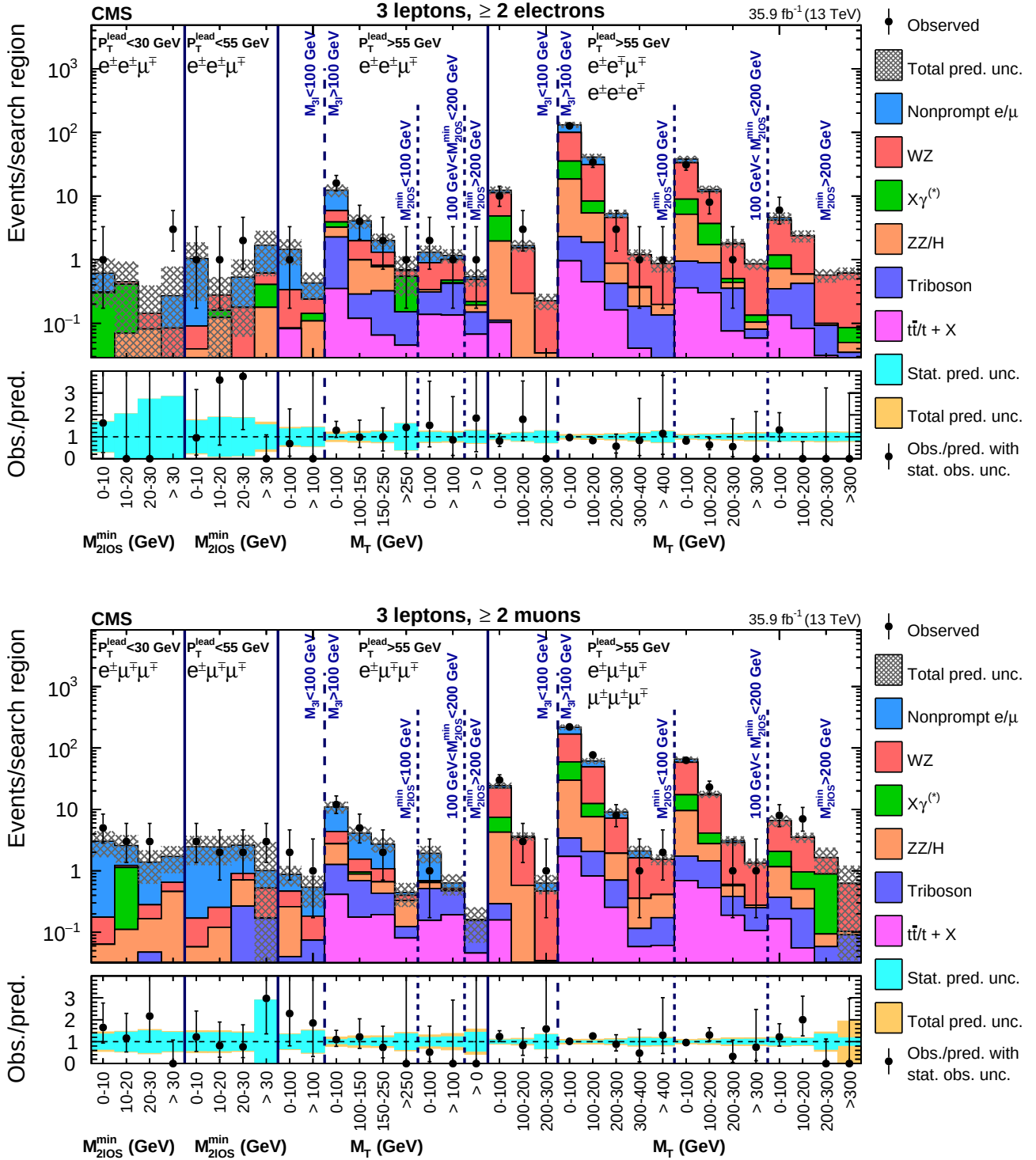


Figure 7.13: Observed and expected event yields in the search region bins for events with at least two electrons (top) and at least two muons (bottom). The first eight bins in each figure correspond to the low-mass region, whereas the other bins display the high-mass region.

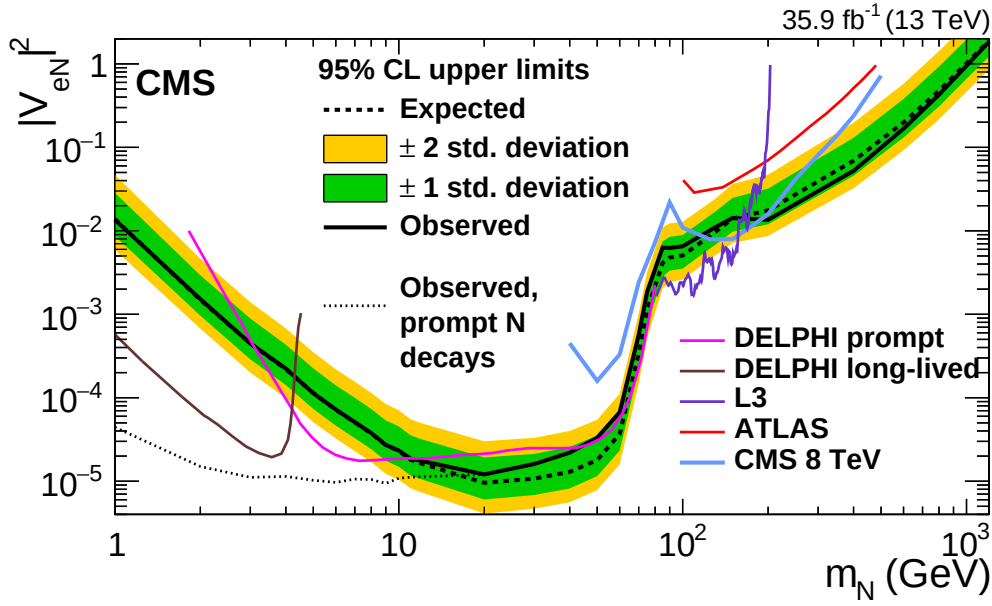


Figure 7.14: 95% confidence level upper limits on $|V_{eN}|^2$ as a function of m_N . Any $|V_{eN}|^2$ value above the curves is considered excluded. The dashed black line represents the lifetime-corrected expected exclusion limits, with the one and two standard deviation bands drawn in green and orange respectively. The lifetime-corrected observed exclusion limits are drawn as a solid black line, while the black dotted line shows the observed limits under the assumption of prompt N decays. Exclusion limits set by the DELPHI experiment at LEP in $Z \rightarrow \nu\nu$ events are drawn as solid brown and magenta lines [266]. Limits by the L3 experiment, also at LEP, but set at higher center of mass energies are drawn in purple [264]. Limits from ATLAS [269] and CMS [274] searches in same-sign dilepton events in 8 TeV pp collisions are drawn as red and blue solid lines.

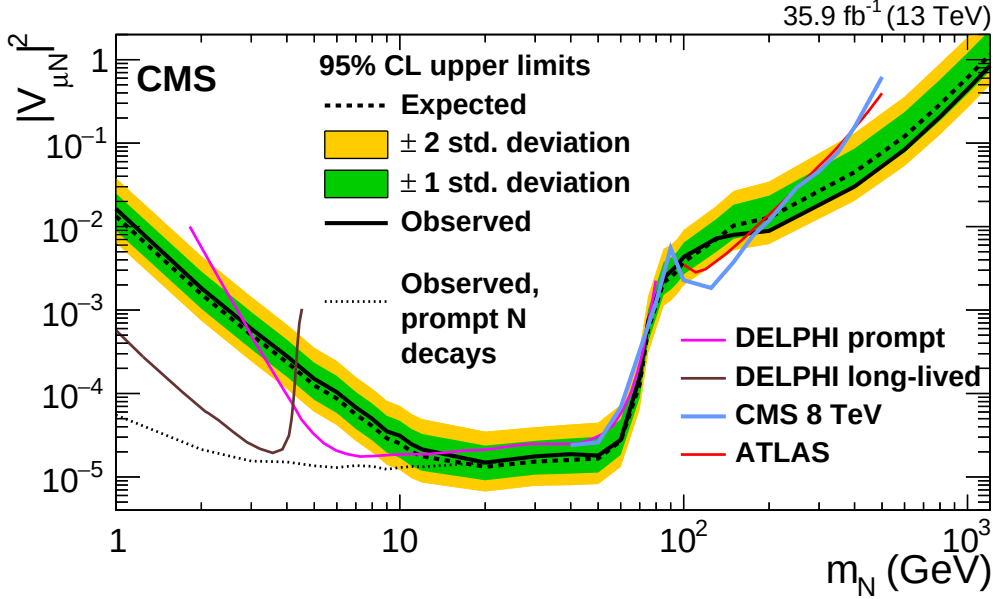


Figure 7.15: 95% confidence level upper limits on $|V_{\mu N}|^2$ as a function of m_N . Any $|V_{\mu N}|^2$ value above the curves is considered excluded. The dashed black line represents the lifetime-corrected expected exclusion limits, with the one and two standard deviation bands drawn in green and orange respectively. The lifetime-corrected observed exclusion limits are drawn as a solid black line, while the black dotted line shows the observed limits under the assumption of prompt N decays. Exclusion limits set by the DELPHI experiment at LEP in $Z \rightarrow \nu\nu$ events are drawn as solid brown and magenta lines [266]. Limits from ATLAS [269] and CMS [273] searches in same-sign dilepton events in 8 TeV pp collisions are drawn as red and blue solid lines.

by the black dotted lines in Fig. 7.14 and 7.15. At an m_N value of 1 GeV, $|V_{eN}|^2$ ($|V_{\mu N}|^2$) values above 4.4 (5.5) $\cdot 10^{-5}$ are excluded under the prompt decay assumption. This upper limit corresponds to a mean N travel distance before decay, $c\tau$, of 8.0 (6.4) m. The CMS tracking detector only extends up to radial distances of 116 cm (Section 4.4), and the ECAL up to about 152 cm [169]. Electrons that originate from several meters from the primary interaction vertex can therefore not be reconstructed at all. Muons might be reconstructed as standalone muons (Section 5.1.1.4), but with a highly reduced efficiency. An N travel distance of just a few cm would be enough to give final state leptons with several missing hits in the pixel detector, bringing about efficiency reductions in the tracking algorithms (Section 5.1.1). Even such relatively small travel distances are well beyond the reach of this analysis. To make the background rate from nonprompt leptons manageable, electrons and muons are explicitly required to have impact parameters with a very small significance and a nominal value of less than about one mm (Section 7.3.1). If sterile neutrinos with a mass of 1 GeV were produced at a mixing parameter value equal to the upper limit under the prompt decay hypothesis, the presented analysis would not detect them. This brings us to the conclusion that the exclusion limits must be corrected for the true N lifetime.

A naive but legitimate approach to correct the limits for the N lifetime would be the following: First compute the exclusion limit at a particular m_N value under the assumption of prompt N decays. Now simulate a new sample of N production events at the $V_{\ell N}$ value of the prompt exclusion limit, this time accounting for the N lifetime in the simulation. Then rerun the analysis with the new sample as the signal model. Because of the large N lifetime most signal events will fail to pass the event selection, and a much larger upper limit is obtained than previously ⁴. The new upper limit is an overestimate

⁴To set the limit one assumes that the excluded signal strength varies linearly with $|V_{\ell N}|^2$, as it does under the prompt hypothesis. This does not hold when the N lifetime is significant, but the iterative nature of the proposed algorithm makes this point moot.

of the true exclusion limit. One can proceed to simulate a new sample of N production events at the $V_{\ell N}$ value of the last exclusion limit. This $V_{\ell N}$ value corresponds to a finite, though probably small, lifetime. Evaluating the analysis with the latest sample once again results in a low upper limit on $V_{\ell N}$. The limit is however higher than the one obtained under the prompt decay assumption because the most recently simulated N decays have a nonzero mean lifetime. One can repeat the procedure above several times, and the resulting exclusion limits will oscillate around the true exclusion limit with an amplitude that becomes smaller at each iteration. Repeating the procedure until a convergence criterion is met would give the corrected exclusion limit. The naive algorithm proposed here is correct, but too computationally demanding for practical use. Simulating a sample of N production events large enough for an accurate computation of the exclusion limit at each iteration of the algorithm is prohibitively expensive. Several observations about the properties of signal events facilitate the design of a much more computationally efficient correction algorithm for the exclusion limits. The basic idea behind the algorithm remains the same. The algorithm proposed below is one of the novelties of the presented analysis.

The first simplification is made by studying the kinematics of the final state in N production events simulated at different mean lifetimes. For several m_N values, simulations are made at a range of mean N lifetimes. As expected, the event selection efficiency varies drastically with the mean N lifetime. However, after applying the prompt lepton selection, the distributions of the analysis variables ($M_{2\ell OS}^{\min}$, M_T , p_T^{leading} , $M_{3\ell}$, and p_T^{miss}) are found to be identical between simulated samples at different mean lifetimes. The analysis selection filters out only those events in which the leptons originate from close to the primary pp interaction vertex. The resolution of the kinematic variables that depend on the leptons therefore stays unaltered. The $\vec{p}_T^{\text{miss}}$ reconstruction also remains unaffected by the mean N lifetime after low impact parameter leptons are selected. This implies that we do not need to process the entire analysis to compare limits at different lifetimes. Instead, we simply have to compare the total event selection efficiency integrated across the search regions. This already considerably simplifies the naive limit correction algorithm introduced above because no complex likelihood fit to all signal region bins must be done at each step.

The second and crucial observation is the fact that it is not necessary to produce multiple simulation samples to evaluate the selection efficiency at different mean N lifetimes for a given m_N . The probability density for a particle with a mean lifetime τ to decay at time t is given by:

$$p_{\text{decay}}(t|\tau) = \frac{1}{\tau} \exp\left(-\frac{t}{\tau}\right) \quad (7.4)$$

A simulated sample of N production events with a mean lifetime of τ_1 can therefore be statistically reweighted to have another mean lifetime value τ_2 , by applying the following weight to each simulated event:

$$W_{\tau_1 \rightarrow \tau_2}(t) = \frac{p_{\text{decay}}(t|\tau_2)}{p_{\text{decay}}(t|\tau_1)} \quad (7.5)$$

$$= \frac{\tau_1}{\tau_2} \exp\left(\frac{t}{\tau_1} - \frac{t}{\tau_2}\right). \quad (7.6)$$

Here it must be noted that the real decay time t of N is known in simulated events. Once a sample of simulated N production events is generated at a single mean lifetime τ_1 , it can be reweighted using Equation 7.6 to effectively become a simulated sample at any arbitrary mean lifetime τ_2 . Due to the direct relationship between $|V_{\ell N}|^2$ and the mean N lifetime, this also implies that we can transform simulated samples at a particular $|V_{\ell N}|^2$ value to another one. The reweighting procedure is shown in Fig. 7.16, where a sample of N production events at one $V_{\mu N}$ value is reweighted to correspond to a different $V_{\mu N}$ value. The only caveat to the procedure is that reweighting samples generated with a low mean lifetime to a higher one will have poor statistical precision at high N travel distance values, rare in the original sample. This is of no concern to this analysis in which the selection only accepts decay products of sterile neutrinos that travel a tiny distance through the detector.

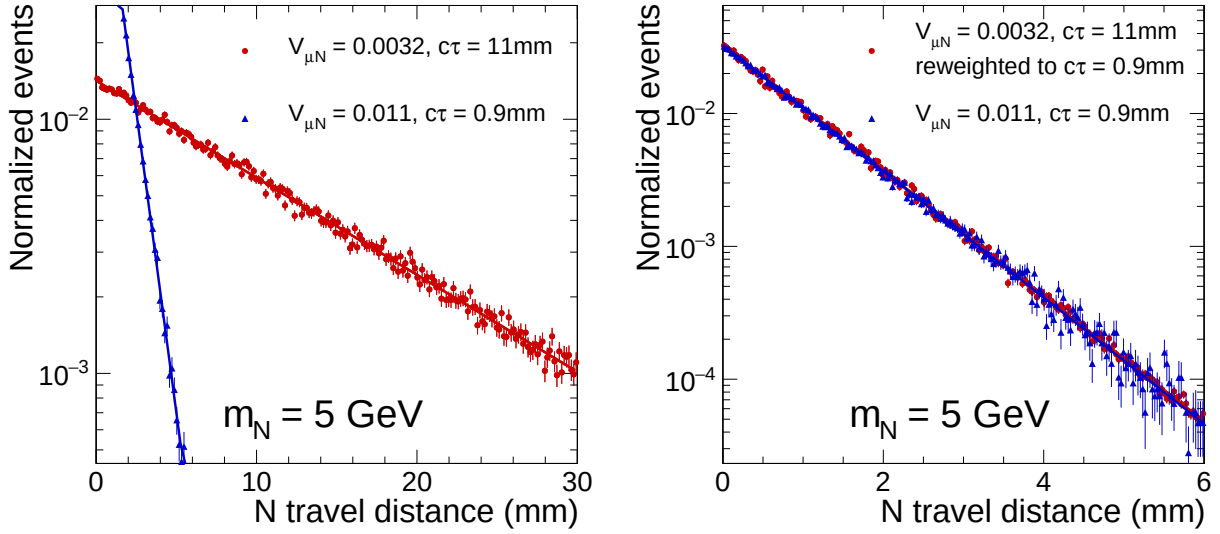


Figure 7.16: Distributions of the travel distance of an N with a mass of 5 GeV before decay at different values of $V_{\mu N}$, and without mixing to the other SM neutrino flavors. Fits of Equation 7.4, in which the lifetime is multiplied by c to obtain the travel distance, are drawn as solid lines. In the left panel no reweighting is applied, while the weights specified by Equation 7.6 are applied in the right panel to transform the simulated sample at $V_{\mu N} = 0.0032$ into an effective $V_{\mu N} = 0.011$ sample.

To summarize, we can transform simulated samples at fixed m_N values to different mean N lifetimes, and comparing the change in the total selection efficiency as a function of the mean N lifetime is equivalent to comparing the change in the exclusion limit. Armed with this knowledge we can formulate the lifetime correction algorithm for the exclusion limits. The correction procedure starts by observing that the uncorrected exclusion limits are essentially a limit on the number of signal events after the analysis selection. The number of signal events that can be excluded is proportional to $|V_{\ell N}|^2$ and the selection efficiency of the analysis. We can write:

$$N_{\text{excluded}} \propto |V_{\text{uncorrected}}|^2 \mathcal{E}_{\text{prompt}}, \quad (7.7)$$

with $|V_{\text{uncorrected}}|^2$ the exclusion limit on $|V_{\ell N}|^2$ we get under the prompt decay assumption, and $\mathcal{E}_{\text{prompt}}$ the event selection efficiency of the analysis under the same assumption. The corrected limit should still correspond to the same number of excluded events, so the problem of correcting the exclusion limits can be cast as solving the following equation:

$$|V_{\text{uncorrected}}|^2 \mathcal{E}_{\text{prompt}} = |V_{\text{corrected}}|^2 \mathcal{E}_{\text{corrected}} \quad (7.8)$$

for $|V_{\text{corrected}}|^2$. If we have a single simulated sample of N production with a nonnegligible lifetime, we can compute the efficiency corresponding to any $|V_{\ell N}|^2$ value with the reweighting procedure of Equation 7.6. We know that $|V|^2 \mathcal{E}$ is a monotonically increasing function because the selection efficiency never decreases as a function of $|V|^2$, or equivalently with decreasing mean N lifetime. It is also guaranteed that $|V_{\text{uncorrected}}|^2 \leq |V_{\text{corrected}}|^2$ because $\mathcal{E}_{\text{corrected}} \leq \mathcal{E}_{\text{prompt}}$. Another constraint on the problem is the physical requirement: $|V_{\text{corrected}}|^2 \leq 1$ ⁵. We can thus state that the following property holds:

$$|V_{\text{corrected}}|^2 \in [|V_{\text{uncorrected}}|^2, 1] \quad (7.9)$$

⁵If this limit does not hold, the analysis has no sensitivity to the studied m_N value. An exclusion limit on $|V_{\ell N}|^2$ might be larger than 1 and still carry meaningful information about how close an experiment comes to being sensitive to the possibility of a sterile neutrino at a given mass, but we do not consider such cases.

We search this interval for the solution. The approach we use is inspired by the binary search algorithm. We take the middle of the interval, $|V_{\text{mid}}|^2 = (1 - |V_{\text{uncorrected}}|^2)/2$, and compute the corresponding selection efficiency $\mathcal{E}_{\text{mid}}$. If $|V_{\text{mid}}|^2 \mathcal{E}_{\text{mid}}$ is greater (smaller) than $|V_{\text{uncorrected}}|^2 \mathcal{E}_{\text{prompt}}$ then $|V_{\text{corrected}}|^2$ must be smaller (greater) than $|V_{\text{mid}}|^2$. We can iterate this procedure to make the search interval exponentially smaller at each step, until convergence is achieved. Algorithm 1 writes out the lifetime correction procedure in pseudocode.

Algorithm 1 Algorithm for correcting an exclusion limit on $|V_{\ell N}|^2$ set under the prompt N decay hypothesis for the true sterile neutrino lifetime. The function `LIMITCORRECTION` returns the corrected limit on $|V_{\ell N}|^2$. The variable “simulated_sample” represents a sample of simulated N production events with a finite mean N lifetime. The limit on $|V_{\ell N}|^2$ obtained by assuming prompt N decays is represented by “limit_prompt”. The convergence criterion is encapsulated in a small numeric constant ϵ . The function `COMPUTE EFFICIENCY` computes the efficiency for a given sample and a desired $|V_{\ell N}|^2$ value. The analysis event selection procedure is represented by a function `PASSSELECTION` which is not given in pseudocode.

```

1:
2: function LIMITCORRECTION(simulated_sample, limit_prompt, efficiency_prompt,  $\epsilon$ ):
3:   low = limit_prompt
4:   high = 1
5:   while (high - low)/low >  $\epsilon$  do
6:     mid = (high + low)/2
7:     efficiency_mid = COMPUTE EFFICIENCY(simulated_sample, mid)
8:     if mid * efficiency_mid < limit_prompt * efficiency_prompt then
9:       low = mid
10:    else
11:      high = mid
12:  return low
13:
14: function COMPUTE EFFICIENCY(simulated_sample, V2_target)
15:    $\tau_{\text{sample}} = \text{simulated\_sample}.\tau$ 
16:   V2_sample = simulated_sample.V2
17:    $\tau_{\text{target}} = (\tau_{\text{sample}} / V2_{\text{target}}) * V2_{\text{sample}}$ 
18:   eff_numerator = 0
19:   eff_denominator = 0
20:   for event in simulated_sample do
21:     t = event.lifetime
22:     weight =  $\tau_{\text{sample}}/\tau_{\text{target}} * \text{EXP}(t/\tau_{\text{sample}} - t/\tau_{\text{target}})$ 
23:     eff_denominator += weight
24:     if PASSSELECTION(event) then
25:       eff_numerator += weight
26:   eff = eff_numerator / eff_denominator
27:  return eff

```

We compute lifetime corrections to all $|V_{\ell N}|^2$ exclusion limits up to $m_N = 20 \text{ GeV}$. At higher m_N values the N lifetime becomes negligible, and no corrections to the limits need to be applied. The corrected exclusion limits vary significantly from those under the prompt decay hypothesis. At $m_N = 1 \text{ GeV}$, the difference is almost three orders of magnitude.

Regardless of the efficiency lost to the N lifetime at low m_N values, the analysis retains a strong sensitivity in the low-mass regime. It is the first search at a hadron collider to probe the range $m_N < 40 \text{ GeV}$, and is sensitive to m_N values as low as 1 GeV . On the other end of the mass spectrum, m_N values up to 1.2 TeV can be successfully probed while earlier analyses reached only up to 500 GeV .

The presented search more than doubled the mass range in which direct N production was studied at the time of writing.

7.9 Summary and outlook

Heavy sterile neutrinos could provide solutions to several of the open questions in the SM, making them ideal candidates for new physics searches. A search for the production of a sterile Majorana neutrino (N) in $W^{(*)} \rightarrow \ell N$ decays is presented. The search targets events with three charged leptons, electrons or muons, in the 13 TeV pp collision data set collected with the CMS detector in 2016. The data set has an integrated luminosity of 35.9 fb^{-1} . It is the first search for sterile neutrinos in the multilepton final state at the LHC. The advantage of using multiple leptons is that they can be reconstructed and identified with great precision, even when they have low p_T values. This was maximally exploited by introducing the lowest possible p_T thresholds in the lepton selection allowed by the trigger paths used to collect the CMS pp collision data set. Several discriminating variables are introduced to separate N production events from the SM background events that pass the analysis selection criteria. A set of orthogonal search region bins is defined as a function of four of these discriminating variables.

No significant excess over the SM expectation is observed in any of the search region bins. Exclusion limits are therefore set on the square of the mixing parameter between N and the SM neutrinos ($|V_{\ell N}|^2$) as a function of m_N . The interpretation is done separately for nonzero values of $|V_{eN}|^2$ and $|V_{\mu N}|^2$. The search provides sensitivity to N production at m_N values between 1 GeV and 1.2 TeV. It is the first ever search to directly probe m_N values above 500 GeV, more than doubling the studied mass range. It is also the first search at a hadron collider to be sensitive to N masses in the range below 40 GeV.

The search explicitly targets promptly decaying sterile neutrinos during the event selection stage. This is done to distinguish signal leptons from those from b or c hadron decays, which can be significantly displaced with respect to the primary pp collision vertex. The assumption of instantaneous N decays is well motivated if $m_N > 20 \text{ GeV}$, but has serious consequences to the analysis sensitivity at very low m_N values. The mean lifetime of sterile neutrinos is proportional to the negative fifth power of their mass, so at the low end of the probed mass range the analysis selection efficiency is markedly reduced. Exclusion limits are initially set under the hypothesis of instantaneous N decays and then corrected to take into account the loss of selection efficiency incurred by the nonzero sterile neutrino lifetime. A novel computationally efficient algorithm is introduced to determine the corrections for the N lifetime to the exclusion limits. A natural next step to extend the sensitivity at low m_N values is to explicitly target events with two leptons originating from a vertex displaced with respect to the primary pp interaction vertex. Such a search requires substantially different analysis techniques from the one presented here which targets prompt leptons. Both ATLAS [276] and CMS [277] carried out such searches several years after the publication of the one presented here. The referenced CMS search was led by researchers at UGent, and I was involved in a peripheral fashion through discussions and brainstorming. That analysis also leveraged the software tools I developed for this and later searches.

Looking back on this search with the experience gained from the later searches presented in the dissertation, it is clear that large improvements can be made. The main source of background events in the most sensitive search bins is the misidentification of nonprompt leptons. Meanwhile, the analysis uses a standard I_{rel} based lepton identification procedure (Section 6.1.1.3). The next publication described in this dissertation, presented in Chapter 8, spectacularly demonstrates the power of machine learning based lepton identification strategies (Section 6.1.2 and 8.2.2.2). Switching from a lepton selection strategy similar to that used in this chapter to the one introduced in Chapter 8 increases the selection efficiency of prompt leptons by roughly 10% per lepton, while the background rate for nonprompt leptons is reduced by a factor two (eight) for electrons (muons). Applying such a procedure in this search could be used to virtually eliminate the nonprompt lepton background at no cost at all to the signal selection efficiency.

Further inspiration might be taken from the search presented in Chapter 9. There we use para-

metric neural networks to search for a new physics scenario with several unknown parameters (Section 9.4.1.1). Using this technique in favor of orthogonal search region bins like those used here, the search sensitivity is increased by more than 100% for most signal hypotheses. This is roughly equivalent to quadrupling the data volume available for analysis. We could replace the search region bins of this search by a single neural network parametric in m_N and would likely obtain similar improvements. By combining a machine learning based lepton identification strategy and a parametric neural network for identifying signal events, one might expect to increase the reach of the search in the 2016 data set by such an extent that it would be greater than what could be achieved by analyzing the full CMS Run 2 data set with the current search strategy. If the increased data volume is added on top of these improvements, the exclusion limits should be expected to be staggering compared to what is presented here.

7.10 Acknowledgements

My contributions to the analysis presented here are the following: I designed the analysis in the high-mass region and parts of the analysis in the low-mass region. The development of the algorithm used to correct the exclusion limits for the sterile neutrino lifetime is also my work. In addition, I produced the final results in all search regions and control regions, made all of the published figures, wrote the bulk of the internal CMS documentation relating to the analysis, and was the main editor of the resulting paper. I also gave most of the internal presentations on the topic, including the critical approval presentation.

Several other people made large contributions to the analysis presented here. They are listed below:

Lesya Shchutska initially started the analysis effort, partially designed the low-mass search regions, and advised me throughout my work on this topic. Lesya also contributed to the design of the lifetime correction procedure, and in the editing of the paper.

Tom Cornelis made all simulations of N production events. In particular, he trailblazed the simulation of sterile neutrinos with a significant lifetime, which proved nontrivial to implement in practice. Without these simulations, the analysis would have been impossible. Tom also helped with the estimation of the trigger efficiencies.

Martina Vit performed the measurement of the lepton fake rate. This facilitates the prediction of the most important background in the analysis. She was also involved in the design of the analysis in the low-mass region. Martina gave the preapproval presentation of the analysis.

As our supervisor, Didar Dobur gave both me and Martina important advice during the analysis and was available for discussing our intermediate results. She motivated and steered us during the work when necessary.

Chapter 8

Observation of tZq production

*“what I set out for I got
and what I hunted I found”*

Elias Lönnrot, *Kalevala*

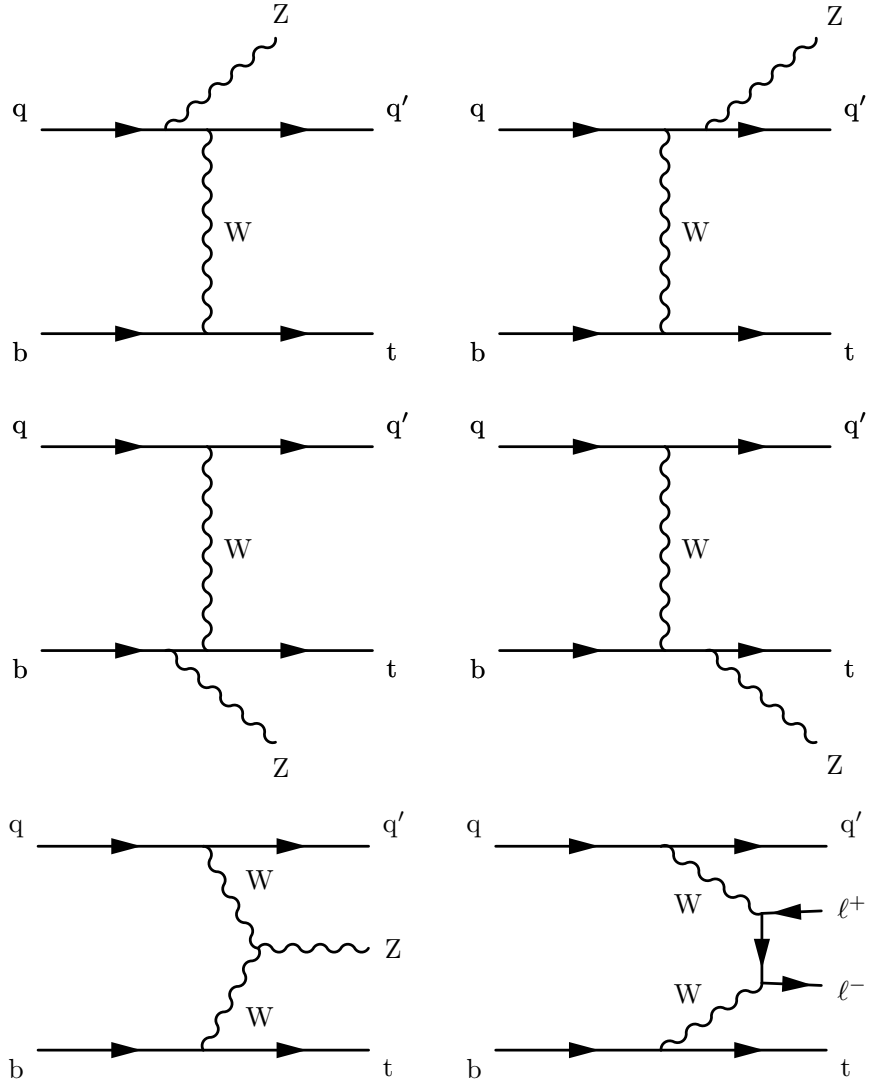
8.1 Introduction

The LHC collided protons at the energy and intensity frontier for several years. During Run 2 an unprecedented volume of data was collected at a center-of-mass energy of 13 TeV. The production cross section of most rare processes that involve heavy particles rises with the collision energy, a fact which is visible in Fig. 6.2. The combination of the large data volume and the increased production cross sections opens the door to the probing of very rare SM processes that were inaccessible to previous colliders and even during Run 1 of the LHC. Measurements of such rare processes might increase our knowledge about the SM or reveal hints about physics beyond the SM, even in the absence of direct production of new particles [295, 296]. New physics might reveal itself as modified integral or differential cross sections of rare SM processes [297].

Production of a single top quark in association with a Z boson and an additional quark ($pp \rightarrow tZq$, where charge conjugation is assumed throughout the chapter) is an extremely rare process that is particularly sensitive to unknown physics modifying the SM. The leading order Feynman diagrams of tZq production are drawn in Fig. 8.1. What makes tZq production so interesting is that it involves many facets of the SM: The WWZ triple-gauge coupling, the $bW \rightarrow tZ$ scattering amplitude, and the ttZ and tbW couplings [298]. New physics that does not modify integral and differential cross section measurements of $t\bar{t}Z$ production nor of inclusive single top quark production could still visibly impact tZq production [299]. Modifications to the SM expectation in tZq production could in particular hint at the presence of flavor changing neutral currents [300–302]. Such currents are greatly suppressed in the SM [303] and their observation would be a clear indication of new physics.

In 13 TeV pp collisions the tZq production cross section is about 1 pb [304]. This makes it about 50 times more rare than the production of an H boson under the same conditions [305]. Prior to the search presented here tZq production remained unobserved. A search by CMS using 8 TeV data yielded an observed significance of 2.4 standard deviations over the background-only hypothesis [306]. Searches in 13 TeV collision data collected in 2016, corresponding to an integrated luminosity of about 36 fb^{-1} , were carried out by both the ATLAS [307] and CMS [244] Collaborations. They respectively achieved observed significances of 4.2 and 3.7 standard deviations above the background-only hypothesis, providing the first evidence for tZq production, but no unambiguous observation.

All the searches for tZq production mentioned above focus on events with three leptons. This final state provides by far the greatest sensitivity for searching and measuring the tZq process. Only a small fraction of the already rare tZq events, about one in 30, contain three leptons. The background in this

Figure 8.1: Leading order Feynman diagrams of tZq production [244].

final state is however orders of magnitude lower than it is in final states with two or fewer leptons. In these final states gigantic backgrounds from W boson, Z boson, $t\bar{t}$, and multijet production effectively eliminate any hope for an observation or precise measurement of tZq production in currently available data sets. In this analysis only events with three light leptons, electrons or muons, are analyzed.

Here we present a revised search for tZq production in the three lepton final state, using the pp collision data set collected with the CMS detector throughout 2016 and 2017. The data set is collected at a center of mass energy of 13 TeV and has a total integrated luminosity of 77.4 fb^{-1} . We introduce a novel lepton identification algorithm to address the nonprompt lepton background that plagued earlier searches. Additionally, several other improvements to the overall analysis strategy are made. These improvements enable the observation of tZq in the same 2016 collision data set which was analyzed by the previous CMS search without a successful observation. On top of this, the integrated luminosity of data set used for the search is more than doubled by adding the data collected in 2017. The analysis achieves the first observation of tZq production and the significance of the observation is well over five standard deviations. The tZq cross section is measured with a relative uncertainty of approximately 15%, more than twice lower than in any previous measurement. The observation made in these results was confirmed about a year later by ATLAS using the full LHC Run 2 data set [308]. The ATLAS analysis attained a similar precision in the measured cross section although roughly 80% more data was analyzed.

A brief outline of the rest of the chapter is given here. Section 8.2 describes the event selection used to target tZq production, and the large improvements made in the identification of leptons. A description of the strategy used to separate the tZq signal from the remaining background in the selected events is given in Section 8.3. Section 8.4 outlines the background estimation procedure. The systematic uncertainties that affect the analysis are covered in Section 8.5. Section 8.6 shows the analysis results: The first observation of tZq production as well as a precise measurement of its production cross section. Section 8.7 summarizes the analysis and presents prospects for tZq measurements in the near future.

8.2 Event and object selection

The final state particles in each event are reconstructed with the PF algorithm (Section 5.1). The requirements imposed on jets are the the topic of Section 8.2.1. The selection of leptons in the analysis is covered in Section 8.2.2, including an in depth discussion on their machine learning based identification in Section 8.2.2.2.

8.2.1 Jet selection

Reconstructed particles are clustered into jets with the anti- k_T algorithm (Section 5.2), and the energy of the resulting jets is corrected for pileup and nonlinearity and nonuniformity in the detector response (Section 5.3). Jets identified as originating from anomalous energy deposits in the detector are removed. To be considered for analysis jets must have $p_T > 25 \text{ GeV}$, $|\eta| < 5$, and be separated from any selected lepton by an angular distance of at least $\Delta R = 0.4$. High $|\eta|$ jets outside of the tracking detector coverage are retained in the analysis because the outgoing quark in tZq events tends to result in a forward jet. This feature can be clearly seen in the event displays of Fig. 8.17 to 8.20 included later in this chapter. These figures display reconstructed events in pp collision data that are highly likely to be from tZq production. At the time the analysis was performed a large amount of noise was present in the very forward region of the ECAL in the data set collected in 2017. This required the use of an increased threshold $p_T > 60 \text{ GeV}$ in the selection of jets falling within the pseudorapidity range $2.7 < |\eta| < 3.0$. This requirement is only applied in the 2017 collision data set. In later reconstructions of the data set this noise has been largely removed by modifying internal parameters of the PF algorithm.

The presence of a top quark in the tZq signal makes b-tagging an important tool for isolating signal

events. Jets are considered b-tagged if they meet the requirements of the medium working point of the DEEPCSV algorithm (Section 5.4). This working point correctly identifies 68% of b jets, while misidentifying 12 (1)% of c quark (light-flavor) jets as a b jets. The identification of b jets primarily relies on information from the tracking detector. Therefore the DEEPCSV b-tagging score is only evaluated for jets that satisfy $|\eta| < 2.4$, a criterion which ensures that most of the jet constituents fall within the coverage of the tracking detector. Jets at higher $|\eta|$ values are assigned a zero probability to contain b hadrons.

8.2.2 Lepton identification

The selection of leptons in this search uses a two-stage procedure. First, leptons are selected with a set of very loose orthogonal requirements that ensure that the leptons are within the detector acceptance and have a minimal quality. These selection criteria are the subject of Section 8.2.2.1. Thereafter, a severe requirement on the output of a machine learning classifier that identifies prompt leptons is applied. This procedure is covered in Section 8.2.2.2.

8.2.2.1 Baseline selection

Reconstructed muons (electrons) are required to fall in the muon system (tracking detector) acceptance of $|\eta| < 2.4$ (2.5). All muons and electrons must have $p_T > 10$ GeV and pass a set of very loose preliminary requirements on several of the variables introduced in Section 6.1. The leptons must be consistent with originating from the primary pp interaction vertex by satisfying $d_{xy} < 0.05$ cm, $d_z < 0.1$ cm, and $d_{3D}/\sigma(d_{3D}) < 8$. Lepton candidates should at least be slightly isolated from other activity in the detector and pass $I_{\text{rel}}^{\text{mini}} < 0.4$, a very loose isolation requirement. Muons must additionally satisfy the medium muon requirements (Section 6.1.1.1) to guarantee a good quality in the matching of the muon tracks in the inner tracker and the muon system. Electrons have to pass a set of loose requirements on the properties of their ECAL supercluter, mirroring those imposed in trigger paths that select multiple leptons. Reconstructed electrons are removed if they fall within $\Delta R < 0.05$ of a muon passing the set of requirements outlined above. The selection described here is just the first stage in the lepton selection. The described requirements have an extremely high efficiency in selecting prompt leptons from tZq production, about 99%, while rejecting many obvious nonprompt leptons. This selection is called the “baseline” lepton selection in the rest of the chapter. The final lepton selection of the analysis adds an additional requirement on a machine learning discriminant, described below, on top of the baseline selection.

8.2.2.2 Machine learning lepton classifier

We target tZq production in the multilepton channel where the decays of both the Z boson and the W boson originating from the top quark decay are leptonic. The target events have three prompt leptons in the final state, making the discrimination between prompt and nonprompt leptons of crucial importance to the sensitivity the analysis can achieve. This is the case more so than in the other analyses presented in this dissertation. The previous 13 TeV search for tZq production conducted by CMS [244] provides an illustration of this. This search has two search regions: Events with a single b-tagged jet and a total of two or three jets, and events with two b-tagged jets. Events without b-tagged jets serve as a WZ control region. Search region events are binned in the output of BDT classifiers, trained to separate the tZq signal from the total background. The resulting observed and post-fit expected distributions are shown in Fig. 8.2. The background from nonprompt leptons is drawn in dark blue and labeled “NPL”. This background source turns out to be difficult to kinematically separate from the tZq signal and remains prominent in the search region bins at high BDT output values. These are the bins where the tZq signal purity is highest. The nonprompt lepton background proved to be the crucial limitation that prevented the observation of tZq in this search.

The earlier CMS search for tZq used a cut based identification for both electrons and muons, in which I_{rel} is the main discriminant between prompt and nonprompt leptons (Section 6.1). This lepton

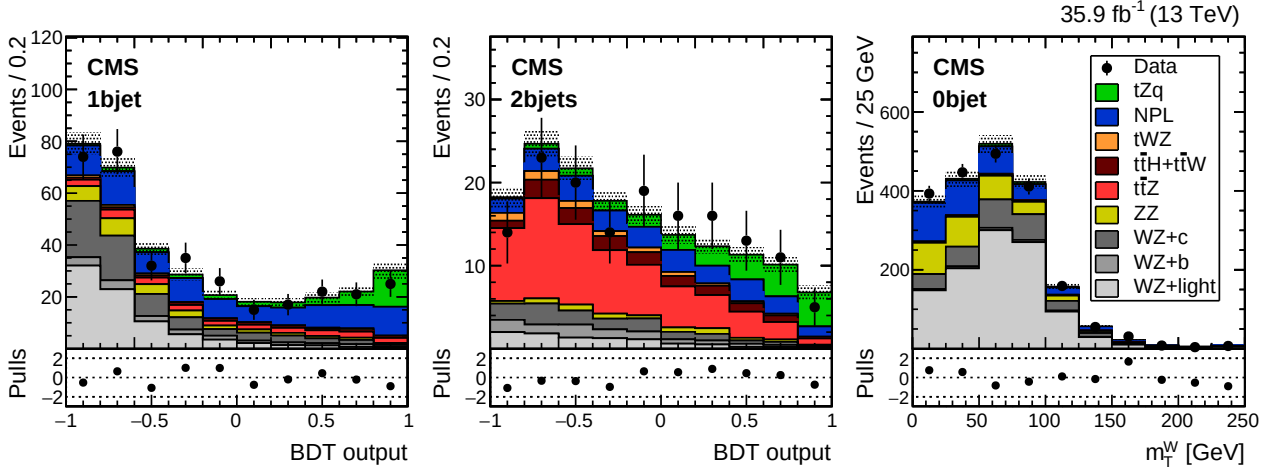


Figure 8.2: Observed and post-fit expected event yields in the search regions (left- and middle) and WZ enriched region of the previous CMS search for tZq production. The background from nonprompt leptons is shown in dark blue and labeled “NPL” [244].

identification strategy is similar to that used in Chapter 7 of this dissertation. To tackle the large nonprompt lepton background we took inspiration from the search that led to the observation of $t\bar{t}H$ production [238, 239]. The multilepton final state proved crucial for observing the $t\bar{t}H$ process, and its sensitivity could to a large extent be credited to the usage of a lepton identification strategy based on BDT classifiers.

This strategy is further refined and optimized in this analysis, and customized to target leptons from tZq production. An introduction to machine learning based lepton identification and an overview of all the input variables used is given in Section 6.1. Several BDT classifiers are trained to separate prompt leptons originating from simulated tZq and $t\bar{t}Z$ production events from nonprompt leptons produced in simulated $t\bar{t}$ events. Events from $t\bar{t}Z$ production are included to increase the size of the training data set. Because of the high similarity of the lepton properties in tZq and $t\bar{t}Z$ events, the added statistical power increases the performance of the BDT classifiers in tZq events, despite the small bias introduced in the lepton properties. Separate BDT classifiers are trained for electrons and muons. The 2016 and 2017 data sets are also treated individually to account for differences in data-taking conditions and the detector, in particular the upgraded pixel detector (Section 4.4.1).

In order to enter the training data set, leptons must pass the baseline selection and no additional requirements. The baseline selection is kept as loose as possible not to spoil the maximum prompt lepton identification efficiency the BDT classifiers can reach. The lepton selection in the analysis mirrors this: Almost no additional requirements¹ are applied after the baseline selection except from a requirement on the trained BDT classifiers. Having no further requirements on the training data set ensures that no identification performance is lost by applying cuts which would be nonoptimal when used without considering their dependence on all other input variables to the lepton selection. The strategy of having a loose baseline selection and letting the BDT classifiers do all the identification work leads to significant gains in the lepton identification efficiency. For electrons an efficiency higher by about 20% per electron is achieved compared to the BDT based lepton identification used for the $t\bar{t}H$ observation, at the same misidentification rate for nonprompt leptons².

The data set available per BDT training consists of about 15 million prompt leptons and 2 million nonprompt leptons. Of these leptons, 80% are used for training and validation of the classifiers, and

¹The only exception to this is a very mild cut on $DEEPCSV_{closest}$ to ensure closure of the fake rate method for predicting the background from nonprompt leptons. In retrospect it would have been more optimal not to apply this cut in the tight selection and further refine the loose selection criteria instead.

²In the case of the $t\bar{t}H$ observation a preselection on the electron MVA is used in the training of the BDT classifiers for electron selection. This turns out to limit the final lepton identification efficiency significantly.

20% for testing. During the optimization of the BDT hyperparameters and the training configuration only half of the training data set is used for the actual training of the algorithms, and the other half to evaluate their performance. When the final training scheme and set of hyperparameters are fixed for each BDT, a new training is performed that uses all available events in the training data set. Not using the test data set for hyperparameter tuning avoids potential overfitting of the BDT hyperparameters on the test data set, which would result in biased conclusions about the classification performance.

The final BDT classifiers used in the analysis are technically gradient boosted random forests of decision trees [309]. Optimizing a single decision tree is NP-hard [310], so a greedy approximation is used: To train a single decision tree, all input variables are scanned individually and the optimal cut on a single variable chosen from all input variables is used to split the tree node. Each tree node is further split until a certain depth or number of events in the leafs is reached. Gradient boosting algorithms aim to iteratively build a powerful discriminator out of a collection of small decision trees. One starts by defining an objective or loss function that is to be minimized. An example of such a minimization objective is the negative log-likelihood in a maximum likelihood fit. Each subsequent tree in the gradient boosted forest fits the local derivative of the loss function with respect to the output of the previous tree. After adding a new tree, the updated forest is given by the previous forest minus the new decision tree multiplied by a scalar value set to minimize the loss function. We can write the update from a forest of n trees F_n to one with $n + 1$ trees F_{n+1} as:

$$F_{n+1}(\mathbf{x}) = F_n(\mathbf{x}) - \gamma_{n+1} \sum_{\text{training data set}} \nabla_{F_n} L(y, F_n(\mathbf{x})) \quad (8.1)$$

$$\gamma_{n+1} = \arg \min_{\gamma} \left\{ \sum_{\text{training data set}} L[y, F_{n-1}(\mathbf{x}) - \gamma \nabla_{F_n} L(y, F_n(\mathbf{x}))] \right\}. \quad (8.2)$$

In these equations the ground truth labels in the training data set are represented by y , in our case whether a lepton is prompt or not. The feature vector $\mathbf{x}$ is the set of input variables to the BDT for each training example, and L is the loss function. The derivatives $\nabla_{F_n} L$ of the loss function with respect to the current boosted forest output are approximated by greedily optimized shallow decision trees. We choose the negative log-likelihood for the binary classification problem as the loss function. This means the BDT trainings effectively perform a maximum likelihood fit of the classification problem.

The final BDT classifiers all consist of 1000 trees with a maximum depth of four. Each node in the trees is required to contain a minimum of 2.5% of the training data set events. This avoids statistical fluctuations from significantly impacting the chosen node splitting criteria, and consequently reduces overfitting. A shrinkage factor, or multiplier of the γ coefficient introduced in Equation 8.1, of 0.1 is used. Shrinkage factors smaller than one are empirically known to improve the generalization of the resulting model³. In every iteration of Equation 8.1 we randomly sample 60% of the training data set to use in the summations. This significantly improves the generalization performance of the algorithm. A decision tree forest trained with random sampling of the training data set in the update step is known as a random forest.

The hyperparameters of the BDT discriminants are chosen based on the performance achieved in the validation data set as mentioned earlier. The performance is quantified by the integral of the receiver operating characteristic (ROC) curve, commonly denoted as area under ROC curve (AUC). The ROC curve evaluates the signal efficiency as a function of the misidentification rate for background events when varying the a threshold on the classifier. The AUC value of a classification algorithm is higher when it achieves a higher degree of separation between signal and background.

After freezing the hyperparameters, the performance and output shape of the classifiers is evaluated in the test data set. No statistically significant differences are found between the output distributions

³It is theoretically not fully understood why this is the case. It is my opinion that this is likely to be related to the fact that learning by means of gradient descent in small steps seems to be an implicit form of regularization that results in good generalization, even when fitting a model with far more parameters than data points. An in depth study of this in the context of deep neural network training can be found in [311].

in the training and test data sets. This indicates that no notable overfitting occurs during the BDT training. The cause for this is the relatively limited parametric size of the BDT classifiers, which contain 1000 small decision trees, compared to the massive training data set of about 17 million leptons available to train each individual BDT.

The BDT training is performed with the TMVA package [312]. As an additional performance cross check similar BDT algorithms are trained in the state of the art XGBOOST library [313]. An equivalent performance to the TMVA training is found. Fully-connected neural networks are also trained and optimized on the lepton classification task using TENSORFLOW [314] with the KERAS [315] interface. The neural networks are found not to give to significant performance benefits in this particular classification problem. This builds additional confidence that the BDT classifiers achieve close to optimal performance on the given set of input variables and training examples.

Fig. 8.3 shows the efficiency in selecting prompt leptons from a mixture of simulated tZq and $t\bar{t}Z$ events as a function of the misidentification rate for nonprompt leptons in a $t\bar{t}$ simulation for continuously varying thresholds on the BDT outputs. A stringent requirement is placed on the BDT output to select leptons for analysis. The requirement has an efficiency of 85 (92)% for correctly identifying prompt electrons (muons) with $p_T > 25$ GeV in simulated tZq events. This corresponds to an increase in identification efficiency of 12 (8)% per electron (muon) compared to the lepton identification strategy in the previous CMS search for tZq production. At the same time, the misidentification rate for nonprompt electrons (muons) is reduced by a factor 2 (8). This improvement in lepton identification is the crucial ingredient in the observation of tZq production presented later in this chapter. The powerful BDT based lepton selection algorithm allows the lowering of the p_T threshold in the lepton selection to 10 GeV in this analysis, down from 25 GeV used in earlier work. In the previous search such a low p_T threshold would have drowned the signal in background events from nonprompt leptons. We expect a far greater tZq signal yield in this analysis than could previously be achieved considering that we target three leptons in the final state and that the efficiency in the selection of each lepton is increased at a given p_T value, while we also manage to drastically reduce the p_T thresholds. At the same time the number of background events with nonprompt leptons is still well below what it was in the previous search, despite the fact that nonprompt leptons predominantly have low p_T .

The lepton selection algorithm developed for this search proved extremely successful. Therefore the decision was made by other CMS analyzers to use it for selecting muons in the search for $H \rightarrow \mu\mu$ decays in the full Run 2 pp collision data set [240]. The same working point as here is used, and the high selection efficiency combined with the rejection of almost all nonprompt leptons was a contributing factor to the analysis seeing the first evidence of the rare $H \rightarrow \mu\mu$ decay mode. The same lepton BDT classifiers were also used in the CMS precision measurement of the $t\bar{t}Z$ production cross section in the 2016 and 2017 pp collision data sets [316].

To enable a data-driven prediction of the background from events with nonprompt leptons with the fake rate method (Section 6.4) a second set of lepton selection criteria are introduced. Leptons passing the BDT selection are called “tight” and this is the nominal selection of the analysis. Leptons that either pass the BDT selection, or fail the criterion on the BDT output but still pass a set of requirements on p_T^{ratio} and $\text{DEEPCSV}_{\text{closest}}$, are labeled “loose”. The loose lepton selection is fine tuned so that the fake rate method precisely predicts the nonprompt lepton background.

8.2.3 Event selection

Events are required to contain three tight leptons, and are vetoed if a fourth loose lepton is found. Events where one or more of the leptons fail the tight selection but pass the loose selection requirements are retained as a sideband used for predicting the background from nonprompt leptons. The selected leptons, ordered by p_T , must have p_T values greater than 25, 15, and 10 GeV. These requirements ensure that all events efficiently pass the trigger requirements.

Events are triggered with a mix of single, two, and three lepton triggers. The efficiency of this trigger cocktail, after applying the specified lepton selection, is measured using a data set collected with orthogonal trigger paths that require a high p_T^{miss} or large amounts of hadronic activity (Section 6.2).

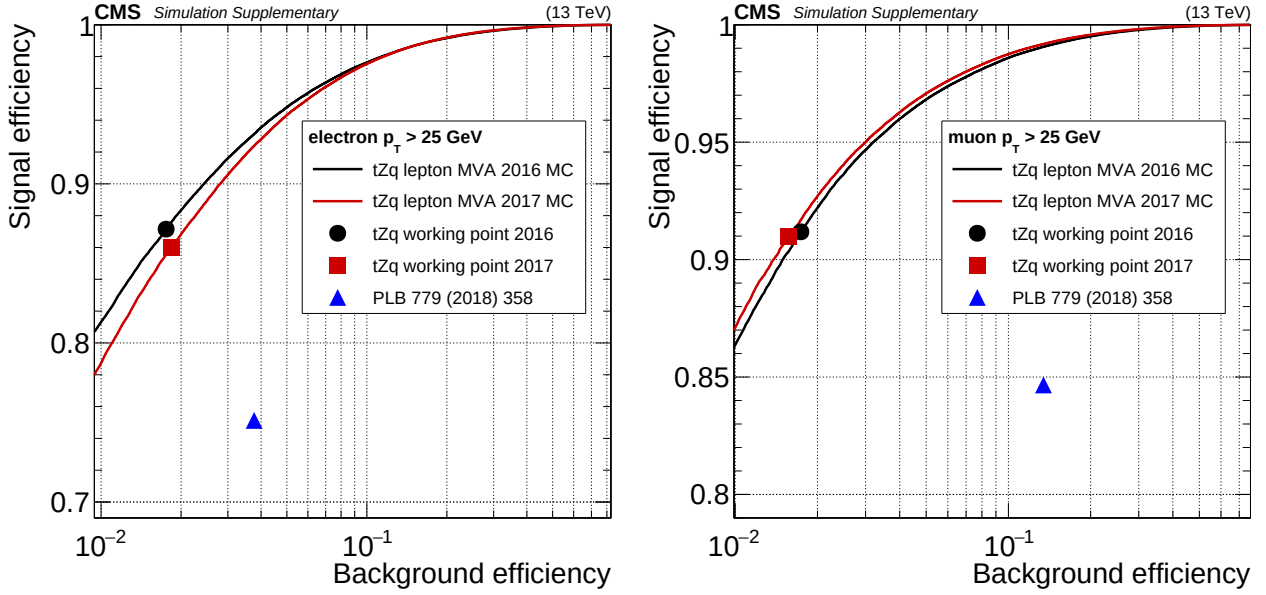


Figure 8.3: Efficiency in identifying prompt leptons from a mixture of simulated tZq and $t\bar{t}Z$ events as a function of the misidentification rate for nonprompt leptons from simulated $t\bar{t}$ events. The left plot contains the results for electrons, while the right plot covers muons. The curves for continuously varying thresholds on the BDT classifier outputs are shown in black and red for the classifiers trained for 2016 and 2017 data-taking conditions respectively. The black circle and red square indicate the selection thresholds chosen in the analysis. The blue triangle represents the selection used in the previous CMS search for tZq production [244]. All leptons are required to have $p_T > 25$ GeV to enable a fair comparison since this was a selection requirement in [244].

The trigger efficiency is found to be compatible with 100% across the kinematic space of the analysis.

Nonresonant $t\ell^+\ell^-q$ production, shown in the bottom right diagram of Fig. 8.1, is considered part of the tZq signal. This component is small compared to resonant tZq production, so we choose to explicitly target events with a Z boson candidate. The three selected leptons must contain at least one OSSF pair with a mass within a 30 GeV window around m_Z . While this requirement reduces the signal efficiency by a minute amount, it significantly lowers the yields of a host of background sources, in particular $t\bar{t}$, $t\bar{t}W$, and $t\bar{t}H$ production, as well as photon conversions.

To target the top quark decay in tZq events at least two jets are required to be present, one of them b -tagged. Events without b -tagged jets, or with less than two jets, have a minuscule signal to background ratio and add no significant signal sensitivity to the analysis. Such events are vetoed. Events with two or more b -tagged jets are kept because the initial state b quark in tZq production typically comes from a gluon splitting into a b quark-antiquark pair. In a nonnegligible fraction of tZq events a second b jet is consequently identified.

8.3 Signal extraction

The selected events are categorized into three distinct search regions, characterized by their jet and b jet content. The set of events with two or three jets, exactly one of them b -tagged (SR-2/3j-1b), contains the largest event yield from tZq production and constitutes the most important event category. Events with four or more jets, only one of them b -tagged (SR-4j-1b), as well as events with two or more b -tagged jets (SR-2b) also have a significant signal to background ratio and are analyzed in parallel.

A distinct BDT classifier is trained and optimized in each event category and data-taking year to maximally separate the tZq signal from the total background. Similar to the case of lepton selection,

discussed in Section 8.2, the discriminants are gradient boosted random forests that do a maximum likelihood fit to the classification task. Note that the BDT classifiers introduced here differ from those used in lepton identification in that they classify entire events. The tZq search and cross section measurements are done by performing a binned maximum likelihood fit to the number of events as a function of the BDT outputs in each event category.

The BDT classifiers are entirely trained on simulated events, including the background contribution from nonprompt leptons. Using simulated events to model the nonprompt lepton background in the BDT trainings prevents overfitting the limited amount of sideband events in data that are used in the actual nonprompt lepton background prediction. Of the simulated data set, half is used for training and half for testing. Each background event is weighted according to the cross section of its parent process so the training data set accurately reflects the background yields expected in the analysis.

For the most part the event-level BDT classifiers share most of the training configuration and hyperparameters of the lepton classifiers. Each classifier is a forest of 1000 decision trees in which each tree has a maximum depth of four. The training of each decision tree and the corresponding update to the boosted forest uses a randomly sampled subset that contains 60% of the training data set events. This smears out statistical fluctuations across the trees to improve the generalization performance. Compared to the lepton classifiers a more stringent limit is placed on the node splitting, reflecting the more limited training data sets: Each leaf in a decision tree must contain at least 5% of the weighted events in the subset of data used to train the tree.

For WZ and ttZ production, the largest background sources, different simulated samples are used during BDT training and in the nominal analysis. This ensures that the prediction of these backgrounds as a function of the BDT output remains unbiased, while we can still utilize the full statistical precision of the simulated samples. No tZq simulation samples with equivalent configurations to the nominal one were available. For the tZq signal and for small background sources with exclusively prompt leptons, the same simulated samples are used during BDT training and analysis. No statistically significant differences between the BDT output distributions in the training and test data sets are seen for any of these processes, so all events are reused for analysis. The statistical uncertainties in the comparison of the output shapes are sufficiently small that any remaining bias that could be caused by reusing events must be negligible compared to the systematic uncertainties that effect the analysis.

A number of distinguishing features of tZq production are used to define the set of input variables on which the BDT classifiers operate. These are described in the rest of this section. First the Z boson decay products are identified by finding the OSSF lepton pair of closest invariant mass to m_Z . Subsequently, the lepton and jet that stem from the top quark decay are identified. The lepton not forming the best Z boson candidate is taken to originate from the W boson in the top quark decay chain. This lepton is called ℓ_W . The jet from the top quark decay is identified by finding the combination of candidate objects, a lepton, a b jet, and a neutrino, that form the invariant mass closest to the top quark mass. The longitudinal momentum of the neutrino from the leptonic W boson decay can be estimated by postulating that the $\vec{p}_T^{\text{miss}}$ in the event is solely induced by this neutrino. One then imposes the requirement that ℓ_W and the neutrino form an invariant mass of m_W , resulting in a quadratic equation with two possible solutions for the neutrino momentum. The invariant masses of each possible combination of a b-tagged jet, ℓ_W , and either of the solutions for the neutrino momentum are evaluated. The b-tagged jet used in the combination giving the mass closest to that of the top quark is hypothesized to originate from the top quark decay.

Studies in simulation show that tZq events very often contain a forward jet recoiling against the top quark and the Z boson. This feature is crystal clear in Fig. 8.17, which displays a tZq event candidate in pp collision data. The forward jet and variables derived from it are the foremost kinematic features for separating tZq events from the background events that remain after the event selection. The jet of highest p_T that is not the b-tagged jet associated to the top quark decay is called the “recoiling jet” in what follows. This naming is chosen because the jet from the final state quark in the tZq matrix element is expected to recoil against the top quark and Z boson system.

A large number of kinematic variables are checked for their discriminating power between tZq and background events. A number of these variables have a clear physical motivation backing them and are found to be useful in the analysis. These variables are listed below:

- Recoiling jet $|\eta|$: The recoiling jet tends to be forward in tZq events whereas it is central in most background events.
- Highest $|\eta|$ among all jets in the event: Signal events almost always have at least one forward jet, background events often only have relatively central jets.
- Maximum invariant mass made up of any pair of jets in the event ($m_{jj}^{\max}$): Because of the presence of a relatively hard forward jet, tZq events usually have far higher $m_{jj}^{\max}$ values than background events.
- Angular separation in terms of ΔR between the b jet from the top quark decay and the recoiling jet: In signal events the recoiling jet typically recoils against the top quark and the Z boson. This results in a large angular separation between the recoiling jet and the top quark decay products.
- Asymmetry of ℓ_W : The asymmetry of ℓ_W is defined as its $|\eta|$ value multiplied by the sign of its charge. The LHC is a pp collider, and the initial state in tZq production consists of a quark and an antiquark at leading order. For these reasons tZq production favors top quarks to top antiquarks, resulting in ℓ_W having an asymmetric charge distribution. Background processes involving top quark pair production or nonprompt leptons have more symmetric charge distributions. Even other charge asymmetric processes such as WZ production have different production mechanisms to tZq and do not have an identical charge asymmetry distribution.
- Highest DEEPCSV value among all jets in the event: In the event selection and categorization we choose a threshold on the b-tagging score above which we consider jets b-tagged. Real b jets generally have a higher DEEPCSV score than misidentified jets that passed the b-tagging requirement. This makes it a powerful variable for separating processes involving one or more top quarks from those without top quarks.
- Angular separation in terms of ΔR between the b jet from the top quark decay and ℓ_W : This quantity is sensitive to the presence of a top quark and tends to be smaller for signal events than for processes without a top quark.
- Scalar sum of the lepton p_T values (L_T) and p_T^{miss} : This variable captures the leptonic momentum scale of the event and provides a degree of separation between tZq and all background processes. It is particularly important for separating tZq production from events with nonprompt leptons that passed the event selection. The latter frequently have low lepton p_T values.
- Scalar p_T sum of all jets in the event (H_T): The hadronic equivalent of L_T , this variable captures the total hadronic momentum scale of the event. Processes involving pair production of top quarks, such as $t\bar{t}Z$, $t\bar{t}W$, and $t\bar{t}H$ have significantly harder H_T distributions than tZq production.
- Number of jets in the event: Pair production of top quarks results in a large number of jets, making this a potent discriminating variable in SR-4j-1b and SR-2b where $t\bar{t}Z$ production is the dominant background source.
- Transverse mass of ℓ_W and $\vec{p}_T^{\text{miss}}$ (M_T): Processes involving a leptonic W boson decay, such as tZq production in the three lepton final state, have an M_T distribution that peaks at m_W and then sharply falls off (Section 6.3.1). Processes which do not include a W boson decay, such as ZZ , $Z\gamma^{(*)}$, and Drell-Yan production in association with a nonprompt lepton have a markedly different M_T distribution.

- Three lepton invariant mass ($M_{3\ell}$): In similar fashion to L_T this variable is sensitive to the total amount of leptonic activity in the event. Note that in the other analyses in this dissertation $M_{3\ell}$ is mainly important for discriminating $Z\gamma^{(*)}$ events from other processes, but this is only a minor background source in this analysis because of the b jet requirement.

In addition to the variables listed above, a number of discriminating variables are found to provide additional discriminating power on a purely empirical basis. The invariant masses, angular separations, and transverse momenta of all combination of the recoiling jet, the jet from the top quark decay, ℓ_W , and the leptons from the Z boson decay are checked for discrimination power between signal and background events. In addition, the minimal and maximal angular separations, invariant masses, and transverse momenta of all combinations of two particles are examined. In this large pool of variables, the following are found to improve the discrimination power of the BDT classifiers when added on top of the variables listed earlier:

- Maximum p_T of any dijet combination in the event ($p_T(j, j)^{max}$).
- Maximum azimuthal separation between any two leptons in the event ($\Delta\Phi(\ell, \ell)^{max}$).
- Minimum angular separation in terms of ΔR between any b-tagged jet and lepton. This variable is strongly correlated with the ΔR between the b jet from the top quark decay and ℓ_W , but is nonetheless found to provide additional sensitivity to the analysis.

Many other variables have discrimination power between signal and background events but do not necessarily further improve the BDT performance. The BDT classifiers can learn complex nonlinear relations between the input variables and might in this way infer the relevant information that would be contained in another input variable. To choose a set of input variables, a performance metric for the BDT classifiers is needed. Like for the BDT classifiers trained for lepton identification, the chosen performance metric is the AUC. As mentioned earlier, half of the available simulated data set is used for BDT training and half for testing the BDT performance. To choose the input variables and hyperparameters only half of the training data, so one fourth of the total data set, is used for actually training the BDT classifiers and the other half for evaluating the results. After the hyperparameters and input variables for a BDT classifier are fixed, it is retrained on the entire training data set. The test data set is only evaluated after freezing the BDT configurations to avoid the statistical bias that might result from fine tuning the BDT configuration to get a better test performance.

The presented set of input variables is selected by iteratively training and evaluating BDT classifiers with one new input variable added at each iteration. This is done separately for each of the three search categories. The variables are ranked in terms of their individual separation power between signal and background. In a first iteration the two most discriminant variables are used for training classifiers, and the performance is evaluated. Then subsequent discriminating variables are added one by one, and each time a new training and evaluation are performed. If the BDT performance does not increase by adding a given variable, it is removed from the list of input variables. The modeling of each input variable is checked in several control regions, introduced in the next section. Poorly modeled variables are removed from the BDT training. Two variables that improve the BDT performance in simulation are found to be unreliably modeled and are removed: The reconstructed top quark mass, and the maximum azimuthal separation between any two jets in the event. The final list of BDT input variables used in each of the analysis search categories is given in Table 8.1.

The distributions of some of the most important discriminating variables in each event category are shown in Fig. 8.4 to Fig. 8.7. It has to be noted that the observed distributions shown in these figures were not unblinded until the entire analysis strategy was finalized. The most important discriminating variables are m_{jj}^{max} and the $|\eta|$ of the recoiling jet. In SR-2/3j-1b the recoiling jet is typically the most forward jet, but in the other search regions this is not the case and the $|\eta|$ of the most forward jet is also among the most important variables.

Table 8.1: List of variables used to build the BDT classifiers that separate tZq production from the total background. Foreach variable the signal enriched categories where it is a BDT input are listed.

Variable	SR-2/3j-1b	SR-4j-1b	SR-2b
$ \eta $ (recoiling jet)	✓	✓	✓
$ \eta $ (most forward jet)	-	✓	✓
$m_{jj}^{\max}$	✓	✓	✓
ℓ_W asymmetry	✓	✓	✓
Highest DEEPCSV value	✓	✓	✓
M_T	✓	✓	✓
$M_{3\ell}$	-	✓	✓
$L_T + p_T^{\text{miss}}$	✓	✓	✓
H_T	-	✓	✓
Number of jets	-	✓	✓
$p_T(j, j)^{\max}$	✓	-	-
$\Delta\Phi(\ell, \ell)^{\max}$	✓	✓	✓
$\Delta R(\ell, b)^{\min}$	✓	✓	-
$\Delta R(\text{b jet from top quark decay, recoiling jet})$	-	✓	-
$\Delta R(\ell_W, \text{b jet from top quark decay})$	-	✓	-

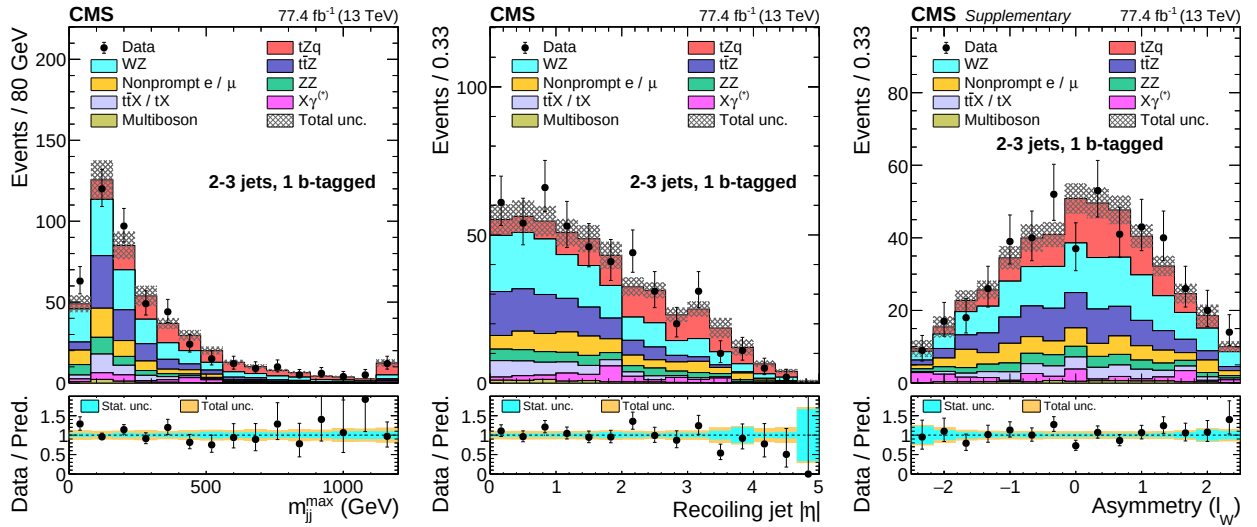


Figure 8.4: Observed and post-fit expected distributions in SR-2/3j-1b of the maximum invariant mass of any pair of jets in the event (left), the $|\eta|$ of the recoiling jet (middle), and the asymmetry of ℓ_W (right).

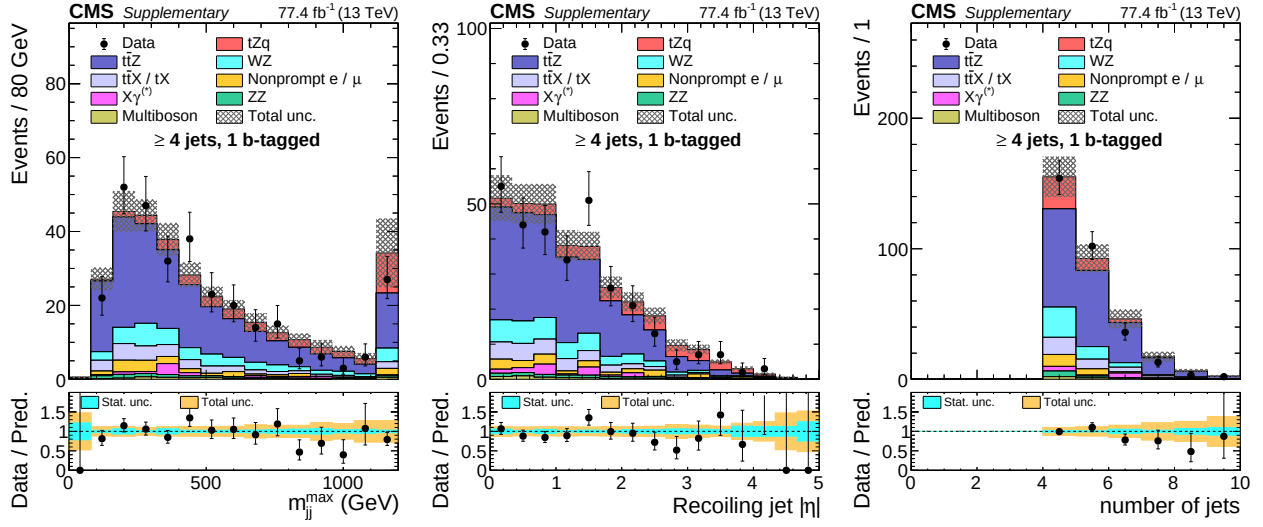


Figure 8.5: Observed and post-fit expected distributions in SR-4j-1b of the maximum invariant mass of any pair of jets in the event (left), the $|\eta|$ of the recoiling jet (middle), and the number of jets in the event (right).

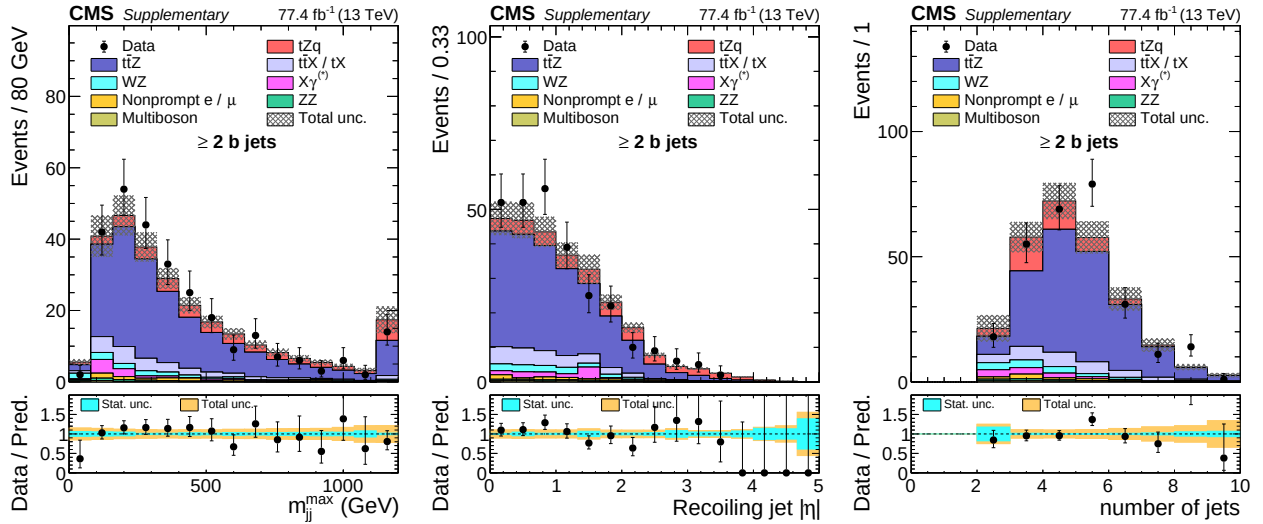


Figure 8.6: Observed and post-fit expected distributions in SR-2b of the maximum invariant mass of any pair of jets in the event (left), the $|\eta|$ of the recoiling jet (middle), and the number of jets in the event (right).

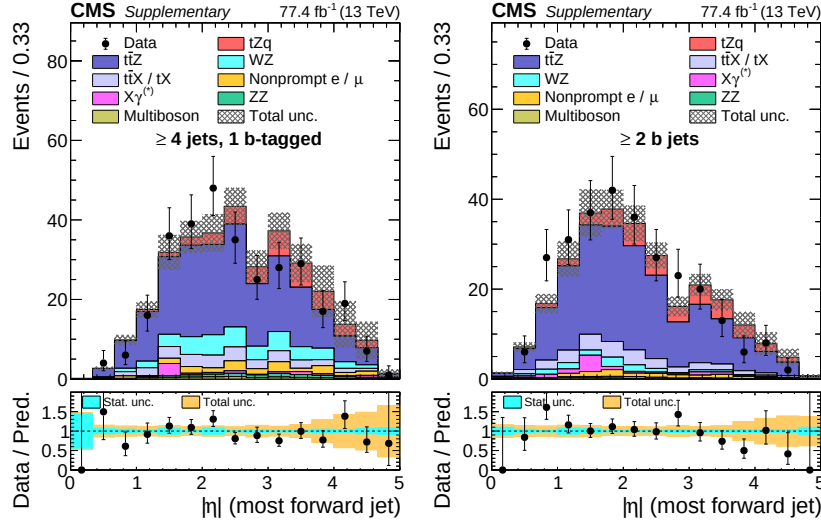


Figure 8.7: Observed and post-fit expected distributions of the highest $|\eta|$ value among all jets in the event in SR-4j-1b (left) and SR-2b (right).

The signal extraction procedure of the analysis consists of a simultaneous binned maximum likelihood fit to the BDT classifier output distributions in each of the search categories. A uniform binning with 15 bins is used in each search category⁴. In general using more bins results in a better sensitivity to the signal as more information is retained in the binned distribution. On the other hand, one must ensure that each bin has a sufficient expected background yield to have a robust background estimation. Using 15 bins is found to be a good compromise between these two objectives.

8.4 Background estimation

An overview of SM processes that contribute to the multilepton final state is given in Section 6.3. The expected event yields of background processes in which all selected leptons are prompt are estimated from simulation. The background contribution from events in which one or more leptons are nonprompt is predicted directly from data with the fake rate method (Section 6.4). In all presented figures we collectively label pair production of top quarks in association with a W or H boson, multiple electroweak bosons, or additional top quarks as $t\bar{t}X$ production. This does not include $t\bar{t}Z$ production, which is considered separately because it is a particularly large background source in this analysis.

In SR-2/3j-1b WZ production gives the greatest background contribution, with $t\bar{t}Z$ production and nonprompt leptons the second and third largest sources of background events. The background composition of SR-2/3j-1b can be seen in Fig. 8.4. Nonprompt leptons form a more important background source than the total corresponding event yield suggests since it is found to be hard to separate this background source from signal events based on kinematic event properties. This was already seen in the previous CMS search for tZq as shown in Fig. 8.2, and is also the case here.

In SR-4j-1b and SR-2b the dominant background source is $t\bar{t}Z$ production. The contribution from nonprompt leptons is much smaller in these event categories. The dominant sources of nonprompt lepton events are $t\bar{t}$ and Drell-Yan production. In $t\bar{t}$ events nonprompt leptons tend to come from the semileptonic decay of one of the b hadrons stemming from the top quark decays. Selected jets

⁴A more optimal result might have been achieved by using a binning scheme that depends on the expected background yield in each bin such as the one used in Chapter 9. The exact output distribution of a BDT classifier is arbitrary in the sense that it depends on the normalization (or equivalently probability priors) of the signal and background events in the training data set. A uniform binning in this semi arbitrary output distribution is unlikely to be optimal. In the limit of a high number of bins the binning scheme becomes irrelevant. Given that the uniform binning leads to several bins that are highly pure in tZq production we conclude that impact on the results of a more refined binning scheme would probably have been minor.

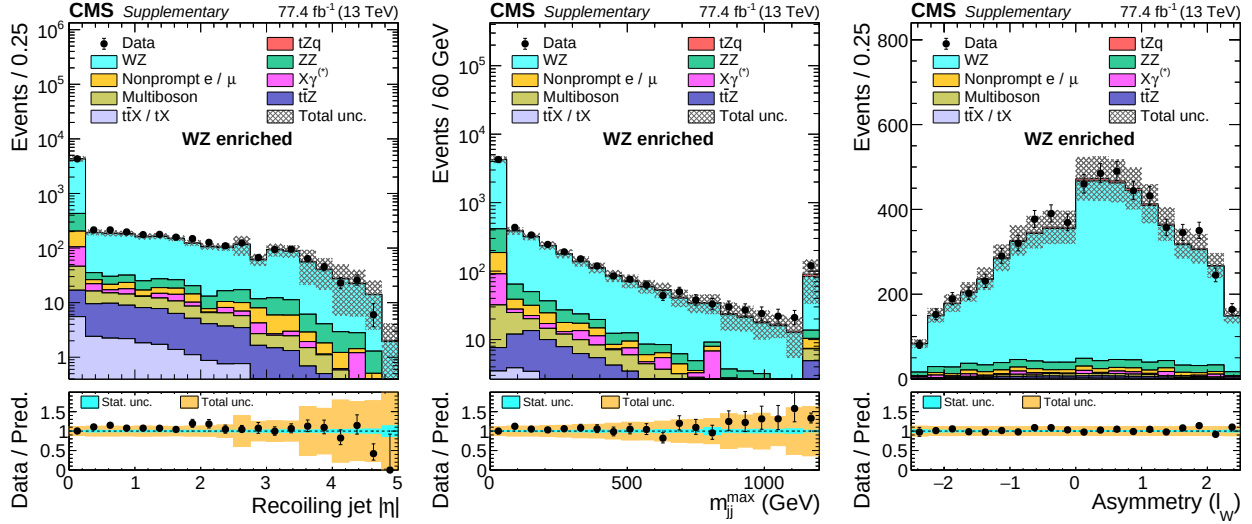


Figure 8.8: Observed and pre-fit expected distributions in the WZ enriched region of the $|\eta|$ of the recoiling jet (left), the maximum invariant mass of any pair of jets in the event (middle), and the asymmetry of ℓ_W (right).

are required not to overlap with any selected lepton, so having two identified b jets in a $t\bar{t}$ event severely reduces the probability to have a nonprompt lepton that passes the selection. Most Drell-Yan production events contain only light quark or gluon jets, and rarely have two jets that manage to pass the b-tagging requirements of the analysis.

Control regions enriched in WZ and ZZ production are used to constrain the normalizations of these processes directly from data, and to verify the accurate modeling of all BDT input variables in simulation. The WZ control region is defined by selecting events that pass the nominal analysis selection criteria, with the requirement on the presence of a b jet inverted. In addition, a requirement of $p_T^{\text{miss}} > 50 \text{ GeV}$ is introduced to reduce the contribution from photon conversions and nonprompt leptons. The ZZ control region is defined by selecting four tight leptons that form two separate OSSF pairs compatible with the Z boson mass. The presence of a b jet is vetoed, and the same lepton p_T requirements as in the rest of the analysis are used. The WZ and ZZ control regions are both included as single bins in the final analysis fit, in addition to the search region events binned in their BDT classification score. Prior uncertainties of 10% are assigned to the normalizations of both the WZ and ZZ processes, and are constrained to smaller values by the control region data.

The distributions of some of the most important input variables to the BDT classifiers used for signal extraction are shown in Fig. 8.8 for the WZ control region and Fig. 8.10 for the ZZ control region. Distributions of the p_T of the leading, subleading, and trailing leptons in the WZ control region are shown in Fig. 8.9. The distribution of the number of jets in both the WZ and ZZ control regions is shown in Fig. 8.11. All analysis variables are found to be well modeled.

There is no dedicated control region enriched in $t\bar{t}Z$ production, but the normalization of this process is implicitly constrained in search region events. Search region events are binned as a function of the BDT output. Bins at low BDT output values in SR-4j-1b and SR-2b are highly enriched in $t\bar{t}Z$ production, the dominant background source in these event categories. These bins are almost depleted of tZq events. This can be seen in Fig. 8.14 and 8.15 when the analysis results are discussed. The $t\bar{t}Z$ enriched search region bins effectively operate as a control region and constrain the normalization of $t\bar{t}Z$ production directly from data.

A control region enriched in nonprompt leptons is defined to verify that the fake rate method gives correct predictions. The same selection criteria as in SR-2/3j-1b are used, aside from the requirement of having a Z boson candidate. The events should either not contain an OSSF lepton pair, or all such pairs should have an invariant mass outside of a 30 GeV window around m_Z . Distributions of m_{jj}^{max} and the $|\eta|$ of the recoiling jet in this control region are shown in Fig. 8.12. The lepton flavor

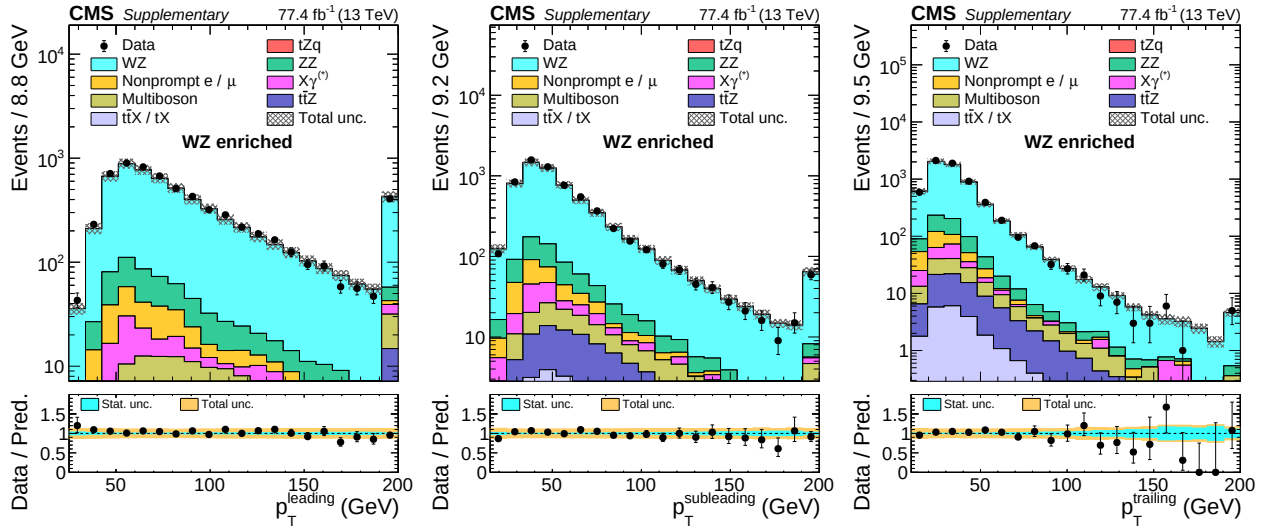


Figure 8.9: Observed and pre-fit expected distributions in the WZ enriched region of the leading (left), subleading (middle), and trailing (right) lepton p_T .

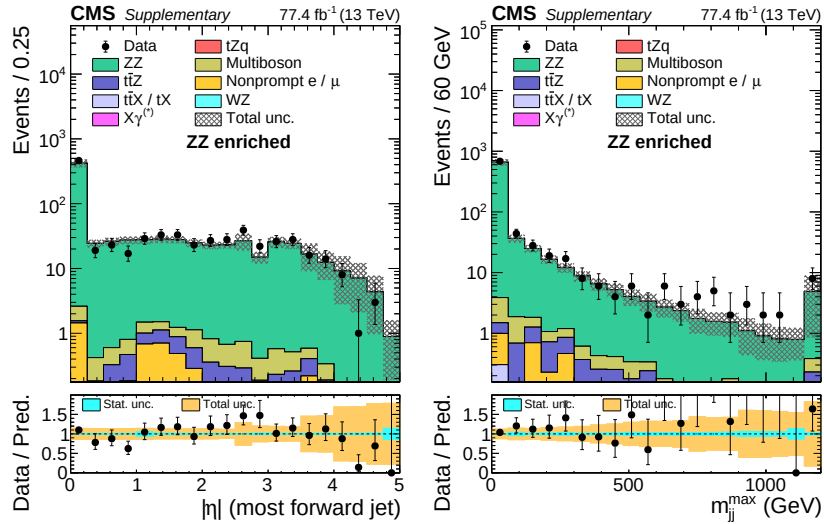


Figure 8.10: Observed and pre-fit expected distributions in the ZZ control region of the highest $|\eta|$ value of any jet in the event (left), and the maximum invariant mass of any pair of jets in the event (right).

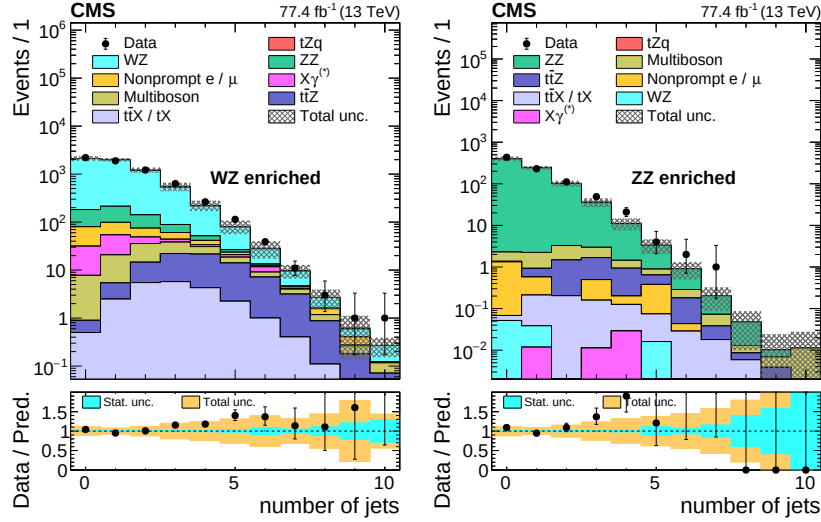


Figure 8.11: Observed and pre-fit expected distributions of the number of jets in the event in the WZ enriched region (left) and the ZZ enriched region (right).

composition of events, and the output value distribution of BDT classifier used for signal extraction in SR-2/3j-1b are shown in Fig. 8.13. The fake rate method is found to give accurate predictions of the nonprompt lepton background within the 30% normalization uncertainty assigned to the method. This 30% uncertainty is based on closure tests in simulated events following the description given in Section 6.4. The purity of the nonprompt control region is lackluster as a result of the power of the BDT based lepton selection. The distributions shown here stand in stark contrast with those of Fig. 7.12 from the previous analysis presented in this work. With a standard I_{rel} based lepton identification similar event selection criteria result in a region dominated by nonprompt leptons.

In addition to the control regions, signal region events depleted of the tZq signal are used to provide an additional verification of the background modeling. In each search category an upper limit is placed on the score of the BDT used for signal extraction. This results in the vast majority of tZq events being vetoed. All of the analysis variables introduced in Section 8.3 and the partial BDT output distributions are checked in the signal depleted versions of the search regions, and in each of the control regions. The full search regions, without an upper limit on the BDT output value, were unblinded after the background modeling was thoroughly verified.

8.5 Systematic uncertainties

A general overview of sources of experimental and theoretical uncertainty affecting the analyses in this dissertation is given in Section 6.5. All the systematic uncertainties listed there are considered here, except for effects from the jet energy resolution (not to be confused with jet energy scale). This section only gives some additional analysis specific details. With a few exceptions all uncertainty sources are treated on an event to event basis. This results in uncertainty sizes that vary across the analysis bins. Variations caused by systematic uncertainties in the integrated luminosity⁵, b-tagging efficiency, trigger efficiency, and jet energy scale are considered to be fully uncorrelated between the data-taking years⁶. Variations due to the lepton selection efficiency are partially correlated between data-taking years, as explained below. All other uncertainty sources are considered to cause variations that are

⁵In reality the uncertainty in the measurement of the integrated luminosity is partially correlated between data-taking years. We refer to Section 6.5.1 for more details. The necessary information to account for the correct correlation in this uncertainty source between the data sets was not available at the time this analysis was performed.

⁶The recommendations by the responsible CMS subgroups were followed at the time of publication. The presented analysis was one of the first to be published that included the 2017 data set. Many of these recommendations were updated for later analyses.

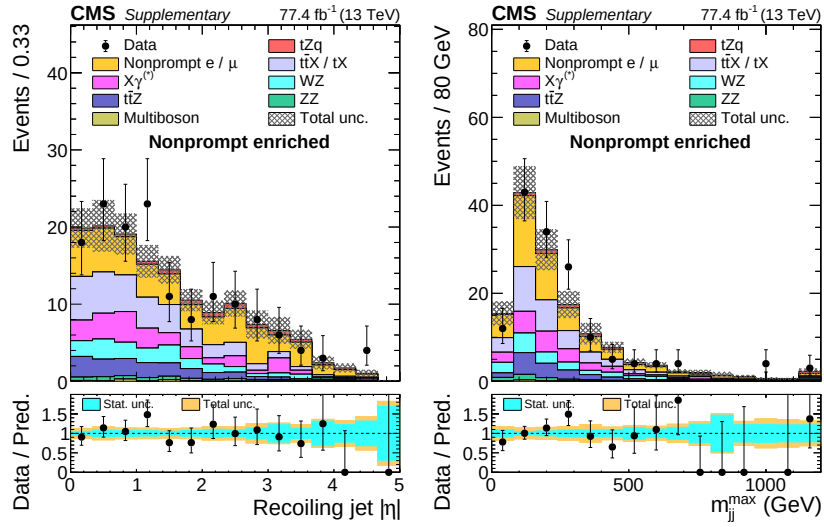


Figure 8.12: Observed and pre-fit expected distributions in the nonprompt lepton enriched region of the $|\eta|$ of the recoiling jet (left) and the maximum invariant mass of any pair of jets in the event (right).

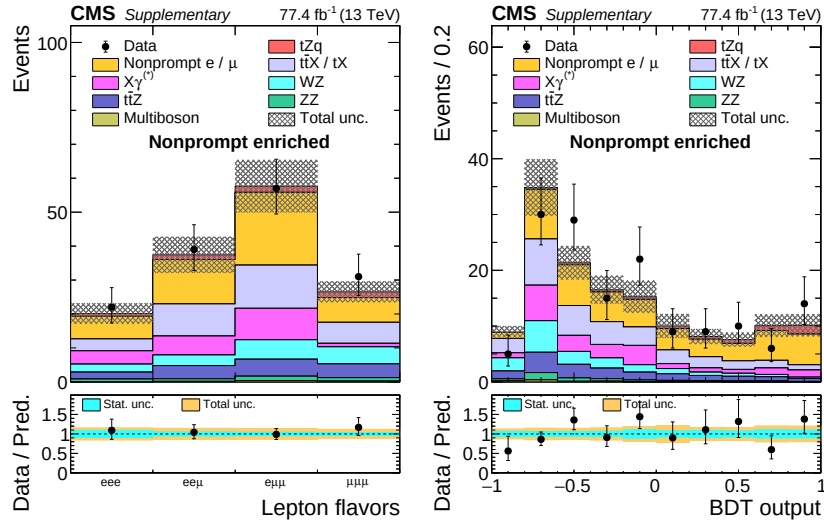


Figure 8.13: Observed and pre-fit expected distributions in the nonprompt lepton enriched region of the lepton flavor composition of the event (left) and the output of the BDT used for signal extraction in SR-2/3j-1b (right).

fully correlated between the data-taking years. Uncertainties in event simulation are also propagated to the subtraction of the prompt contamination to the sideband data events used to predict the nonprompt lepton background. In this way, all of the uncertainty sources affecting simulation also affect the nonprompt lepton background prediction, albeit to a far smaller extent.

When the analysis was published the jet energy scale parametrization and the corresponding uncertainties (Section 5.3) were not finalized for the forward detector region in the 2017 data set. The uncertainties are therefore much larger than what they would be if the analysis was repeated with a more recent version of the jet energy scale parametrization. The large jet energy scale uncertainties are clearly visible in the distributions on the left of Fig. 8.8 and 8.10. The jet energy resolutions were not calibrated yet for 2017 data-taking conditions when the analysis was performed. The effect is therefore not taken into account, but is expected to be minute.

Trigger efficiencies are measured in an unbiased data set collected by triggering on p_T^{miss} and the hadronic transverse momentum in events. The efficiency of the trigger cocktail is found to be compatible with 100% in both data and simulation across the phase space of the analysis. A flat normalization uncertainty of 2% is applied to simulated events, corresponding to the statistical uncertainty in the trigger efficiency measurement at low lepton transverse momenta.

Uncertainties in the lepton identification efficiency are the most important source of experimental uncertainty in this analysis. As such, they are treated with slightly more care than in the other chapters of this dissertation where they are of relatively minor importance. As usual, the lepton identification efficiencies are measured in dilepton events enriched in Z boson production (Section 6.1.3). The systematic and statistical parts of the uncertainty in the measured efficiencies are separately propagated to the analysis results. The statistical uncertainties in the measured efficiencies are uncorrelated between muons and electrons, between the data-taking years, and in different p_T and $|\eta|$ bins. We therefore treat the statistical uncertainties in the lepton identification efficiencies as a large number of separate sources of systematic uncertainty.

The systematic uncertainty in the lepton identification efficiency measurement in $Z \rightarrow \ell\ell$ events is partially correlated⁷ between muons and electrons, and almost fully correlated between data-taking years. Aside from uncertainties in the lepton efficiency measurement itself, uncertainties in the extrapolation of the measured efficiencies to the analysis signal regions should also be considered. Extrapolation uncertainties arise from differences in the density of the underlying event and the probability for accidental overlap between leptons and jets in various processes. These extrapolation uncertainties are expected to be largely correlated between electrons and muons and should be treated as such to not underestimate their impact.

This analysis uses $e\mu$ events enriched in $t\bar{t}$ production to gauge the effect of the extrapolation uncertainties in the lepton identification efficiencies. The efficiency for selecting leptons is measured and compared to a simulation in which the data to simulation scale factors measured in events enriched in Z boson production are applied. Good agreement is observed, except for small discrepancies at low lepton p_T values. These discrepancies however fall within the systematic uncertainties in the lepton selection efficiencies measured in $Z \rightarrow \ell\ell$ events⁸. The analysis signal regions explicitly target $Z \rightarrow \ell\ell$ decays with one additional lepton. We therefore expect differences between per-lepton selection efficiencies measured in $t\bar{t}$ and $Z \rightarrow \ell\ell$ events to be an upper limit on the extrapolation uncertainties of the measured lepton selection efficiencies to the signal regions. We conclude that no additional extrapolation uncertainties are to be applied. The systematic uncertainties in the lepton efficiencies are assumed to be fully correlated between electrons and muons and between data-taking years. This

⁷The extent of this correlation was a hot topic of debate in CMS at the time the work presented in dissertation was performed. The efficiency measurements are based on fitting the mass shape of Z boson candidates. The systematic variations considered in the fits are similar for electrons and muons which should cause some degree of correlation. On the other hand, the energy measurements of electrons and muons, which determine the shape of the reconstructed Z boson mass distribution, rely at least partially on different subdetectors. This last fact could significantly reduce the correlation between the uncertainties in the measured muon and electron identification efficiencies.

⁸It should be noted that the precision of lepton efficiency measurements in $Z \rightarrow \ell\ell$ events has improved in recent analyses. The extrapolation uncertainties are typically no longer covered by them in analysis involving top quarks.

stems largely from the expectation that uncertainties in the extrapolation are largely correlated and that we should ensure our uncertainties are not underestimated.

Theoretical uncertainties from the limited knowledge of the proton PDFs and the choice of the QCD renormalization and factorization scales cause variations in the cross sections and shapes of simulated processes. The shape effects are taken into account for all processes, while there are three exceptions in which the effect on the cross section is not taken into account: For the WZ and ZZ processes a 10% prior normalization uncertainty is used instead of the exact theoretical uncertainty. This eases the interpretation of the control region results and has no impact of any kind on the analysis. The normalization of both processes is measured in data, and the cross section uncertainties get constrained accordingly. The analysis aims to measure the cross section of tZq , and therefore no theoretical normalization uncertainties in the tZq cross section itself are considered in the analysis fit. Note that theoretical uncertainties that affect the expected shape or selection efficiency of tZq production are taken into account.

In the control region used to extract the WZ normalization the presence of a b jet is vetoed. The WZ events that form the main background in SR-2/3j-1b contain a b jet on the other hand, which is not present in the bulk of WZ events. For this reason we consider an additional uncertainty source which accounts for potential mismodeling of the cross section of WZ production in association with real b jets. Simulations indicate that real b jets in WZ events mainly come from gluon splitting. The modeling of gluon splitting is studied by selecting Drell-Yan events with two additional b jets. At a low angular separation between the b jets, these events predominantly originate from gluon splitting. In this regime the Drell-Yan simulation is found to agree with collision data within 20%. The contribution from gluon splitting in simulated WZ events is then varied by 20% and found to induce an 8% variation in the number of WZ events with at least one b-tagged jet. This 8% uncertainty is added as a source of systematic uncertainty in the WZ contribution in the signal regions. Note that the uncertainty due to misidentification of light jets as b jets is also taken into account as part of the set of experimental systematic uncertainties.

Uncertainties in the parton shower simulation are considered for the $t\bar{t}Z$ and tZq processes. The estimation of these uncertainties is briefly explained in Section 6.5.2. For processes other than $t\bar{t}Z$ and tZq production the necessary information to gauge the size of these uncertainties is not available in the simulated samples and no corresponding sources systematic uncertainties are taken into account.

A full list of the systematic uncertainties considered in the analysis and the variations they cause in the expected yields across the signal region BDT bins, is given in Table 8.2. More important than the variations in the expected event yields caused by an uncertainty source is its impact on the measured tZq cross section. Table 8.3 lists the relative uncertainty in the measured tZq cross section resulting from the most important uncertainty sources. Only sources causing uncertainties larger than 1% in the tZq cross section measurement are shown. The impact factor of an uncertainty source is computed by removing it from the analysis fit and reevaluating the measured cross section. The difference between the resulting uncertainty in the cross section measurement and that of the nominal analysis comes from the uncertainty source that was removed. This procedure implicitly assumes that the sources of systematic uncertainty are uncorrelated and therefore only provides an approximation. Details about how the cross section measurement is done are given in the next section.

Table 8.2: Overview of the systematic uncertainties in the analysis. For each uncertainty source the minimum and maximum variations caused in the expected event yields across BDT bins are listed for both data-taking years. The table also includes which processes are affected by a given uncertainty source and whether it affects only the normalization of processes or also their shape.

Source	Affected processes	Range in 2016 (%)	Range in 2017(%)	Treatment
General experimental uncertainties				
Jet energy scale	all	1.5 - 15	1.8 - 38	shape
Unclustered energy	all	0.8 - 7.9	0.4 - 33	shape
Pileup	all	0.7 - 4.3	0.7 - 5.0	shape
Light flavor b-tagging efficiency	all	0.1 - 3.0	0.1 - 4.4	shape
Heavy flavor b-tagging efficiency	all	0.4 - 3.0	0.3 - 4.4	shape
Luminosity measurement	all	2.5	2.3	normalization
Trigger efficiency	all	2	2	normalization
Trigger prefiring	all	0.02 - 1.1	0.07 - 1.8	shape
Lepton reconstruction efficiency	all	0.54 - 1.2	0.7 - 0.9	shape
Lepton identification efficiency systematic	all	2.5 - 4.9	2.7 - 3.4	shape
μ identification efficiency statistical	all	0.2 - 0.9	0.3 - 0.7	shape
e identification efficiency statistical	all	0.3 - 0.8	0.2 - 0.4	shape
Theoretical uncertainties				
QCD scale cross section	$t\bar{t}X$ and multiboson	0.2 - 1.6	0.16 - 1.5	shape
PDF cross section	$t\bar{t}X$ and multiboson	0.1 - 0.9	0.01 - 0.08	shape
QCD scale shape	all	0.6 - 6.8	0.8 - 9.6	shape
PDF shape	all	0.3 - 1.5	0.05 - 0.9	shape
ISR scale	tZq and $t\bar{t}Z$	0.1 - 6.5	0.2 - 6.0	shape
FSR scale	tZq and $t\bar{t}Z$	0.3 - 7.3	0.3 - 6.6	shape
Process specific uncertainties				
WZ normalization	WZ	10 (of WZ)	10 (of WZ)	normalization
ZZ normalization	ZZ	10 (of ZZ)	10 (of ZZ)	normalization
$t\bar{t}Z$ normalization	$t\bar{t}Z$	15 (of $t\bar{t}Z$)	15 (of $t\bar{t}Z$)	normalization
$X\gamma^{(*)}$ normalization	$X\gamma^{(*)}$	10 (of $X\gamma^{(*)}$)	10 (of $X\gamma^{(*)}$)	normalization
WZ + b jet extrapolation	WZ	0.07 - 5.0	0.02 - 4.6	normalization
Nonprompt lepton normalization	nonprompt e/μ	30 (of nonprompt)	30 (of nonprompt)	normalization

Table 8.3: Mean impact of the major sources of systematic uncertainty on the measured tZq signal strength. The computation of the impact of an uncertainty source is done by removing this uncertainty source and reevaluating the measured cross section. When assuming the uncertainty sources to be uncorrelated, the squared impact of an uncertainty source is given by the difference between the squares of the total uncertainty in the measured tZq cross section with and without the uncertainty source in question. Uncertainty sources with an impact factor smaller than 1% are grouped in the “other” category.

Uncertainty	Impact (%)
Experimental	
Lepton selection	3.2
Trigger efficiency	1.4
Jet energy scale	3.3
b-tagging efficiency	1.7
Nonprompt lepton normalization	4.1
$t\bar{t}Z$ normalization	1.0
Luminosity	1.7
Pileup	1.9
Other	1.3
Theoretical	
Final-state radiation	2.0
tZq QCD scale	2.0
$t\bar{t}Z$ QCD scale	1.4

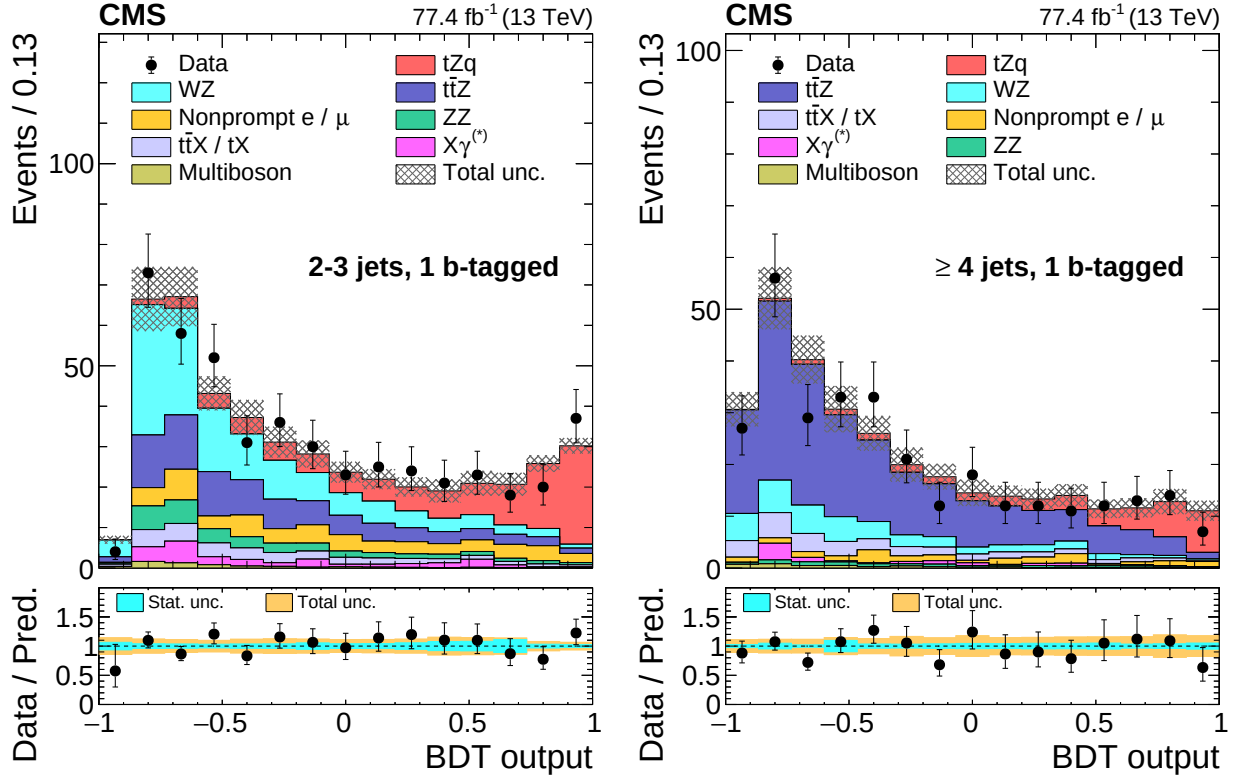


Figure 8.14: Observed and post-fit expected number of events as a function of the BDT used for signal extraction in SR-2/3j-1b (left) and SR-4j-1b (right). A large expected tZq signal, compatible with the collision data, can be seen at high BDT scores.

8.6 Observation and measurement

The observed and post-fit expected yields in SR-2/3j-1b and SR-4j-1b (SR-2b) as a function of the BDT classifier outputs are shown in Fig. 8.14 (8.15). The total observed and post-fit expected number of events in each of the search regions are listed in Table 8.4.

The tZq signal strength, defined as the ratio of the observed to the expected number of tZq events, is measured in a simultaneous binned maximum likelihood fit to the BDT classifier output distributions in the search regions. The WZ and ZZ control regions are each included as a single bin in the fit. The best fit signal strength and the 68% confidence interval are computed with the methods explained in [246]. Systematic uncertainties are included as nuisance parameters in the likelihood function. Sources of systematic uncertainty which exclusively affect the normalization of one or more processes are taken to have log-normal probability density functions. Uncertainty sources whose effect varies bin by bin are incorporated in the likelihood with the template morphing technique [289] with Gaussian

Table 8.4: Overview of the post-fit expected yields for the background and tZq production in each of the signal regions, as well as the observed yields in data. The total uncertainty in the expected yields is given.

Source	2-3 jets, 1 b-tagged SR-2/3j-1b	≥ 4 jets, 1 b-tagged SR-4j-1b	≥ 2 b jets SR-2b	WZ enriched	ZZ enriched
Exp. background	357 ± 34	278 ± 32	228 ± 25	6308 ± 478	847 ± 69
Exp. tZq	103 ± 5.1	38 ± 5.3	37 ± 1.8	67 ± 2.8	0.01 ± 0.004
Total exp.	460 ± 37	316 ± 35	265 ± 25	6375 ± 478	847 ± 69
Observed	475	310	278	6373	852

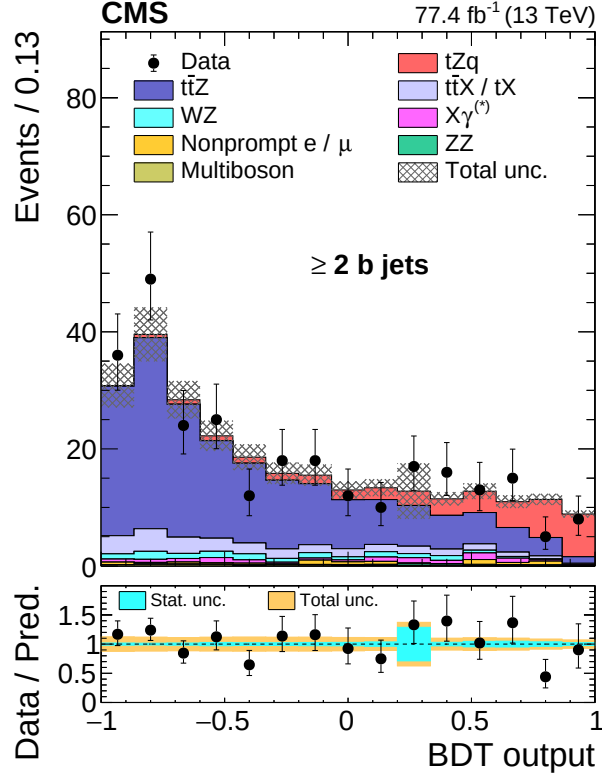


Figure 8.15: Observed and post-fit expected number of events as a function of the BDT used for signal extraction in SR-2b. A large expected tZq signal, compatible with the collision data, can be seen at high BDT scores.

probability density functions. The Barlow-Beeston method [290] is used to account for the finite size of simulated samples and of the sideband data set used in the prediction of the nonprompt lepton background.

A large excess of events over the background-only hypothesis can be seen in Fig. 8.14 and 8.15. To quantify this excess we define the following test statistic:

$$q = -2 \ln \left[\frac{L(\text{data} | r = 0, \hat{\theta}_0)}{L(\text{data} | r = \hat{r}, \hat{\theta})} \right], \quad (8.3)$$

in which L represents the likelihood of the observed data, θ the set of nuisance parameters, r the tZq signal strength, and the notation $\hat{x}$ is used to represent best-fit values of different parameters obtained by maximizing the likelihood of the observed data. The likelihood in the numerator is determined under the hypothesis of a zero signal strength, and $\hat{\theta}_0$ represents the best-fit nuisance parameter values under this assumption. The significance of an excess is given by the p -value of the test statistic of the observed data. This is the probability that a test statistic value larger or equal to the observed one occurs under the background-only hypothesis. To estimate the p -value of the observed test statistic, one must generate MC pseudoexperiments to model the expected test statistic distribution. When generating the pseudoexperiments, the nuisance parameter values are frozen to their best-fit values under the background-only assumption and in each pseudoexperiment the corresponding constraint terms are randomized when evaluating the likelihood. The expected test statistic distribution obtained by generating 510 million pseudoexperiments is shown in Fig. 8.16. The observed test statistic value is so large that it exceeds all values found in the generated pseudoexperiments by a large margin. In other words, the excess of events over the background-only hypothesis is of such significance that it can not be quantified without making approximations. The background-only hypothesis is conventionally

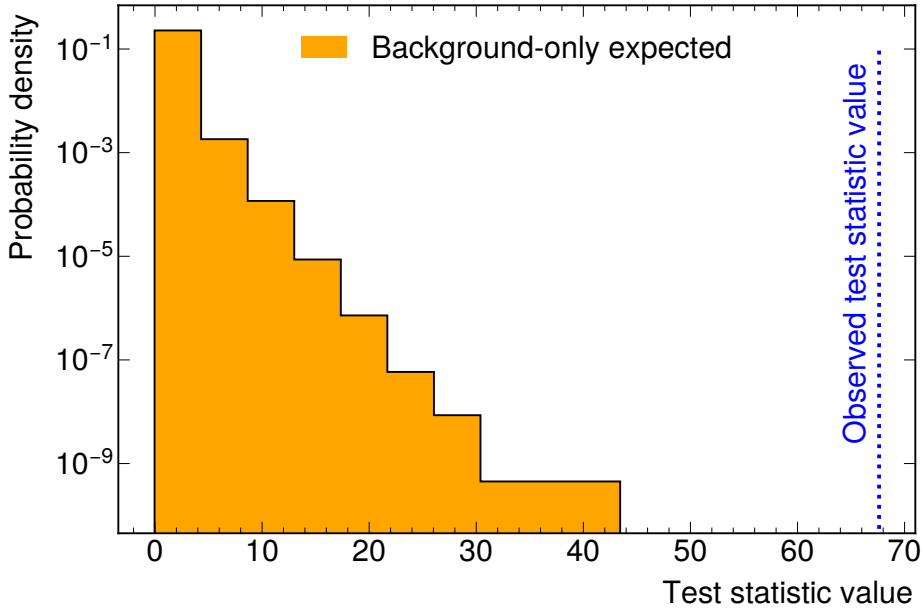


Figure 8.16: Expected distribution of the test statistic from Equation 8.3 under the background-only hypothesis compared to the observed test statistic value. A total of 510 million MC pseudoexperiments were generated to compute the expected test statistic distribution, and it seems that orders of magnitude more would be needed before any pseudoexperiment reaches the observed test statistic value.

considered rejected when the p -value for the observed test statistic is smaller than 2.87×10^{-7} . This corresponds to the probability of drawing a value from a Gaussian distribution that is at least 5 standard deviations larger than the expectation value. The excess shown here is far more significant than that.

The asymptotic approximation of the profile likelihood test statistic [291, 292] is used to compute the significance of the tZq signal. The observed (expected) significance is found to be 8.2 (7.7) standard deviations over the background-only hypothesis. A significance of 8 standard deviations is approximately equivalent to a p -value of 10^{-15} , so more than 10^{15} MC pseudoexperiments would have to be simulated to compute the exact significance. Analyzing the 2016 and 2017 data sets separately results in observed (expected) significance values of 7.2 (5.7) and 5.4 (6.0) standard deviations above the background-only hypothesis. Here we also used the asymptotic approximation of the test statistic. We conclude that tZq production is observed with a significance well over five standard deviations. In addition, the process is observed in the 2016 data set alone, in which it remained elusive in an earlier CMS search. The result presented here is the first ever observation of the tZq process.

The tZq cross section is calculated from the fitted signal strength relative to the SM expectation computed at NLO in QCD with MADGRAPH5_aMC@NLO. The cross section is found to be:

$$\sigma(pp \rightarrow tZq \rightarrow t\ell^+\ell^-q) = 111 \pm 13(\text{stat})_{-9}^{+11}(\text{syst}) \text{ fb}, \quad (8.4)$$

for ℓ an electron, muon, or τ lepton and dilepton invariant masses above 30 GeV. The dilepton invariant mass requirement is a fiducial cut used in the calculation of the reference cross section. The abbreviations “stat” and “syst” refer to the statistical and systematic uncertainty components. The corresponding signal strength is:

$$\mu = \frac{\sigma(pp \rightarrow tZq \rightarrow t\ell^+\ell^-q)}{\sigma^{\text{SM}}(pp \rightarrow tZq \rightarrow t\ell^+\ell^-q)} = 1.18_{-0.13}^{+0.14}(\text{stat})_{-0.10}^{+0.11}(\text{syst}) \pm 0.04(\text{theo}), \quad (8.5)$$

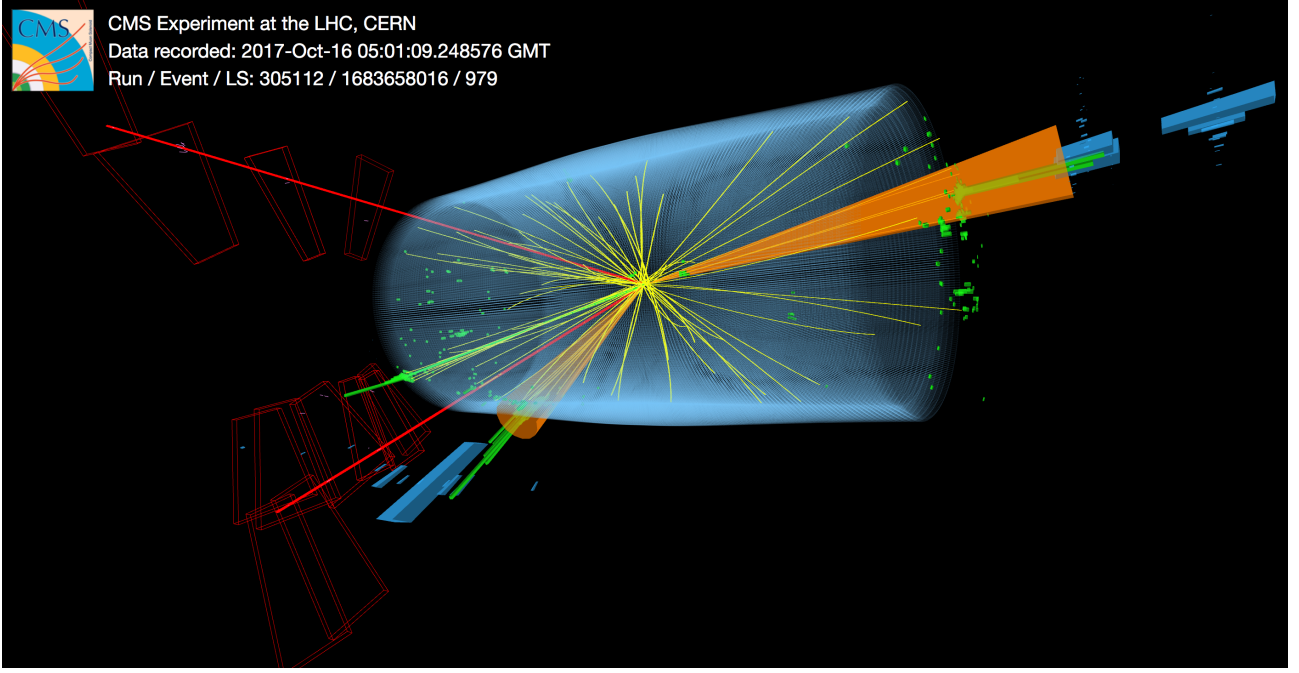


Figure 8.17: Display of the SR-2/3j-1b event with the highest BDT score in 2017 collision data. The high BDT score signifies that this event is highly likely to originate from tZq production. Electrons (muons) are drawn in green (red), and the two muons form a Z boson candidate. Jets are represented by orange cones. The jet on the right has high a $|\eta|$ value and recoils against the rest of the event. The jet on the left passes the b-tagging requirements of the analysis.

where “theo” stands for the effect of the theoretical uncertainty in the expected tZq production cross section. The measured cross section is compatible with the SM expectation. The signal strength values obtained by analyzing the 2016 and 2017 data separately are:

$$\mu_{2016} = 1.36^{+0.22}_{-0.20}(\text{stat})^{+0.14}_{-0.12}(\text{syst}) \pm 0.04(\text{theo}) \quad (8.6)$$

$$\mu_{2017} = 1.03^{+0.18}_{-0.17}(\text{stat})^{+0.14}_{-0.12}(\text{syst}) \pm 0.03(\text{theo}). \quad (8.7)$$

Both measurements are compatible with each other and with the expectation of the SM. The systematic uncertainty is relatively larger in the 2017 data set because the jet energy scale corrections available at the time of analysis were preliminary and had larger associated uncertainties. The combined cross section measurement has a relative uncertainty of 15%, more than twice the precision achieved in any previous measurement. Before this result the most precise measurement of the tZq production cross section came from the ATLAS and CMS searches in the 2016 pp collision data sets. Both of these obtained a relative uncertainty of around 35% in the tZq production cross section. In the 2016 data set alone this analysis has a relative uncertainty of just under 20%, demonstrating the considerable improvements in analysis techniques that were made.

Events at extremely high BDT classification scores are likely tZq candidates. To visualize tZq production, event displays are made of the SR-2/3j-1b events that have the highest BDT score in the 2016 and 2017 collision data sets. The top scoring 2017 collision event is shown in Fig. 8.17, while the 2016 event is shown from different angles in Fig. 8.18 to 8.20.

The tZq process is exceptionally sensitive to the presence of new physics. Small modifications to the SM might result in the top quark and Z boson in tZq production events having a far harder p_T spectrum than expected in the SM. To investigate this possibility, the p_T spectrum of the Z boson in tZq events is studied in the three search region categories. To achieve a high purity in tZq production a stringent requirement that the BDT classifier output is larger than 0.5 is applied in each search

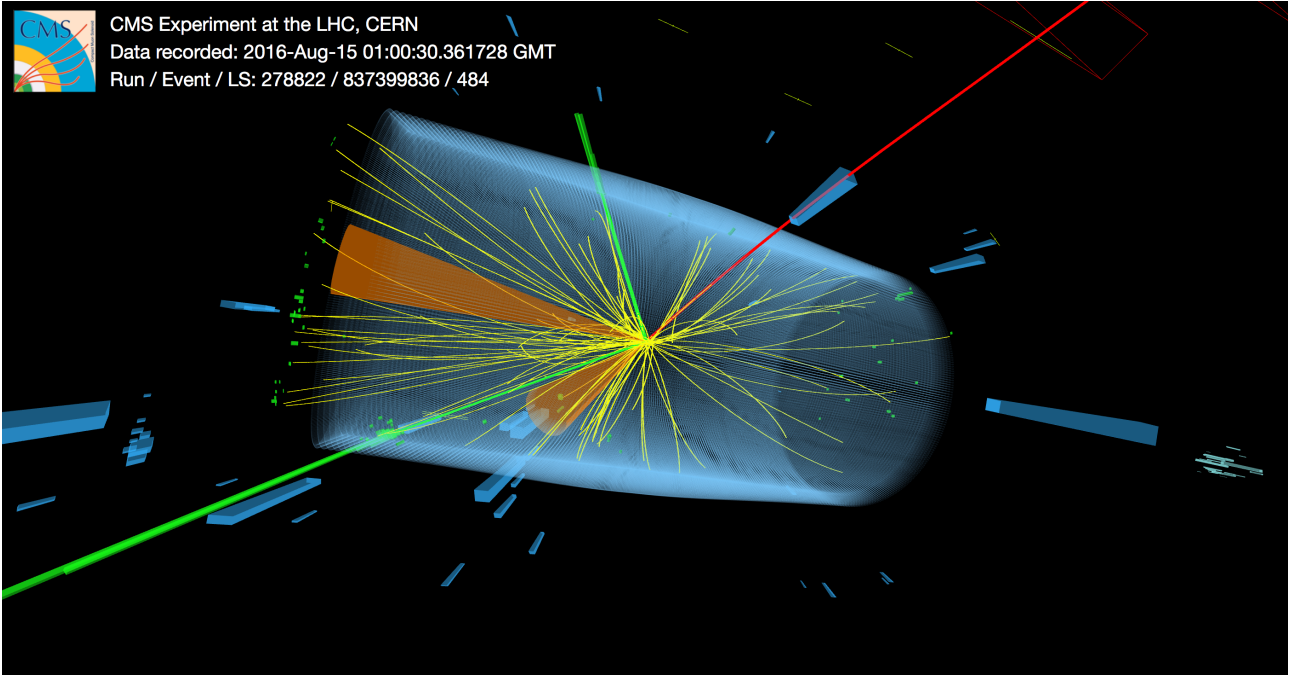


Figure 8.18: Display of the SR-2/3j-1b event with the highest BDT score in 2016 collision data. The high BDT score signifies that this event is highly likely to originate from tZq production. The contents of the figure are as described in the caption of Fig. 8.17. The two electrons form a Z boson candidate, one jet has a high $|\eta|$ value, and the central jet passes the b-tagging requirements.

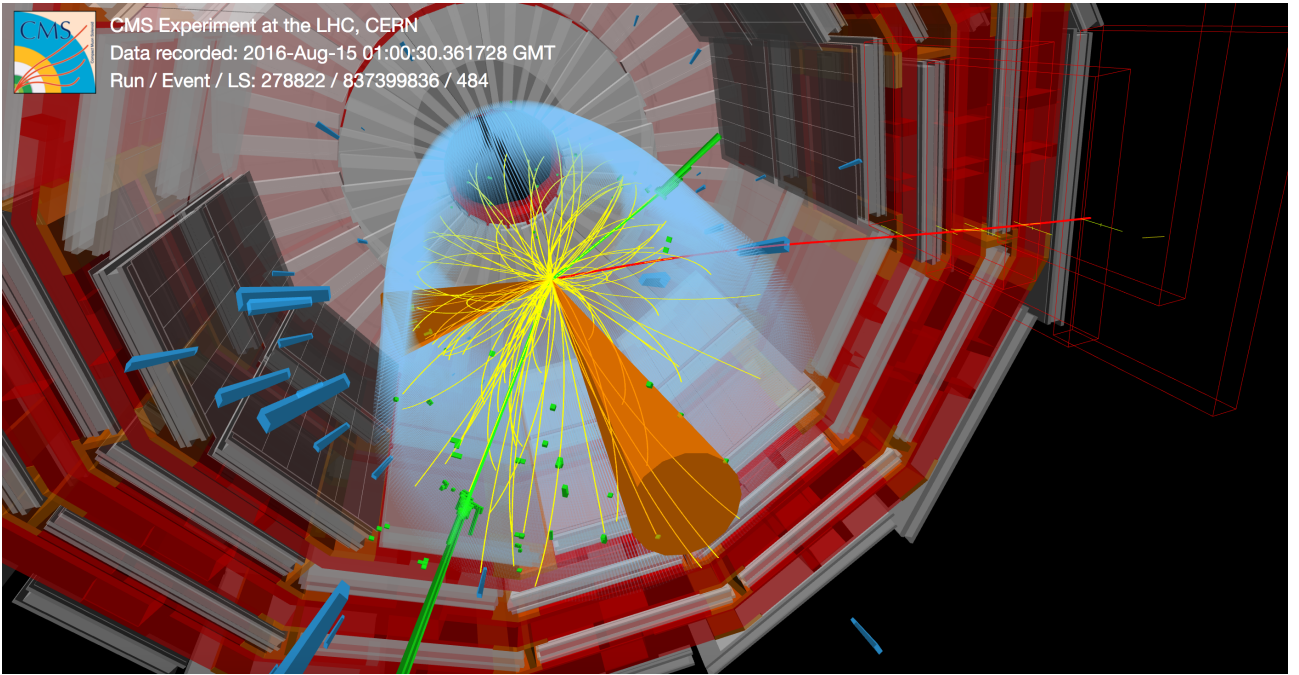


Figure 8.19: Display of the SR-2/3j-1b event with the highest BDT score in 2016 collision data. This is a display of the same event as in Fig. 8.18 seen from a different angle, and an explanation of the contents is given in the corresponding caption.

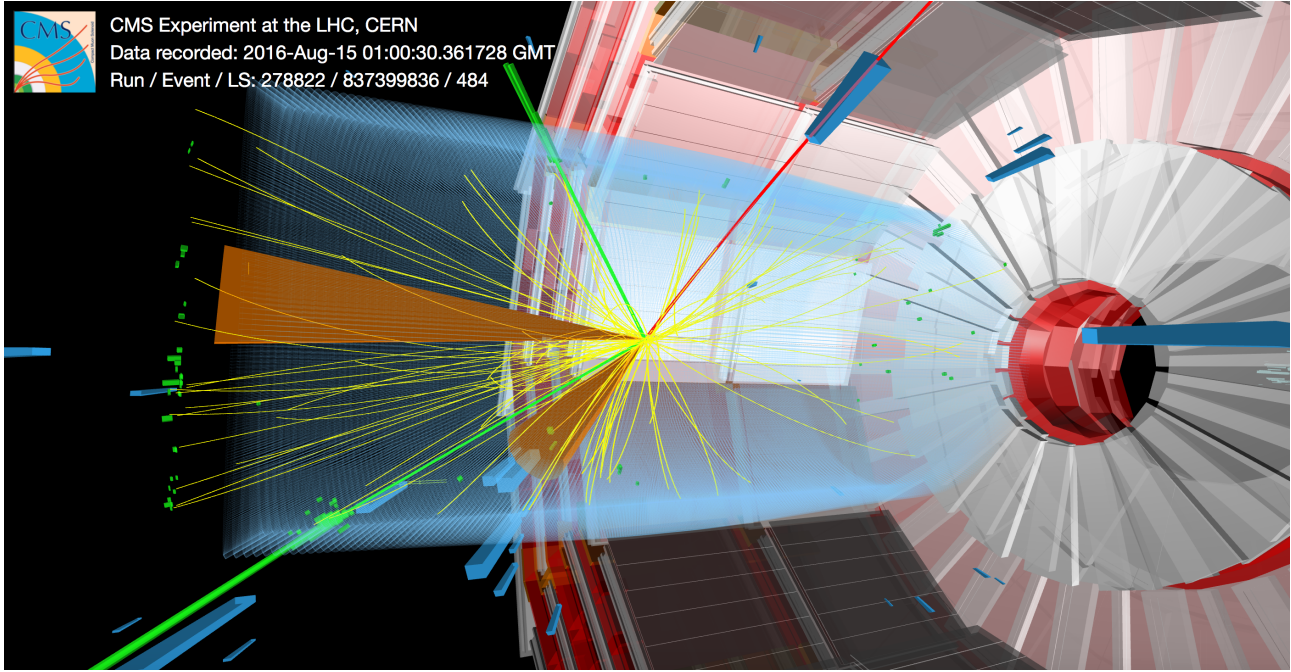


Figure 8.20: Display of the SR-2/3j-1b event with the highest BDT score in 2016 collision data. This is a display of the same event as in Fig. 8.18 seen from a different angle, and an explanation of the contents is given in the corresponding caption.

category. The results are shown in Fig. 8.21 for SR-2/3j-1b, in Fig. 8.22 for SR-4j-1b, and in Fig. 8.23 for SR-2b. The p_T spectrum of the Z boson is found to agree well with the SM expectation. The statistical precision is limited by the rarity of tZq production. Future studies of this p_T spectrum in a larger data set are needed. The aforementioned figures also show the lepton flavor composition in tZq enriched events and the $|\eta|$ distribution of the recoiling jet. In each case the observed distributions match the SM expectation well.

8.7 Summary and outlook

A search for single top quark production in association with a Z boson (tZq) was presented. The tZq process is an extremely rare SM process with an exceptional sensitivity to deviations from the SM. Given its rarity, the process remained unobserved prior to this search. Out of all the possible tZq decay modes, the one in which both the top quark and the Z boson decay to leptons is the most promising for an observation and accurate measurement of tZq production in the current data set. Other decay modes have higher branching fractions, but orders of magnitude more background events. We therefore search for tZq in events with three light leptons, electrons or muons. The analyzed data set consists of 13 TeV pp collisions collected with the CMS detector throughout 2016 and 2017. The total integrated luminosity of the data set is 77.4 fb^{-1} .

We identify an aspect of earlier searches for tZq production that can be drastically improved: The lepton identification algorithms used in those searches are not satisfactory, resulting in a large background from nonprompt leptons. A machine learning based lepton identification algorithm tailored for the identification of leptons from tZq decays is introduced. Compared to the lepton identification strategy used in the previous CMS search for tZq production, this algorithm increases the identification efficiency by 12 (8)% per signal electron (muon) with $p_T > 25 \text{ GeV}$, while reducing the background rate of nonprompt electrons (muons) by a factor 2 (8). The new lepton identification algorithm has the added bonus that it allows one to introduce far lower lepton p_T thresholds in the event selection than those used in earlier work. This further increases the signal selection efficiency.

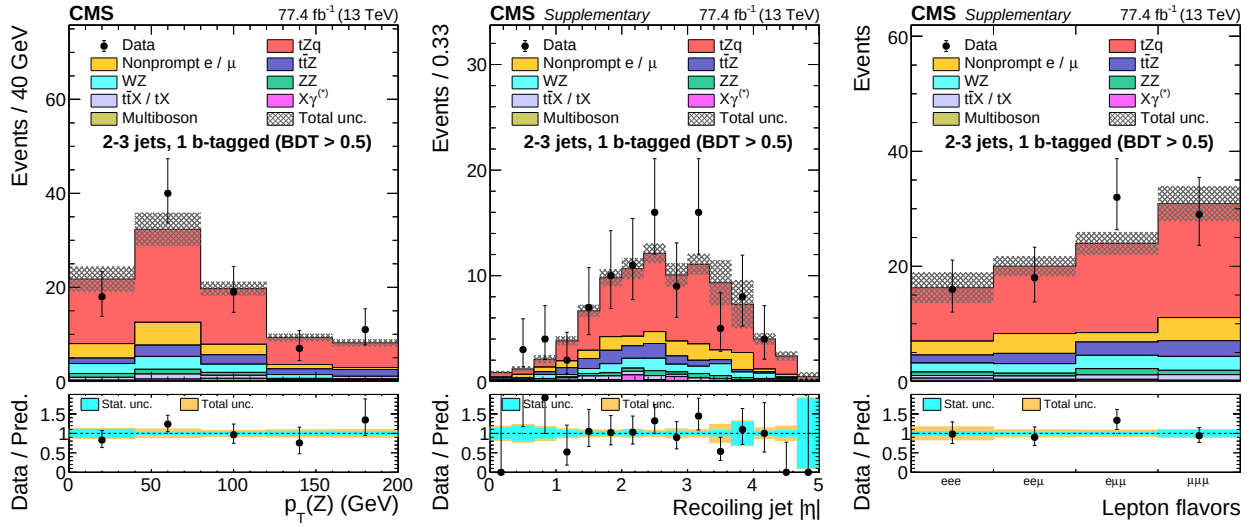


Figure 8.21: Observed and expected yields for events in SR-2/3j-1b with BDT scores in excess of 0.5, as a function of the p_T of the reconstructed Z boson candidate (left), the $|\eta|$ of the recoiling jet (middle), and the lepton flavor content of the event (right).

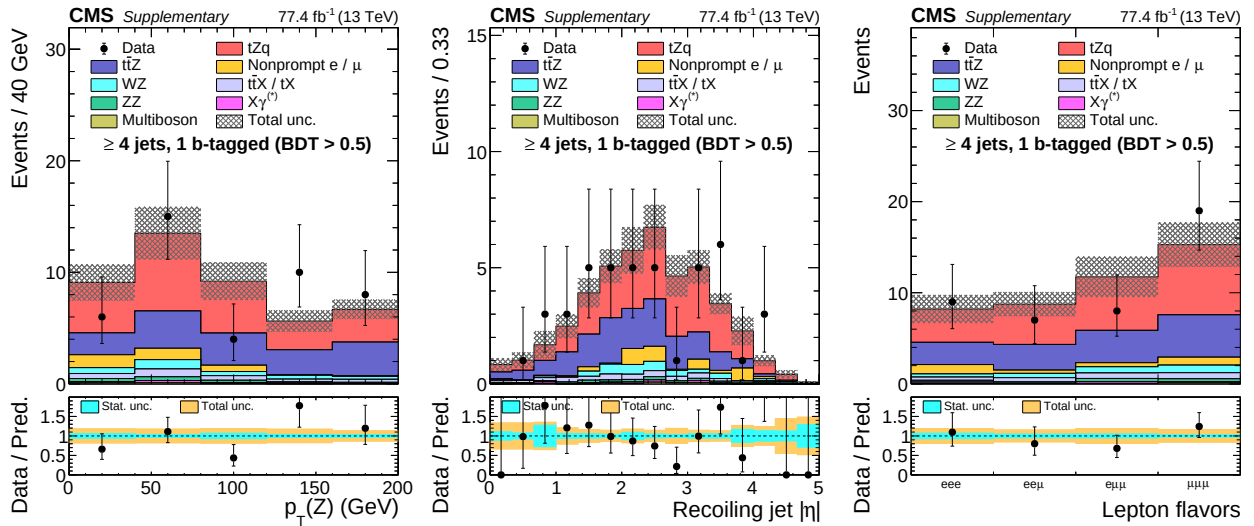


Figure 8.22: Observed and expected yields for events in SR-4j-1b with BDT scores in excess of 0.5, as a function of the p_T of the reconstructed Z boson candidate (left), the $|\eta|$ of the recoiling jet (middle), and the lepton flavor content of the event (right).

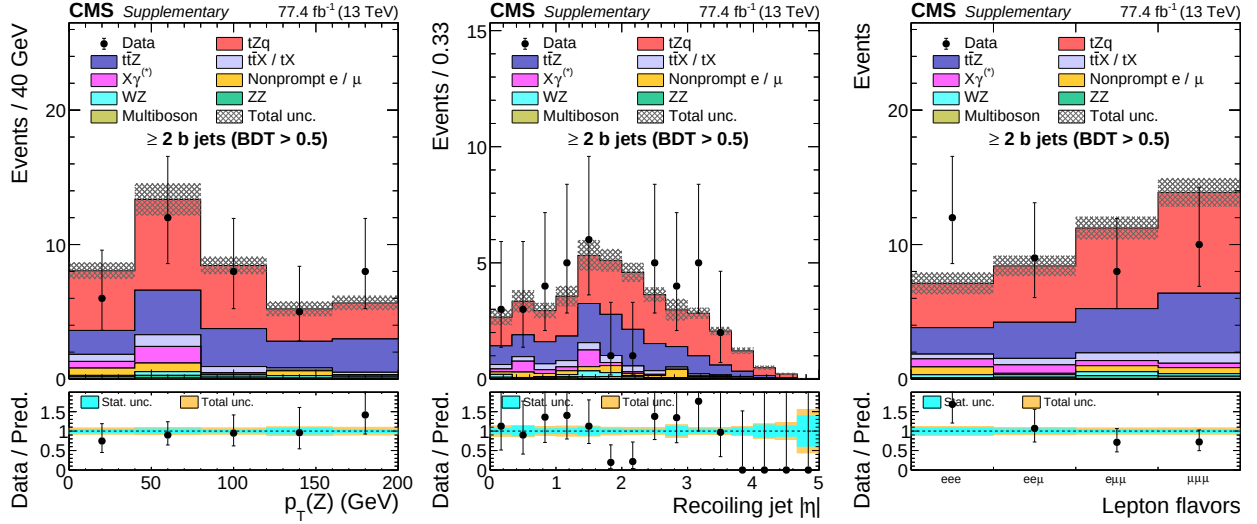


Figure 8.23: Observed and expected yields for events in SR-2b with BDT scores in excess of 0.5, as a function of the p_T of the reconstructed Z boson candidate (left), the $|\eta|$ of the recoiling jet (middle), and the lepton flavor content of the event (right).

The general analysis strategy is redesigned compared to earlier search iterations. Events are categorized into three search categories defined by their jet and b jet content, and several control regions are introduced to constrain the rates of background processes. One of the three search categories is made up of events with four or more jets, one of them b-tagged. This event topology has a significant signal sensitivity but was not analyzed in earlier searches. In each of the search categories a BDT classifier is trained to identify events from tZq production, and the set of input variables to these BDT classifiers is entirely reoptimized compared to earlier work.

The analysis of the data is done by simultaneously fitting the BDT distributions in the search categories and the event yields in the control regions. An excess over the background-only hypothesis is observed, and its significance is so large that it can not be quantified without making approximations given the available computational resources. Using the asymptotic approximation of the profile likelihood test statistic, we find an observed (expected) significance of 8.2 (7.7) standard deviations over the background-only hypothesis. Separately analyzing the 2016 and 2017 collision data sets yields observed (expected) significances of 7.2 and 5.4 (5.7 and 6.0) standard deviations. We conclude that tZq production is unambiguously observed for the first time.

The total cross section $\sigma(pp \rightarrow tZq \rightarrow t\ell^+\ell^-q)$ is measured to be $111 \pm 13(\text{stat})_{-9}^{+11}(\text{syst})$ fb for dilepton invariant masses exceeding 30 GeV, with “stat” and “syst” referring to the statistical and systematic uncertainty components. This is compatible with the SM expectation. Separate analyses of the 2016 and 2017 collision data sets give consistent results. The relative uncertainty of 15% in the tZq cross section measurement is more than twice lower than in any previous measurement.

The kinematics of tZq production events are highly sensitive to modifications to the SM by new physics, even if direct production of the involved particles would be out of range of the LHC. The distribution of the Z boson p_T is studied in very tZq -like signal region events, and the shape is found to be compatible with the SM expectation.

Several improvements to the current analysis could still be made, but the low hanging fruits have been picked. Some minor improvements would be to use the DEEPIET algorithm (Section 5.4) for identifying b-tagged jets, which was not available at the time the presented results were obtained. Deep neural networks could be used in the identification of leptons, though preliminary studies indicate that improvements would be minor compared to the current BDT based strategy. The BDT classifiers used to identify tZq events might also be replaced by a neural network, and instead of separate classifiers in each event category a single classifier could be used. Several sources of systematic uncertainty might be reduced, the most obvious being the uncertainty in the jet energy scale. The analysis presented here

was one of the first published using the 2017 pp collision data set, and the jet energy scale corrections for this data set were not finalized at the time. As a consequence, the uncertainties are significantly larger than they would be nowadays.

A new analysis by the CMS Collaboration using the full Run 2 pp collision data set has recently been published [245]. The work was led by UGent researchers, and relied heavily on software frameworks I wrote for this analysis and the one presented in Chapter 9. My direct involvement was limited to tips and discussions. The analysis strategy of this new tZq measurement is a slightly modified version of the one presented here, and with a data volume roughly 80% greater than what was available in the presented measurement, a relative uncertainty of 12% in tZq production cross section is obtained. Extrapolating the presented analysis to the same data volume yields a relative uncertainty of roughly 13%, showing that improving the uncertainty in the total tZq production cross section measurement is challenging at this point. Meanwhile, the ATLAS Collaboration achieved a relative uncertainty of 15%, the same as the presented search, by analyzing the full Run 2 pp collision data set [308]. In recent measurements of the total tZq production cross section the size of the statistical uncertainty has become comparable with the total systematic uncertainty. Increasing the collision data volume further will give quickly diminishing results without improvements to the analysis techniques.

An important area where the presented analysis can be extended and where an increased data volume would bring great benefits is in the measurement of the kinematic distributions of tZq events. A differential measurement is in order and has been performed in the full Run 2 CMS analysis mentioned above [245]. All distributions are found to be compatible with the SM expectation. As the data volume increases, so will the precision of such differential measurements, and surprises might yet be revealed.

8.8 Acknowledgements

This analysis was largely my effort, though I received help from others in some parts of the work. All internal CMS documentation on the analysis is my writing, and I was the main editor of the paper and the author of the initial drafts. I gave all internal CMS presentations during the publication process of these results, and was the contact person of the analysis. I designed the analysis strategy from beginning to end, including the improvements to the lepton identification strategy. Here it must be noted that existing analyses gave a good basis for starting, and I had the benefit of hindsight to see the limitations of those analyses. Minor changes were also made during the publication process following insights and comments from other people. All of the presented results, values, and figures related to this analysis are produced by me.

In two important segments of the analysis my colleagues Tom Cornelis and Illia Khvastunov helped. Tom measured the lepton identification efficiencies in collision data. These are usually readily available, but not in this case where novel BDT classifiers were used for the selection of leptons. The efficiency measurements are absolutely crucial for the practical usage of the BDT based lepton selection algorithm. Illia measured the lepton fake rate in collision data, which is used for the estimation of the background from nonprompt leptons.

The role of my supervisor Didar Dobur was as important as ever in this analysis, given that for the most part I was the sole analyzer in this search. She was the main person with whom I could bounce ideas about the analysis when needed. Her advice was particularly important to navigate the complex publication procedure. Given the nature of the analysis, an observation of a previously undiscovered process which also strongly improved upon earlier CMS results, it was often difficult to find internal agreement in the CMS Collaboration on the manner in which the results should be presented.

Two more people deserve to be mentioned for their help in the publication process of this work: Louise Skinnari and Greg Landsberg. Louise was the chair of the analysis review committee for this publication. She went above and beyond to facilitate a quick approval of the results and the corresponding paper. Greg worked as the language editor of the paper, and sped up the writing process significantly. He also put in great effort to support the timely publication of the results.

Chapter 9

Search for charginos and neutralinos

*“They sought one to strike
one to knock the great bull down
in Karelia the fair
on the great farms of Finland
in Russia’s mild land
in the bold land of Sweden
in Lapland’s broad distances—
the mighty land of Turja;
they sought one in Tuonela
the Dead Land, underground too—
they sought but they did not find
they searched but they did not see.”*

Elias Lönnrot, *Kalevala*

9.1 Introduction

Several open problems in fundamental physics can be addressed by supersymmetry, one of the most popular extensions of the SM. Supersymmetry is a symmetry under transformations that turn bosonic states into fermionic ones and vice versa. The addition of such a symmetry to the SM results in the prediction of the existence of a host of new particles, “superpartners” of the SM particles [317–321].

The problem of the quadratic dependency of the squared H boson mass on the ultraviolet energy scale, the hierarchy problem (Section 2.3.6), is naturally solved in supersymmetry. The mass correction terms stemming from bosons and fermions have opposite signs. The introduction of superpartners therefore resolves the mass divergence of the H boson. Dark matter can be accounted for in supersymmetric models by introducing the conservation of R-parity, a multiplicative quantum number that is positive for SM particles and negative for superpartners [119]. Consequences of R-parity conservation are that superpartners are always produced in pairs, and that the lightest supersymmetric particle (LSP) is stable. The decay chain of every heavy supersymmetric particle ends with an LSP. The LSP is usually predicted to be weakly interacting. Most often it is the lightest neutralino, a mass eigenstate made up of a mixture between the superpartners of the electroweak bosons. A stable and weakly interacting LSP provides an excellent dark matter candidate. The search presented here explicitly targets supersymmetric models in which R-parity is conserved. Scenarios in which the LSP is a neutralino are considered, as well as those in which it is a gravitino, superpartner of the hypothetical graviton that carries gravitational interactions.

The existence of supersymmetry could affect a plethora of measurements in experimental high energy physics, but only the direct observation of superpartners can unambiguously confirm the existence of supersymmetry. At the LHC, the experimental search for direct superpartner production has

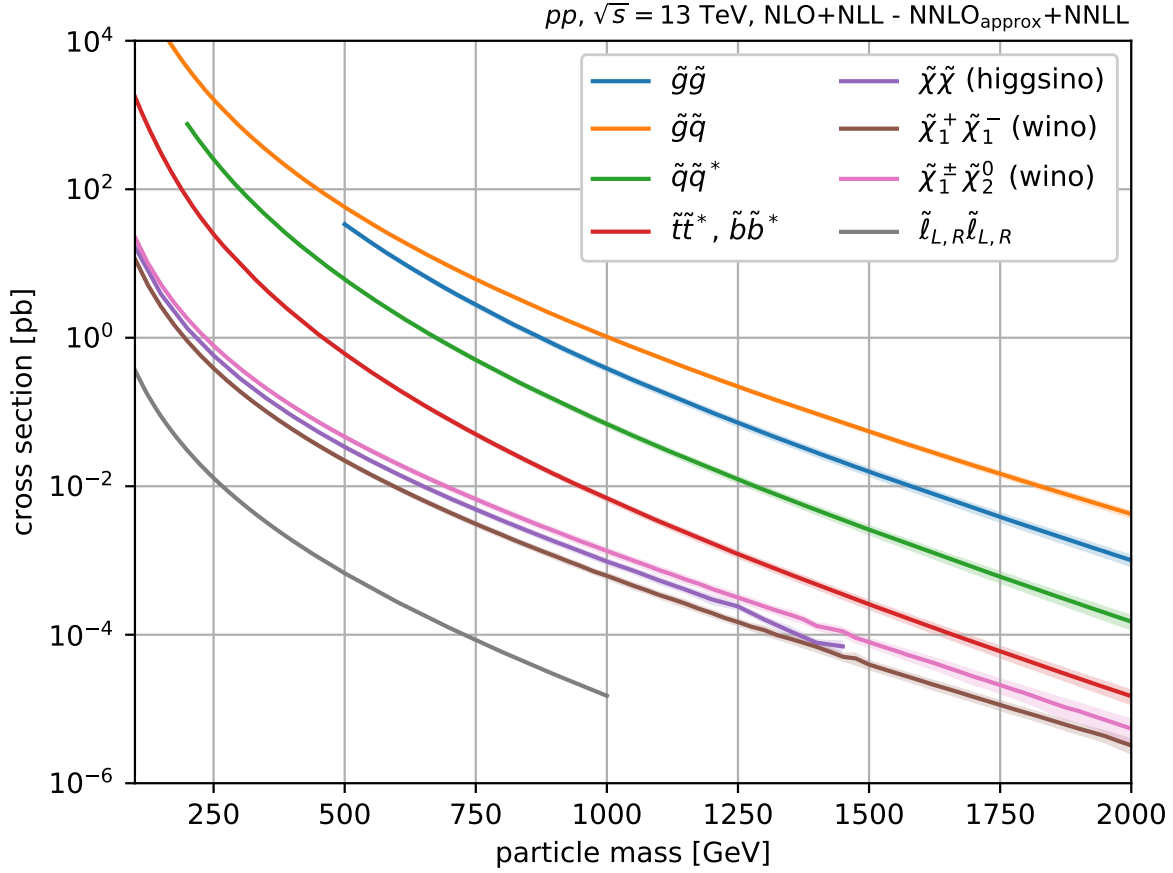


Figure 9.1: Overview of production cross sections of different superpartner pairs in 13 TeV pp collisions [331]. The cross sections for electroweak superpartner production are computed with the techniques described in [332] while those for squarks and gluinos are from [333]. Superpartners are represented by the same symbols as their SM partners, but with an additional tilde. Charginos and neutralinos, mixed states of the superpartners of the electroweak bosons, are represented by the symbol χ .

been carried deep into the TeV frontier. Most searches at both ATLAS [322–325] and CMS [326–330] set their sights on the production of strongly interacting superpartners. The production cross sections of these superpartners are expected to be much higher at the LHC than those of superpartners that only have electroweak couplings. This becomes clear in Fig. 9.1 where the expected production cross sections of different superpartner pairs in 13 TeV pp collisions are shown. The large production cross sections of strongly interacting superpartners are a consequence of the LHC being a pp collider and the fact that strong superpartners can be produced through gluon fusion. At a $p\bar{p}$ collider the production of superpartners through electroweak interactions would be favored because the necessary initial states consist of quarks and antiquarks [119]. All of the referenced searches have returned null results and stringent constraints have been set on the existence of strongly interacting superpartners. An overview of the exclusion limits on the production of gluinos, the superpartners of gluons, is given in Fig. 9.2. The exclusion limits on their mass have been pushed beyond 2 TeV under some model hypotheses.

The lack of experimental evidence for strong superpartners could indicate that they are heavy and out of reach of the LHC. Under these circumstances electroweak superpartner production provides an alternative path to the discovery of supersymmetry. Because of the relatively low cross sections of

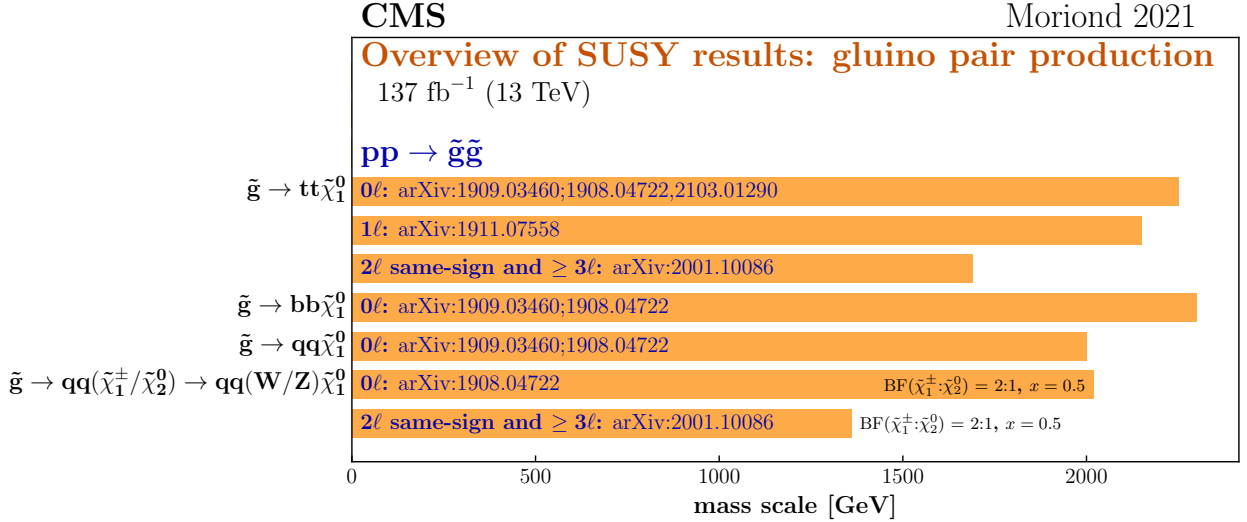


Figure 9.2: Overview of the current most stringent exclusion limits set on gluino production by the CMS Collaboration [334]. Similar results by the ATLAS Collaboration can be found in [335].

electroweak superpartner production, shown in Fig. 9.1, the associated exclusion limits are nowhere near as stringent as those on strong superpartner production. An overview of the limits on electroweak superpartner production set by the CMS Collaboration at the time of writing is presented in Fig. 9.3. Compared to the limits on strong superpartner production shown in Fig. 9.2 these are much lower. Moreover, all of the limits exceeding 1 TeV (labeled with SUS-19-012) are results of the analysis presented in this chapter. Several other searches for electroweak production of superpartners have recently been performed by the ATLAS [336–341] and CMS collaborations [342–345].

Here a search for the direct production of charginos ($\tilde{\chi}_1^\pm$) and neutralinos ($\tilde{\chi}_2^0, \tilde{\chi}_1^0$), admixtures of the superpartners of the electroweak bosons, is presented. Events with two light leptons of the same sign and those with three or more leptons, out of which up to two are τ_h candidates (Section 5.5), are analyzed. The search covers a wide variety of chargino and neutralino decay hypotheses, introduced in Section 9.2, necessitating the analysis of a large number of final states. Only final states with multiple leptons are considered because of the relatively low SM background rates encountered there. The lower purity in the identification of τ_h candidates compared to light leptons drives the upper limit of two τ_h candidates in analyzed events. The full Run 2 data set of 13 TeV pp collisions, collected by CMS from 2016 to 2018, is analyzed. The data set corresponds to a total integrated luminosity of 137 fb⁻¹. This search is an updated version of the one presented in [5], the first project I worked on in the context of this dissertation. The use of parametric neural networks, a complete overhaul of all search region definitions, and a four times larger data set significantly increase the sensitivity of this search compared to the previous iteration. The improvements made in the analysis techniques extend the reach of the search more than the increased data volume does.

If R-parity is conserved, superpartner production at the LHC always results in two LSPs in the final state. In the majority of supersymmetric theories, including those we probe, LSPs are weakly interacting and are expected to escape detection. This means that each superpartner production event contains at least $2m_{\text{LSP}}$ of missing energy. In a hadron collider this translates into a tendency of these events to have a large p_T^{miss} . Supersymmetry searches at LHC almost always leverage p_T^{miss} or related quantities.

In most supersymmetry searches p_T^{miss} and a host of other event properties are used to define a set of orthogonal search region bins. The bins are designed to maximally isolate a large variety of potential manifestations of supersymmetry from SM background events. This is similar to the approach taken

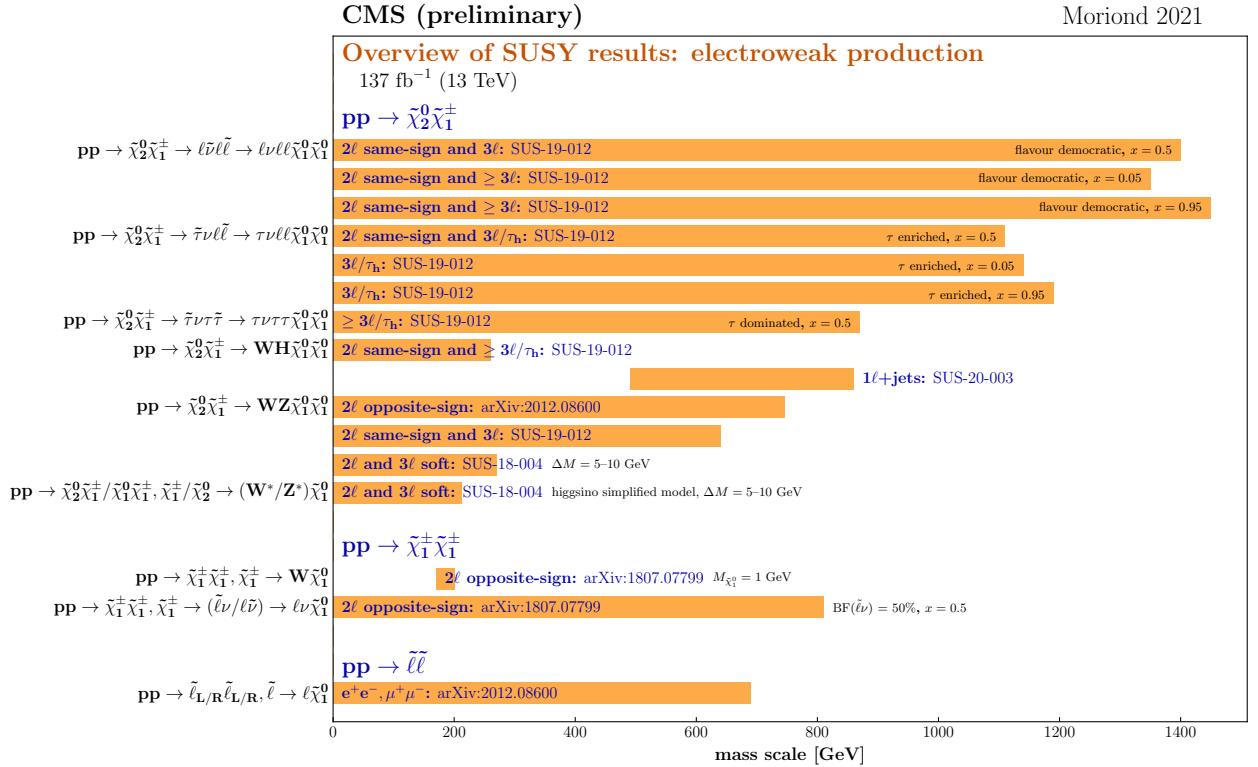


Figure 9.3: Overview of the current most stringent exclusion limits set on chargino-neutralino pair production by the CMS Collaboration [334]. Similar results by the ATLAS Collaboration can be found in [335].

in Chapter 7 to search for sterile neutrinos. The search regions must account for a very large number of model hypotheses spanning a range of possible particle masses and couplings. This makes it a highly challenging optimization problem. In practice, there is no straightforward method to optimize search region bins. Performance metrics based on a few signal hypotheses are typically defined. One then starts from one or a few initial guesses for search region bins and evaluates the performance metrics. The initial guess is refined through trial and error until no further improvements are found. At this point the search region bins are considered optimal. Such an approach covers only a tiny fraction of the space of possible search region bins and is likely to be suboptimal for all or several signal models.

This search is among the first to address the aforementioned optimization problem. A novel search approach, parametric machine learning [346], is used for the first time to probe superpartner production. Parametric machine learning introduces one or more variables that represent unknown parameters of a new physics theory into machine learning models. The resulting models¹ can optimally separate the SM background from the signal at every possible hypothesis of the unknown new physics parameters. The details of the approach are given in Section 9.4.1.2. Parametric machine learning is only used in events with three light leptons containing at least one OSSF pair. Such events dominate the search sensitivity under several signal hypotheses and suffer from a much larger irreducible SM background than any other type of event that is analyzed. In all other event categories the classical approach of orthogonal search region bins is used.

This chapter is structured similarly to chapters 7 and 8: First the concrete signal models that we search for are introduced in Section 9.2, followed by the description of the event selection scheme used to target these signals in Section 9.3. The strategy for separating selected signal and background events is laid out in Section 9.4. This covers both the machine learning and search region bin based approaches. Section 9.6 discusses the systematic uncertainties that affect the search. The final results are presented in Section 9.7, and the interpretation of these results in terms of the signal models is the topic of Section 9.8. Section 9.9 succinctly summarizes the search and gives an outlook.

9.2 Signal models

The search focuses on pair production of charginos and neutralinos resulting in final states with three or more leptons. There is a vast range of supersymmetric models with many free parameters. Even the minimal supersymmetric extension of the standard model (MSSM), contains far too many unknown parameters for a complete interpretation of the LHC data. To facilitate interpretations of the search results, several simplified models are constructed. These serve as benchmarks for different manifestations of supersymmetry that might be found at the LHC [347, 348]. In each simplified model the only free parameters are the masses of two or three of the superpartners.

A first class of models that we consider in this search is the production of chargino-neutralino pairs ($\tilde{\chi}_1^\pm \tilde{\chi}_2^0$). In these models the chargino and heavy neutralino are taken to be admixtures of the superpartners of the $SU(2)_L$ gauge field. The lightest neutralino ($\tilde{\chi}_1^0$) couples mainly to the $U(1)_Y$ field and is the LSP. The neutralino $\tilde{\chi}_2^0$ and charginos $\tilde{\chi}_1^\pm$ are assumed to have equal masses and are the next-to-lightest supersymmetric particles (NLSPs). The masses $m_{\tilde{\chi}_1^\pm} = m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$ are free parameters in the models. More details on the models in this class are given in sections 9.2.1 and 9.2.2.

Scenarios in which $\tilde{\chi}_1^\pm$, $\tilde{\chi}_2^0$, and $\tilde{\chi}_1^0$ are nearly mass-degenerate and mostly made up of the H boson superpartners form a second class of models targeted by this search. The LSP is assumed to be a nearly massless gravitino and the only free parameter in the model is the mass of the mass-degenerate charginos and neutralinos. This family of models is discussed in Section 9.2.3.

The MSSM contains five scalar bosons that have similar properties to the SM H boson. In all signal models we assume one of these five scalar bosons to have an equal mass to the SM H boson and the same couplings. We refer to it simply as the H boson. The other scalars are assumed to be much

¹Trained machine learning algorithms are often referred to as models. Model can also be used to indicate a particular set of hypotheses for the supersymmetry signal we search for. Throughout this chapter the context should make it clear which of the two meanings is attached to the word model.

heavier and of no consequence to our search. All superpartners that are not explicitly considered in the production diagram of a particular model or the subsequent decay chain are assumed to be out of reach of the LHC and decoupled. These superpartners therefore do not affect any of the model predictions.

The kinematic properties of superpartner production events vary considerably depending on the mass difference between the superpartners that take part in the decay chain. In the discussions that follow we use the umbrella term “compressed” to refer to models in which the mass splitting between two or more of the superpartners is small. Final state particles in compressed scenarios tend to have low p_T values, even if the superpartners that were produced are relatively heavy. Models in which the mass differences between superpartners are large have high- p_T particles in the final state and are called “noncompressed”.

The rest of this section defines the specifics of each model this analysis is optimized to search for.

9.2.1 $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with slepton-mediated decays

We consider models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production in which the superpartners of the charged leptons and neutrinos, sleptons ($\tilde{\ell}$) and sneutrinos ($\tilde{\nu}$) are lighter than $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$. The $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ pair decays to $\tilde{\chi}_1^0$ via intermediate sleptons and sneutrinos, as illustrated in Fig. 9.4. We only consider decay modes with three charged leptons in the final state. The branching fraction of $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ to sleptons and sneutrinos is taken to be 100%. The sleptons and sneutrinos are assumed to have equal masses, lying in between $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$. A parameter x is introduced that governs the slepton mass via the following relation:

$$m_{\tilde{\ell}} = x \cdot m_{\tilde{\chi}_2^0} + (1 - x) \cdot m_{\tilde{\chi}_1^0} \quad (9.1)$$

Three different x values are considered in the analysis, each of them serving as a benchmark for a part of the spectrum of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production models with slepton-mediated superpartner decays that might occur in nature:

- $x = 0.5$: The mass of the slepton lies exactly in the middle of $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$. All of the final state leptons shown in Fig. 9.4 carry a similar amount of energy from the $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ decays, resulting in noncompressed models as long as the mass difference between $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$ is sufficient.
- $x = 0.95$: The mass of the slepton is close to $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^\pm}$. Leptons and neutrinos coming from the initial decay of $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ to sleptons and sneutrinos are relatively soft. At this x value models are generally compressed unless the mass difference between $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$ is huge.
- $x = 0.05$: The mass of the slepton is close to $m_{\tilde{\chi}_1^0}$. Depending on the mass splitting between $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$ this results in compressed models with one or two soft leptons. Sleptons and sneutrinos are emitted with significant momentum in the decays of $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$, making models in this scenario generally less compressed than those which have $x = 0.95$.

For this family of models the target final state of the analysis always contains three leptons. In compressed models the probability that one of the final state leptons fails to be reconstructed or identified is relatively large. To partially recover events in which a lepton failed to be identified, we also analyze events with two leptons of the same sign, in addition to those with three or more leptons. Same-sign dilepton events have a low SM background rate and add significant sensitivity to highly compressed models of slepton-mediated superpartner decays. These events also add some search sensitivity for compressed models with WZ- and WH-mediated superpartner decays (Section 9.2.2).

The decays of $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ might be more likely to result in τ leptons than electrons or muons [119]. Whether this is the case depends on the masses of $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ relative to those of the other superpartners, and the superpartner admixtures making up $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$. We consider three different hypotheses:

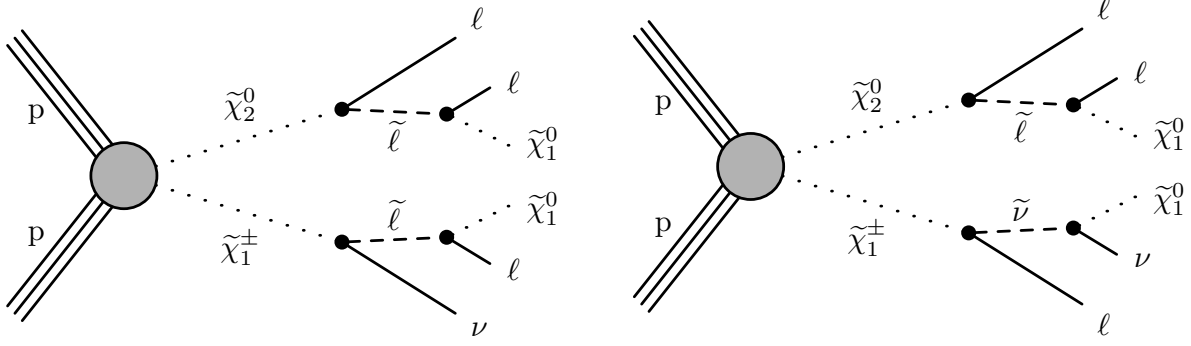


Figure 9.4: Production of a $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ pair with slepton-mediated decays. Only decay chains with three leptons in the final state are targeted in the search.

- Flavor-democratic scenario: In this scenario $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ decay to left-handed sleptons. All lepton flavors occur in the final state with equal probability.
- τ -enriched scenario: The left-handed sleptons and sneutrinos are assumed to be heavy and decoupled so that $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ must decay to right-handed sleptons. The decay of $\tilde{\chi}_1^\pm$ strongly favors τ leptons since right-handed sleptons only couple to $\tilde{\chi}_1^\pm$ via its Higgsino component and the τ lepton has a far larger Yukawa coupling than the light leptons. The decay of $\tilde{\chi}_2^0$ still proceeds with equal branching fraction to all lepton flavors.
- τ -dominated scenario: This scenario covers the possibility that the first and second generation of sleptons and sneutrinos are heavy and decoupled, and only the τ lepton superpartners are lighter than $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$. The decays of both $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ now exclusively result in τ leptons. Final states with light leptons still provide sensitivity to models in this class because of leptonic τ decays.

The sensitivity to flavor-democratic models is entirely driven by events that exclusively contain light leptons in the final state. Compared to τ_h candidates light leptons are reconstructed and identified with a far higher efficiency and a lower misidentification rate. In the τ -enriched and τ -dominated scenarios, events containing τ_h candidates significantly boost the reach of the search.

9.2.2 $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with W, H, and Z boson-mediated decays

If all slepton generations are heavy and out of reach, $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ decay to $\tilde{\chi}_1^0$ by respectively emitting a W, and an H or Z boson. These decay chains are illustrated in Fig. 9.5. The branching fraction to three leptons is far smaller than in the models with slepton-mediated superpartner decays introduced earlier: 3.3% for WZ-mediated superpartner decays and 2.9% for WH-mediated superpartner decays. Because of lepton universality the sensitivity to this family of models is dominated by events with only light leptons in the final state.

9.2.3 Production of mass-degenerate neutralinos and charginos

Lastly, we consider models of neutralino pair production in which the LSP is an almost massless gravitino [349–351]. Direct neutralino pair production has a negligibly small cross section [332, 352, 353]. Effective neutralino pair production is possible if $\tilde{\chi}_1^\pm$, $\tilde{\chi}_2^0$, and $\tilde{\chi}_1^0$ are nearly mass-degenerate. In such a scenario production of chargino pairs, or chargino-neutralino pairs, becomes effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production with additional soft particles. We assume the soft particles escape detection. The remaining $\tilde{\chi}_1^0$ pair decays to two LSPs by emitting H or Z bosons. The different decay chains possible in this model family are drawn in Fig. 9.6.

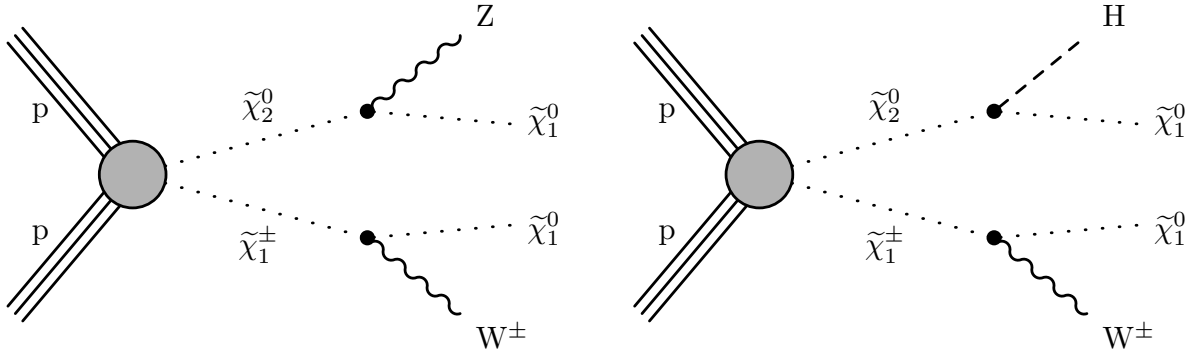


Figure 9.5: Production of a $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ pair with subsequent decays of $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ via the emission of a W and a Z boson (left), or a W and an H boson (right).

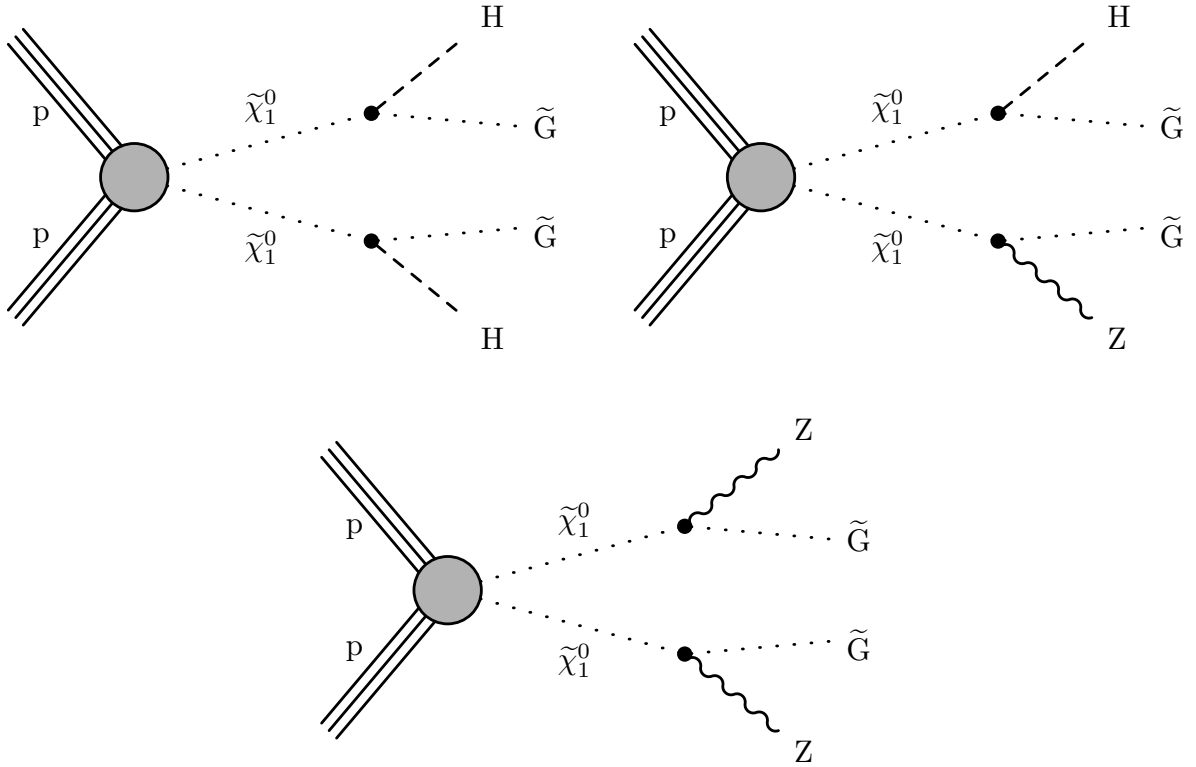


Figure 9.6: Effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ pair production with subsequent decays of the neutralinos via two H bosons (top left), an H and a Z boson (top right), or two Z bosons (bottom).

9.2.4 Simulation and cross sections

Simulated signal samples are generated with the MADGRAPH5_aMC@NLO generator at LO in perturbative QCD. Up to two additional partons are considered in the computation of the matrix elements. Signal cross sections are computed with Resummino [332, 352, 353] at NLO in perturbative QCD including resummation of soft gluon emission at next-to-leading logarithmic precision.

A vast number of simulated signal events must be generated to probe the entire superpartner mass spectrum accessible to the analysis for all signal classes. For this reason the simulation of the detector and its response are performed with the CMS fast simulation package [354]. This package makes the trade-off of giving slightly less accurate results than the GEANT4 based detector model used in the simulation of background events, in return for having a much faster run time. For several signal hypotheses and mass points the agreement between the predictions of the fast simulation and a GEANT4 based simulation are made. No large differences are observed in any variable of importance to the analysis, and systematic uncertainties are introduced to cover any remaining discrepancy. Details on the related uncertainties are provided in Section 9.6.

9.3 Event selection

As elsewhere in the dissertation all particles in the event are reconstructed with the PF algorithm (Section 5.1).

The selection criteria imposed on light leptons in this search are conceptually very similar to those used in the observation of tZq (Chapter 8). The selection consists of two stages: The leptons must first pass a preselection based on a set of independent cuts. Almost every lepton from a superpartner decay passes this preselection stage. Leptons that passed the preselection are required to satisfy a threshold on a BDT classifier trained to optimally separate prompt from nonprompt leptons. All variables used in the light lepton selection are introduced in Section 6.1.

To pass the baseline selection muons must have $p_T > 5$ GeV and be within the muon system acceptance of $|\eta| < 2.4$. They must also pass the loose muon identification requirements (Section 6.1.1.1). Electrons have to satisfy $p_T > 7$ GeV and be within the longitudinal tracking detector coverage $|\eta| < 2.5$. The number of missing hits in the tracker is not allowed to exceed two in the baseline electron selection. All light leptons must pass a set of loose requirements on their impact parameters to ensure they are compatible with originating from the primary pp collision vertex: $d_{xy} < 0.05$ cm, $d_z < 0.1$ cm, and $d_{3D}/\sigma(d_{3D}) < 8$. The most obviously identifiable nonprompt leptons are removed with a very loose isolation requirement: $I_{\text{rel}}^{\text{mini}} < 0.4$. Electrons within $\Delta R < 0.05$ from a muon passing the baseline selection are vetoed. These requirements have an efficiency of more than 99% for prompt leptons stemming from W and Z boson decays and reject most nonprompt leptons.

In addition to the baseline selection we define a tight and a loose selection in similar fashion to what is done in chapters 7 and 8. The tight selection represents the nominal analysis selection, while the loose selection is fine tuned to facilitate a data-driven prediction of the background from nonprompt leptons with the fake rate method (Section 6.4).

Leptons coming from superpartner decays have very similar properties to prompt leptons from decays of W, H, or Z bosons. In both cases the final state leptons are generally isolated from other activity in the detector and originate from the primary pp interaction vertex because of the short lifetime of the mother particle. The BDT classifiers developed for lepton identification in the CMS measurement of the $t\bar{t}H$ process with the full Run 2 data set [60] are found to be highly effective at selecting final state leptons in the signal models of this analysis ². For this reason the tight lepton

²This selection is less fine tuned for the specific target signals than that used in Chapter 8 to search for tZq . In the tZq search the lepton selection algorithm is trained to provide a maximum likelihood classification between prompt leptons from signal events and nonprompt leptons. In this search an algorithm is used that is trained on $t\bar{t}H$ events which have a different underlying event structure and lepton p_T spectrum compared to the signal events of this supersymmetry search. In practice, the $t\bar{t}H$ lepton classifier identifies the final state leptons in any of our signal hypotheses with high efficiency and low background rates. The selection is much more powerful than any standard cut based identification

selection is based on these BDT classifiers. Separate classifiers are used for muons and electrons, and for 2016 and 2017-2018 data-taking conditions. The input variables to the BDT classifiers are the same as those used in the lepton selection BDT of the tZq search (Chapter 8) with the exception that $\text{DEEPJET}_{\text{closest}}$ is used instead of $\text{DEEPCSV}_{\text{closest}}$. The more powerful DEEPJET b-tagging algorithm makes the lepton selection slightly more effective. The DEEPJET algorithm was not available at the time of the tZq search.

The thresholds on the output of the BDT classifiers used for lepton selection are optimized by evaluating the expected upper limits that can be set on the cross sections of several representative signal models in the absence of a signal. These expected upper limits are computed by running the entire search on only simulated events. During this procedure differences in the lepton selection efficiency and trigger efficiency in data and simulation are not accounted for, and the yields of all background sources are predicted directly from simulation. The BDT output thresholds resulting in the lowest expected upper limits are chosen for lepton selection in the analysis. Separate lepton selection working points are used in events with three or more leptons and in those with two light leptons of the same sign, with or without an additional τ_h candidate. In same-sign dilepton events nonprompt leptons make up a larger fraction of the background than in events with three or more light leptons, making a more stringent lepton selection desirable.

To be considered tight, leptons must pass the baseline selection, a threshold on the output of the appropriate BDT classifier, and some additional requirements listed here. Muons must pass the medium muon requirements which ensure a good matching between the track in the muon system and tracking detector (Section 6.1.1.1). Electrons have to satisfy a loose requirement on the electron MVA (Section 6.1.1.2), which is applied in the training data set of the $t\bar{t}H$ lepton selection BDT, and on the properties of their ECAL supercluster to ensure efficient triggering. Both electrons and muons must have $p_T > 10$ GeV to be considered tight.

The total tight selection efficiency for prompt muons typically ranges from 90-99%, depending on their p_T and $|\eta|$. Meanwhile, the misidentification rate for nonprompt muons that pass the baseline selection is between 5 and 10%. Prompt electrons are selected with an efficiency of 75% with the working point used in events with three or more leptons. The corresponding misidentification rate for nonprompt electrons passing the baseline selection is around 5%. In events with two leptons of the same sign the tighter selection working point results in an efficiency of 50–60% for prompt electrons with a typical misidentification rate of 2% for nonprompt baseline electrons³.

To pass the loose selection electrons and muons must pass the same requirements as for the tight selection, with the exception of the BDT classifier. They can either pass the BDT requirement or in case they fail it a lower bound on p_T^{ratio} and a p_T -dependent upper bound on $\text{DEEPCSV}_{\text{closest}}$. Loose electrons are additionally required to satisfy a tighter requirement on the electron MVA than the one used in the tight selection, but only if they fail the BDT selection. The loose selection definition is strictly looser than the tight one, so leptons that are tight are also loose by construction. The loose requirements are fine tuned to enable an accurate prediction of the background from nonprompt leptons with the fake rate method. A single loose selection definition is used despite the different

strategy. It is unlikely that more than small improvements to the lepton selection could be made by tailoring it exactly to our signal. Given the fact that the exact properties of the signal we search for are unknown, doing so would be highly nontrivial. In addition, this search is not as crucially dependent on a completely optimized lepton selection as the search for tZq because the major backgrounds in events with light leptons are irreducible.

³A careful reader might notice that this value is considerably lower than the electron selection efficiency in the tZq search of Chapter 8, even though the latter selection algorithm had a lower misidentification rate. The reason for this is that the BDT classifiers trained for the tZq search have less requirements on the electrons entering the training data set. In particular, no direct cut on the electron MVA is used in the tZq lepton selection, while it is used in the lepton selection developed for the $t\bar{t}H$ measurement. When training machine learning algorithms it is preferable not to apply any preselection on the input variables. Doing so will only reduce the maximum efficiency the algorithm can achieve because predefined cuts do not take into account the relation between all of the variables accessible to the algorithm. The decision to use the $t\bar{t}H$ BDT classifier was made because the training was available under the latest reconstruction conditions and for 2018 collision data. This analysis is less sensitive to the lepton selection than the tZq search, so we decided it would be overkill to train and validate our own lepton identification algorithms to alleviate the electron selection efficiency issue.

BDT requirements imposed in events with two leptons of the same sign and events with three or more leptons. This makes the tuning of the loose selection working point more complex but enables an unambiguous categorization of events depending on their lepton flavor and sign content. The event categorization is based on the number of loose leptons of each sign and flavor in events and is explained in Section 9.4.

In this analysis τ_h candidates are part of the target final state. This stands in contrast to the analyses in chapters 7 and 8 that are entirely based on light leptons. The reconstruction of τ_h candidates is done with the hadron-plus-strips algorithm (Section 5.5). The candidates must be compatible with either the one- or three-pronged hadronic decay modes of the τ lepton. They must satisfy $p_T > 20$ GeV, and have $|\eta| < 2.3$ so that the bulk of the jet from which they are reconstructed falls inside the tracking detector volume. A large number of quark and gluon jets are misidentified as τ_h candidates. This background is reduced by using a BDT classifier that aims to identify prompt τ_h candidates (Section 5.5). The tight working point of this BDT is used for τ_h candidate selection⁴ (not to be confused the nominal lepton selection which is also called tight in this dissertation). Prompt τ_h candidates from W and Z boson decays have a 50% efficiency in passing this selection requirement, while the misidentification rate for QCD jets is 0.2% [232].

In Section 5.5 it was explained that both muons and electrons can have a significant misidentification rate as τ_h candidates. Because of their similarity to charged hadrons, muons might be mistaken for single-prong hadronic τ lepton decays. Electrons, on the other hand, emit bremsstrahlung photons that can resemble the π^0 mesons from a τ lepton decay.

The rate of misidentification of muons as τ_h candidates is reduced by requiring all τ_h candidates to pass a selection aimed at vetoing misidentified muons. If track segments (Section 5.1.1.4) in two or more muon detector planes are found within $\Delta R < 0.3$ of a τ_h candidate, or when the energy deposited in the calorimeters by the highest- p_T track in a τ_h candidate is lower than 20% of the track momentum, the τ_h candidate is vetoed. The efficiency for prompt τ_h candidates from $Z \rightarrow \tau\tau$ decays to pass these requirements is larger than 99%. Prompt muons from $Z \rightarrow \mu\mu$ decays are misidentified as τ_h candidates at a rate of about 0.35% after the application of the aforementioned selection criteria [232].

Another BDT classifier is used to distinguish genuine prompt τ_h candidates from misidentified electrons. The inputs to this classifier are several properties of the photons clustered into the jet that makes up the τ_h candidate. We use the loose working point of this BDT selection, which has a selection efficiency of slightly more than 80% for prompt τ_h candidates in $Z \rightarrow \tau\tau$ decays. The misidentification rate for prompt electrons from $Z \rightarrow ee$ decays is approximately 1%. Any τ_h candidate within $\Delta R < 0.4$ of an electron or muon that passes the baseline selection is vetoed to ensure orthogonality between the selected objects in the analysis.

All τ_h candidates that pass the selection outlined above are considered tight. The loose selection criteria are identical, except that the requirement on the BDT that discriminates prompt τ_h candidates from QCD jets is relaxed. The loose definition of the BDT defined in [232] (again not to be confused with the nominal meaning of the word loose used in this dissertation) is used in the loose τ_h candidate selection. The loose selection of light leptons is heavily fine tuned for closure of the fake rate method. For τ_h candidates the fine tuning options in the selection are limited but the fake rate for τ_h candidates is far easier to measure than for light leptons. Separate measurements of the τ_h fake rate in a $t\bar{t}$ enriched region and a Drell-Yan enriched region are performed. More details on this are given in Section 9.5.

Lepton sign misidentification is an important background source in events with two same-sign light leptons. Even a small sign misidentification rate can lead to a large pollution in the same sign dilepton event category because of the copious rate of events with two opposite-sign leptons. To reduce the rate of this background source we impose additional requirements on the leptons in events with two same-sign light leptons, whether or not the event contains an additional τ_h candidate. For muons the relative uncertainty in the p_T measured purely from the tracking detector track must be smaller

⁴A significant improvement to the τ_h candidate selection in the analysis can be made by replacing the BDT based selection with the neural network based DEEPTAU algorithm [355]. This algorithm was unfortunately not ready for use in physics analyses at the time the search presented here was developed.

than 20%. After imposing this requirement the sign misidentification rate for muons is found to be below 10^{-6} , negligibly small. The remaining background contribution from muon sign misidentification is estimated directly from simulation. The probability that a prompt muon passes the aforementioned requirement is greater than 99.9%.

The sign of electrons is determined from the full GSF fit used in their reconstruction (Section 5.1.1.3). The nominal sign measurement is required to be consistent with a simplified sign measurement: The direction of the electron when crossing the pixel detector is linearly extrapolated to the calorimeter surface. The relative position of the extrapolation to the calorimeter deposit of the electron gives a measurement of its sign because of the bending of electron trajectories in the CMS magnetic field. If the two sign measurements give different results, the electron is vetoed. The consistency requirement pushes the sign misidentification rate for electrons down to values that range from $3 \cdot 10^{-3}$ to well below 10^{-5} depending on the p_T and η of the electron. The sign measurement consistency requirement has an efficiency of over 99% for prompt electrons that pass the tight selection. The sign misidentification rate is at its maximum for high p_T and η electrons in 2016 data. In the 2017 and 2018 data sets the sign misidentification rate is lower by more than a factor two at almost any electron p_T and η value. This is due to the new pixel detector which is lighter than the previous one, resulting in less electron bremsstrahlung, and which has an additional measurement layer (Section 4.4.1).

We target the signal models introduced in Section 9.2 in their decays to three or more leptons. To mirror this, events are selected for analysis if they contain three or more tight leptons. To increase the selection efficiency in compressed scenarios where it is likely that at least one lepton fails to be identified, events with two tight light leptons of the same sign are also retained. If more than two τ_h candidates are present in an event, it is discarded. In such events the signal sensitivity is drowned by a gigantic background from nonprompt τ_h candidates. Events are required to pass at least one unrescaled trigger path out of those selecting one, two, or three light leptons. Triggers selecting τ_h candidates have a markedly lower efficiency than those selecting light leptons and including them is found not to give meaningful improvements to the analysis. They are therefore omitted to reduce complexity of the analysis.

Additional p_T thresholds are introduced on light leptons to ensure all events selected for analysis can efficiently pass one or more of the chosen trigger paths. These requirements are applied on top of those mentioned above in the selection of individual leptons. In events with two, three, or more light leptons, the leading light lepton must have $p_T > 20(25)$ GeV if it is a muon (electron). If there is only a single muon in the event, this p_T threshold is increased to 25 GeV. The subleading lepton must satisfy $p_T > 10(15)$ GeV in case it is a muon (electron). No additional p_T requirements are imposed on any of the other leptons present in the event. In events with only a single light lepton and two τ_h candidates, the p_T threshold for the light lepton is increased to $p_T > 25(30)$ GeV if it is a muon (electron).

As in the other analyses, the total trigger efficiency in each event category is measured in a data set collected with independent triggers requiring the presence of significant p_T^{miss} or hadronic transverse momentum (Section 6.2). The presence of τ_h candidates has a negligible impact on the trigger efficiency since the chosen trigger paths only impose requirements on light leptons. Trigger efficiencies are therefore measured in events with one, two, or three light leptons. The trigger efficiency in events with τ_h candidates and light leptons is taken to be that of the measurement with the corresponding number of light leptons⁵. The trigger efficiency in events with two same-sign light leptons is measured without applying criteria on the sign of the leptons since the trigger efficiency should be independent of the lepton signs.

Data to simulation scale factors in the trigger efficiency are used to correct simulated events.

⁵A small caveat here is that the single lepton data set used for the trigger efficiency measurement was preselected with the requirement of having two τ_h candidates in each event. This reduces the statistical power of the trigger efficiency measurement, but it is still more than high enough for an accurate measurement. This is augmented by the fact that the background from nonprompt τ_h candidates is dominant in events with two τ_h candidates and one light lepton. This background is predicted directly from data, thus requiring no data to simulation scale factor for the trigger efficiency.

Table 9.1: Total trigger efficiency in data and simulation and the corresponding scale factor as a function of the number of light leptons in the event. The uncertainties in the efficiency measurements are purely statistical, while the data to simulation scale factors also have a systematic uncertainty. The systematic uncertainty on a given scale factor is taken to be the difference in the scale factors obtained from separate measurements in each data-taking year.

Data set	Total trigger efficiency	Data to simulation scale factor
1 ℓ data	0.795 ± 0.009	$0.927 \pm 0.010(\text{stat.}) \pm 0.07(\text{syst.})$
1 ℓ simulation	$0.858^{+0.009}_{-0.010}$	
2 ℓ data	0.953 ± 0.001	$0.980 \pm 0.002(\text{stat.}) \pm 0.02(\text{syst.})$
2 ℓ simulation	0.973 ± 0.002	
3 ℓ data	$0.995^{+0.002}_{-0.004}$	$1.001 \pm 0.004(\text{stat.}) \pm 0.001(\text{syst.})$
3 ℓ simulation	$0.994^{+0.0008}_{-0.0009}$	

Trigger efficiencies in simulation are obtained from large samples of simulated WZ events, one for each data-taking year. The trigger efficiency measurements are done in an integrated fashion across the p_T and η of the leptons. Data sets taken with orthogonal triggers lack statistical power for a parametrization of the trigger efficiency in terms of the p_T and η of the leptons. Separate measurements of the trigger efficiency scale factors are done for 2016, 2017, and 2018 collision data. Differences in the resulting scale factors between the data-taking years are taken as a systematic uncertainty. The measured trigger efficiencies in data and simulation, their statistical and systematic uncertainties, and the resulting scale factors are given in Table 9.1. Because of the large redundancy between trigger paths in events with three light leptons, the trigger efficiency in these events is compatible with 100%. The analysis also uses events with four light leptons. Such events are too rare to be able to perform an accurate trigger efficiency measurement in a data set collected with independent trigger paths. The trigger efficiency in events with three light leptons, which is already 100%, serves as a lower limit for the trigger efficiency in events with four or more light leptons. The motivation for this is that four lepton events have one additional lepton which results in an even higher degree of redundancy in the trigger selection. So we can safely conclude that four light lepton events have a trigger efficiency of 100%.

A typical feature of superpartner production in supersymmetric models with R-parity conservation is the presence of a large p_T^{miss} caused by the LSPs escaping detection. Many background events, on the other hand, have a small p_T^{miss} . For this reason a threshold of $p_T^{\text{miss}} > 50 \text{ GeV}$ is introduced in events with three or more leptons. In events with two light leptons of the same sign the more stringent requirement $p_T^{\text{miss}} > 60 \text{ GeV}$ is applied. This is motivated by the relatively larger background contribution from events with a small p_T^{miss} , for instance due to electron sign misidentification.

Events containing an OSSF pair of leptons passing the baseline selection with an invariant mass below 12 GeV are vetoed. This criterion reduces the background yield from photon conversions and low-mass resonances, while leaving the signal selection efficiency nearly unaffected.

The anti- k_T algorithm with a distance parameter of 0.4 is used for jet clustering (Section 5.2), and these jets are further corrected for the nonuniformity and nonlinearity of the detector response (Section 5.3). Jets must have $p_T > 25 \text{ GeV}$, $|\eta| < 2.4$, and be separated from any loose lepton by a distance greater than $\Delta R = 0.4$. Jets that are compatible with the hypothesis of originating from noise or anomalous energy deposits are rejected.

None of the signals targeted by this analysis involve top quarks, so the vast majority of signal events is expected not to contain b jets. Jets in the analysis are considered b-tagged if they pass the tight working point of the DEEPCSV algorithm. This criterion has an approximate efficiency of 50% in identifying genuine b jets while having a misidentification rate of 0.1 (2.4)% for light quark and gluon (c quark) jets. Events are vetoed if they contain one or more b-tagged jets.

Table 9.2: Overview of the different event categories used in the search.

Event category	Requirements
2 ℓ SS	Two same-sign light leptons
3 ℓ A	Three light leptons at least two of which form an OSSF pair
3 ℓ B	Three light leptons, none of which form an OSSF pair
3 ℓ C	$\mu^+\mu^-$ or e^+e^- pair and a τ_h candidate
3 ℓ D	$e^\pm\mu^\mp$ pair and a τ_h candidate
3 ℓ E	Two same-sign light leptons and a τ_h candidate
3 ℓ F	Two τ_h candidates and one light lepton
4 ℓ G	Four light leptons, including two separate OSSF pairs
4 ℓ H	Four light leptons, including one or no OSSF pairs
4 ℓ I	Three light leptons and a τ_h candidate
4 ℓ J	Two light leptons and two τ_h candidates, including two OSSF pairs
4 ℓ K	Two light leptons and two τ_h candidates, including one or no OSSF pair

At the time this analysis was performed the newer DEEPJET b-tagging algorithm was already available and it is an input to the BDT classifiers used for selecting light leptons. The older DEEPCSV algorithm is however found to be the better choice in this search because the associated uncertainties in the misidentification rates of light quark and gluon jets are far smaller. The vast majority of signal and background events in this search have no genuine b jets. In this situation keeping the uncertainty in the misidentification rate small is more important than the slight decrease in the background yield that would result from using the DEEPJET algorithm.

9.4 Signal extraction

The signal models listed in Section 9.2 lead to a large variety of final states with different lepton flavor and sign combinations across a wide kinematic spectrum. The masses and couplings of superpartners are unknown so all possibilities must be explored. With this search we aim to cover a set of manifestations of supersymmetry as broad as possible. This is achieved by allowing almost any combination of multiple light leptons and τ_h candidates in the selected final states. Events are categorized based on their lepton flavor and sign content. Each event category adds search sensitivity to particular supersymmetric scenarios. Table 9.2 gives an overview of the different event categories used in the search and the labels we use to refer to each of them. All of the categories are motivated in more detail in the following sections.

The events in each category are further divided into an array of search regions based on their kinematic properties. The search regions are defined with the aim of separating signals at varying superpartner masses from background events. This strategy is identical to the one followed in Chapter 7 to isolate a signal from hypothetical sterile neutrinos. The complexity of this search is however far greater because of the vast amount of free parameters in supersymmetry.

In events with three light leptons, at least one of them forming an OSSF pair, a more powerful search strategy based on parametric neural networks is used. With this novel search strategy we try to address innate problems in the optimization of search region bins and the fact that only one set of bins is defined to simultaneously probe all signal hypotheses. Further motivation and details are provided in Section 9.4.1. An overview of the search region definitions in all the other event categories is given in sections 9.4.2 to 9.4.5.

9.4.1 Three light leptons with at least one OSSF pair

Light leptons can be reconstructed and measured with much greater efficiency than τ_h candidates while at the same time having far lower background rates in their identification. Consequently, if

superpartners do not favor decays to τ leptons over the other lepton flavors, the sensitivity of the analysis is almost completely determined by events which exclusively contain light leptons. This is the case for models of flavor-democratic slepton-mediated superpartner decays (Fig. 9.4), and WZ-mediated superpartner decays (Fig. 9.5). In both of these scenarios superpartner decays to light leptons, without an intermediate τ lepton, always result in the presence of an OSSF lepton pair. Events with three light leptons at least two of which form an OSSF pair are therefore of extraordinary importance to this search. We refer to this event category as $3\ell A$. The $3\ell A$ category has the highest background yield out of any event category in the analysis by a large margin. The main background component is SM WZ production, which is irreducible.

Much of the effort in previous iterations of this analysis [5, 356] was concentrated on $3\ell A$ events. A large and complex set of search region bins was optimized in this event category. These search regions aim to be sensitive to both WZ- and slepton-mediated superpartner decays. A vast range of kinematic possibilities must be covered for the search to be exhaustive. The kinematics of the event strongly depend on $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$, with larger mass differences resulting in more energetic final state particles. The search must probe mass differences ranging from a few GeV to more than a TeV. In slepton-mediated superpartner decays, the p_T spectra of the final state leptons and the distribution any variable derived from them strongly depend on whether $m_{\tilde{\ell}}$ is close to $m_{\tilde{\chi}_2^0}$ ($x = 0.95$), $m_{\tilde{\chi}_1^0}$ ($x = 0.05$), or in between them ($x = 0.5$). Each of these scenarios should be simultaneously targeted by the search. Superpartner production with WZ-mediated decays is kinematically similar to SM WZ production in many respects, one example being the invariant mass of the Z boson decay products. Exploring this decay scenario requires a very different set of search regions from those one would need for probing slepton-mediated superpartner decay models, which do not contain a Z boson.

Particularly difficult are WZ-mediated superpartner decay scenarios in which the mass difference between $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$ approaches the Z boson mass. In this case the kinematics of the signal become highly similar to those of SM WZ production. The superpartner mass difference is just enough to produce a real W and Z boson, and the LSPs are emitted at low p_T . The bulk of the $\vec{p}_T^{\text{miss}}$ comes from the neutrino from the W boson decay, just like in SM WZ production. This region is called the “WZ corridor” for the characteristic gap, or corridor, that appears in exclusion limits on $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production as the difference between $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$ approaches m_Z . The exclusion curves in the previous iteration of the analysis provide a striking example of this [5]. There are however small differences between the signal in the WZ corridor and the WZ background: SM WZ production has interference with $W\gamma^{(*)}$ production. This interference alters the event kinematics, such as the invariant mass distribution of the Z boson decay products, somewhat. The electrically neutral $\tilde{\chi}_2^0$ does not couple to photons so the interference effect is not present in signal events. Another difference is that while the final state is similar between SM WZ production and superpartner production, this is not necessarily the case for the underlying event in which the initial state particles were produced. Different amounts of ISR can alter the observed kinematics and the amount of hadronic activity in the events is generally different. Nonetheless, separating signal and background events in the WZ corridor remains difficult.

Given the possibilities outlined above it becomes clear that defining a set of search regions that satisfies the requirement of being as sensitive as possible to all the possible signal scenarios is a daunting task. At their core search regions are an attempt to approximate a, often highly nonlinear, multivariate function that separates signal and background events optimally with a set of discrete sequential cuts. In practice, there is no straightforward or computationally tractable method of finding a truly optimal set of search regions. In a typical search for new physics one starts by defining a set of search regions based on physical intuition, experience from prior work, a guess, or a combination of these. One then evaluates the performance of the search regions in terms of some metric of choice. This metric is typically the expected exclusion limit for a few cherry-picked signal mass and decay hypotheses. The set of search regions is then iteratively refined by varying bin definitions, increasing or decreasing the number of bins, and changing the kinematic variables in terms of which the bins are defined. This procedure has often given good analysis results but it is extremely unlikely that the search region bins of any search for new physics could not be further improved if more possibilities are explored.

Conceptually one can imagine the standard search region optimization procedure as starting at a not entirely random point in a high dimensional space that covers all possible sets of search regions one could define. A small space around the initial guess is explored during the optimization procedure but the true optimum might lie somewhere else completely. One has no guidance about what direction to move in to find the optimum. Out of the explored points the optimum one is chosen in terms of only a select few signal hypothesis, not all of them.

In this analysis we explore an alternative to the classical approach of search region bins. Machine learning classification algorithms can be used to separate signal and background events in a nearly optimal fashion. The main difference with the approach of defining discrete search region bins is that some machine learning algorithms can efficiently find an optimum for their objective. Given the current configuration of the learning algorithm and the performance on some data examples, there is a well defined rule to update the configuration in a direction that brings it closer to the optimum. Such algorithms essentially have a sense of direction in the configuration hyperspace they are exploring. A good example of this is the gradient descent algorithm that is used in the optimization of neural networks. A training example is passed through the neural network and the objective, also called loss function, is evaluated for the resulting output. The partial derivatives of the loss function with respect to the weights in the neural network are computed in a backwards pass through the neural network, and the weights are updated with this gradient to move the neural network one step closer to optimal performance. Another example is a gradient boosted forest in which each tree added to the forest represents a gradient of the current output in the direction of the objective (Equation 8.1). Gradient based optimization strategies stand in stark contrast with the semi-directionless strategy used for choosing search region bins. Because of the discrete nature of search region bins gradient based optimization approaches can not be applied to search region optimization in a straightforward manner.

To increase the reach of the search, neural network classifiers are introduced to separate signal and background events in $3\ell A$. The neural networks use parametric machine learning [346], a technique that is applied for the first time in a supersymmetry search. This approach is explained in Section 9.4.1.2. We also repeat the analysis of $3\ell A$ events with the classical orthogonal search region approach to facilitate an easier reinterpretation of the results by other researchers. The search region based analysis in $3\ell A$ is outlined in Section 9.4.1.4. To evaluate the analysis results in $3\ell A$, the neural network and search region approaches are always evaluated separately because the same data events are used in both approaches. The results of the neural network analysis are better than those of the search region analysis for every signal mass point and decay hypothesis for which a neural network is used. The nominal analysis results are therefore those obtained with the neural networks.

9.4.1.1 Parametric machine learning

When the properties of the signal one searches for are unambiguously defined it is straightforward to train a machine learning classifier to separate signal and background events. This is exactly what is done to search for tZq production in Chapter 8. Being a SM process, tZq production has well defined properties and a classifier can be trained to identify simulated tZq events. Trying to apply such an approach in our supersymmetry search presents us with a clear problem: Supersymmetry has many unknown parameters. In this analysis the free parameters in each of the signal models are $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$, and the kinematics of signal events are highly dependent on these masses. One can think of several naive approaches of training a machine learning classifier to separate superpartner signals for from the background. Some are listed here.

- A first tentative solution might be to train a single machine learning model on a training data set that contains a mixture of simulated signal events covering the range of $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$ one wants to have sensitivity to. This approach has the problem that the classifier is trained to find the average signal in an optimal ⁶ way, not the individual signals. The average is arbitrary because

⁶When the term optimal is used, it is only with respect to machine learning models of the same family, with a fixed

it depends on the number of simulated signal events for each mass configuration. Mass points with a larger number of events in the training data set will be favored by the machine learning model and will have a relatively better classification performance. In addition, the classifier is in general nonoptimal for each individual mass point since it must take all of them into account at once.

- An alternative to the aforementioned solution is to train the machine learning algorithm on a single mass point. This yields a classifier that is optimal for the mass point used during training, but is suboptimal for all other mass points. The further separated a mass point is from the one used for training, the worse the classification performance for this mass point generally is.
- Another straightforward solution is to train a separate classifier for each signal mass point that is available in simulation. This approach results in an optimal classifier for each simulated signal mass point. Obvious drawbacks are that a great deal of complexity is introduced in an analysis in this way and that training all machine learning models is costly in terms of computational time. A more fundamental problem in this method is that one can only simulate and evaluate the signal at a discrete set of mass points. It is not obvious how to treat signal mass points for which there are no events in the training data set. For a signal falling in between two points present in the training data set there are no guarantees on the performance of any of the trained classifiers. In the worst case none of the classifiers provides any sensitivity at all and a new physics signal gets missed. More likely is that the results for signal points not seen in the training data set are simply suboptimal. An attempt to solve this could be to use a linear interpolation between the outputs of the classifiers trained to identify the two nearest trained mass points. In general however, the classification output varies in a nonlinear way as a function of the signal mass point.

Parametric machine learning, introduced in [346], fills the holes in the tentative solutions outlined above. The unknown parameters of a new physics model are added as input variables to a machine learning algorithm, on top of the reconstructed event quantities nominally used as inputs. Each simulated signal event has a well defined value for the new physics parameters: The parameter values at which it was simulated. On the other hand, the new physics parameters are meaningless for SM background events. For each background event random parameter values are sampled from the parameter distribution over the set of simulated signal events. This means that the overall probability distribution of the new physics parameters for signal and background events seen by machine learning algorithm are identical by construction. The machine learning algorithm can therefore not use the new physics parameters directly to classify events. During training the machine learning algorithm learns that no relation exists between the new physics parameters and the properties of background events. Meanwhile, the algorithm learns the dependency of signal event properties on the new physics parameters.

By learning how the properties of a signal event change with the new physics parameters, a single trained machine learning algorithm can achieve optimal performance at all parameter values. Parametric machine learning algorithms also have the ability to achieve optimal performance at parameter values not present in the training data set. This is an important property to be able to discover a signal at parameter values not represented in the set of simulated signal events used to train the machine learning algorithm. Optimal performance for each seen and unseen point is not guaranteed for an arbitrary algorithm training and must be explicitly verified. It should also be noted that a general property of supervised machine learning algorithms is that they are powerful tools for interpolation but can not accurately extrapolate to points outside the manifold spanned by the training data set. It is therefore important to simulate signal events up to the extreme values of the new physics

set of hyperparameters, and a given amount of events in the training data set. It will always be possible to further increase the performance, although probably only slightly, with alternative machine learning models and more training data.

parameters the analysis might have sensitivity to. For parameter points exceeding the range spanned by the training data set there are no guarantees whatsoever on what a machine learning algorithm will output.

Neural networks are preferable to other machine learning algorithms for practically achieving the idealized properties of parametric machine learning sketched above. In their simplest form, so called fully-connected neural networks, they are a sequence of layers, each of which transforms the output of the previous layer $\mathbf{y}_{L-1}$ in the following way:

$$\mathbf{y}_L = f_L(W_L \cdot \mathbf{y}_{L-1}). \quad (9.2)$$

Here f_L can be any simple nonlinear function applied to each element in a vector, and W_L is an $m \times n$ matrix with n the size of vector $\mathbf{y}_{L-1}$ and m the size of vector $\mathbf{y}_L$. The set of input variables forms the vector y_1 and the weight matrices W of every layer are set to minimize an error function on the training data set (more details are given in Section 9.4.1.3). Neural networks are smooth functions of their input variables and can learn arbitrarily complex multivariate nonlinear functions. The only limits are the number of learnable parameters in the neural network and the size and quality of the training data set. The output of a parametric neural network varies smoothly as a function of the new physics parameters, which is important to be able to evaluate the neural network at any arbitrary parameter value. In addition, a neural network with a sufficient amount of internal weights is able to learn any dependency of event kinematics on a set of new physics parameters, however complex and nonlinear. Tree based methods, such as the BDT classifiers used elsewhere in this work, are piecewise functions that do not vary smoothly as a function of their input variables. Their ability to learn nonlinear relationships is also more limited than that of neural networks. Deep neural networks are particularly powerful because they tend to learn nonlinear relationships by building a hierarchical structure throughout their layers. They can even learn to solve complex problems while only having access to low-level inputs without human feature engineering. Deep neural networks are behind the recent revolution in fields such as computer vision and natural language processing, but are equally powerful when applied to high energy physics [357, 358]. For the reasons listed above we choose to use parametric neural network classifiers to search for various signal models in this analysis.

One drawback of neural networks compared to tree based methods is that properly training them and choosing their configuration requires more care. Section 9.4.1.3 is included for this reason. It is a relatively dense and technical section on the design choices and optimizations that went into the neural networks used for the analysis.

In a standard machine learning classifier used in high energy physics the output for a given event only depends on the values of the reconstructed quantities serving as inputs to the algorithm. A single event has a single output value for the classifier. This is the case in Fig. 8.14 where each selected event has a single output value for one of the BDT classifiers used to search for $t\bar{t}Zq$ production. In a parametric machine learning model the output depends on both the reconstructed event properties and the new physics parameters. In other words, given an event with fixed reconstructed quantities the output is a function of the parameters. This is visualized for a parametric neural network in Fig. 9.7.

9.4.1.2 Parametric neural networks to search for superpartner production

The signal models targeted in the $3\ell A$ event category have several unknown parameters. For models of WZ-mediated superpartner decays the free parameters are $m_{\tilde{\chi}_2^0} (= m_{\tilde{\chi}_1^\pm})$ and $m_{\tilde{\chi}_1^0}$. The kinematic properties of signal events under this decay hypothesis are in approximation governed by a single parameter:

$$\delta m = m_{\tilde{\chi}_2^0} - m_{\tilde{\chi}_1^0} = m_{\tilde{\chi}_1^\pm} - m_{\tilde{\chi}_1^0}. \quad (9.3)$$

The mass gap δm between $\tilde{\chi}_1^0$ and $\tilde{\chi}_2^0$ determines the momenta of the particles produced in the decays of $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$. A $\tilde{\chi}_1^0$ is only observable through the p_T^{miss} it induces, and LSPs of different

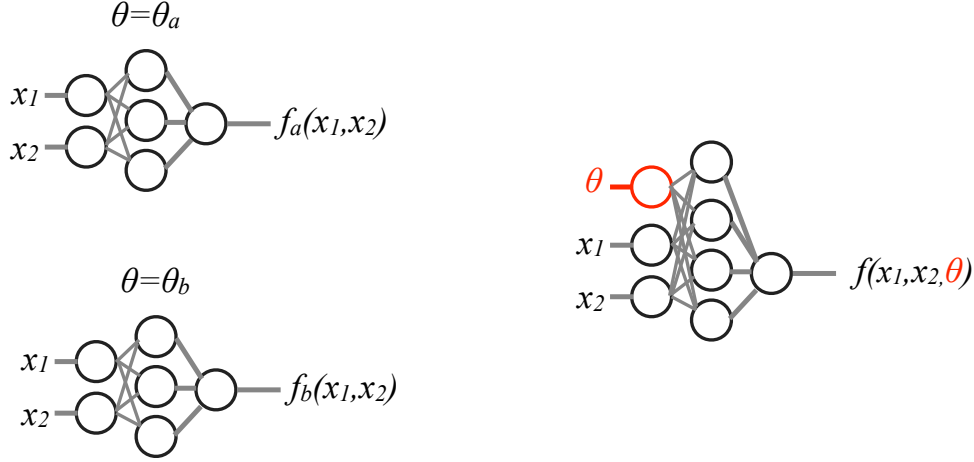


Figure 9.7: Illustration of a simple parametric neural network. Given a new physics parameter θ one could train separate neural networks to target fixed values of θ (left) or a single parametric model that uses θ as an input (right). Given an event with the reconstructed scalar quantities x_1 and x_2 classical neural network classifiers have a single output value, while the output of a parametric neural network is a function of the parameter θ [346].

mass but equivalent $p_T^\pm$ look identical in a reconstructed event. Small kinematic differences with varying superpartner masses at a given δm do occur because the underlying interaction needed to produce a $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ pair of high and low mass differs. These differences are minor and found to be largely inconsequential in the search.

For models of slepton-mediated superpartner decays $m_{\tilde{\ell}}$ is another free parameter. We express it in terms of the quantity x introduced in Equation 9.1, which governs the proximity of $m_{\tilde{\ell}}$ to $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$. Three values of x are studied. These serve as benchmarks for scenarios in which $m_{\tilde{\ell}}$ is close to $m_{\tilde{\chi}_2^0}$, close to $m_{\tilde{\chi}_1^0}$, or in the middle between both.

We opt to train four neural networks parametric in δm : One to search for WZ-mediated superpartner decays and three to probe slepton-mediated superpartner decays, one for each value of x . Each neural network has one output value: The predicted probability that a collision event stems from superpartner production. For slepton-mediated superpartner decay models a single neural network parametric in both δm and x could be trained. The three x values available in simulation are however not enough to verify the properties of a neural network parametrization in terms of x . We therefore opt to train separate neural networks. An alternative would be to train a neural network parametric in $m_{\tilde{\chi}_2^0}$, $m_{\tilde{\chi}_1^0}$, and $m_{\tilde{\ell}}$, but this is unfeasible for reasons laid out below.

Using a neural network parametric in the single parameter δm has some clear practical advantages over a parametrization in the masses of all superpartners present in a particular signal model. As discussed in Section 9.4.1.1, given an event the output of a parametric neural network is a function of the parameters. Standard new physics searches bin events in a set of search regions or a single machine learning classifier. This results in a unique distribution of data and simulated events. When using a parametric neural network this is no longer the case: The distribution of data and simulated events as a function of the neural network output changes when evaluated at different parameter values. In other words, the set of selected collision events produces a continuum of output distributions instead of a single one. To illustrate this point, the expected distribution of events from SM WZ production as a function of the output of the parametric neural network targeting WZ-mediated superpartner decays (slepton-mediated superpartner decays at $x = 0.5$) is shown in Fig. 9.8 (Fig. 9.9). Each δm slice contains the same events.

To practically search for a signal, the parametric neural networks must be evaluated at a discrete set of parameter values. This set must be large enough to probe the entire parameter space where

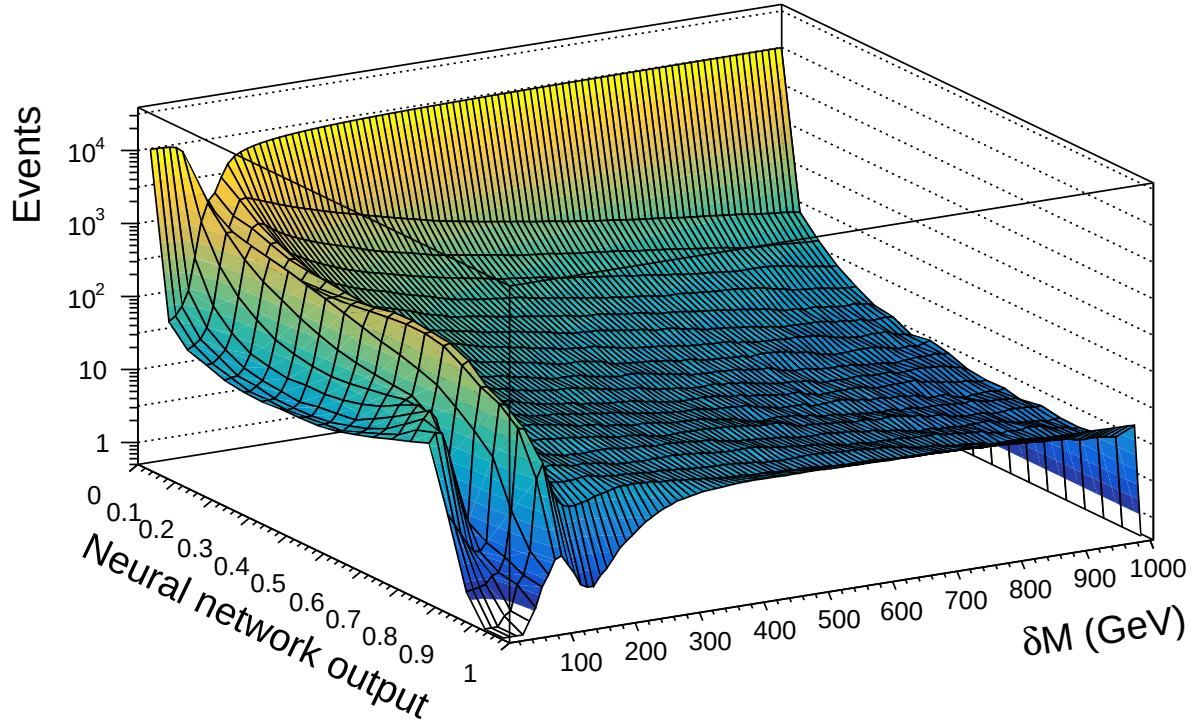


Figure 9.8: Expected yield of $3\ell A$ events from SM WZ production as a function of δm and the output of the parametric neural network targeting WZ-mediated superpartner decays. Each slice at a constant δm value contains the same set of events. A clear ridge can be seen around $\delta m = m_Z$, indicating that the discrimination power of the neural network between signal and background events is reduced in this δm region. This is the WZ corridor.

sensitivity to superpartner production might be achieved without risking that a signal at a parameter value in between those evaluated is missed. Each evaluation of the neural network requires processing the entire set of selected collision events. We evaluate the neural networks at each δm value for which simulated signal events exist. Doing so gives a maximal resolution in the exclusion limits of the analysis while a significant signal would not be missed. At neighboring points in the set of evaluated δm values the neural networks still vastly outperform the search region based analysis introduced later. The full set of δm points at which the analysis is evaluated can be found later in this chapter in Fig. 9.33 to 9.36.

With a single new physics parameter in the neural network the analysis becomes considerably more demanding in terms of computing resources than typical analyses, hundreds of times more so than the analyses presented in chapters 7 and 8. Evaluating the analysis for two or more parameters was unfeasible with the computational setup available for this work.

One can get a rough idea ⁷ of the computing resources necessary to evaluate the analysis by estimating the number of floating point operations needed to compute the neural network output at each evaluated parameter. We look at the neural network used to identify slepton-mediated superpartner decays at $x = 0.95$ as an example. A single evaluation of the neural network requires roughly 10^9 floating point operations. The amount of events the neural network has to be evaluated on is of the order 10^7 , including collision data, and simulated signal and background events. At a fixed δm value

⁷The reasoning that follows is simplistic in the sense that it ignores every other computational step in the analysis other than the evaluation of the neural networks. The real increase in execution time that would result from adding multiple new physics parameters to the neural networks depends on the structure of the analysis software. Nonetheless, one can get a notion of the orders of magnitude in the relative analysis execution times for different parameter choices.

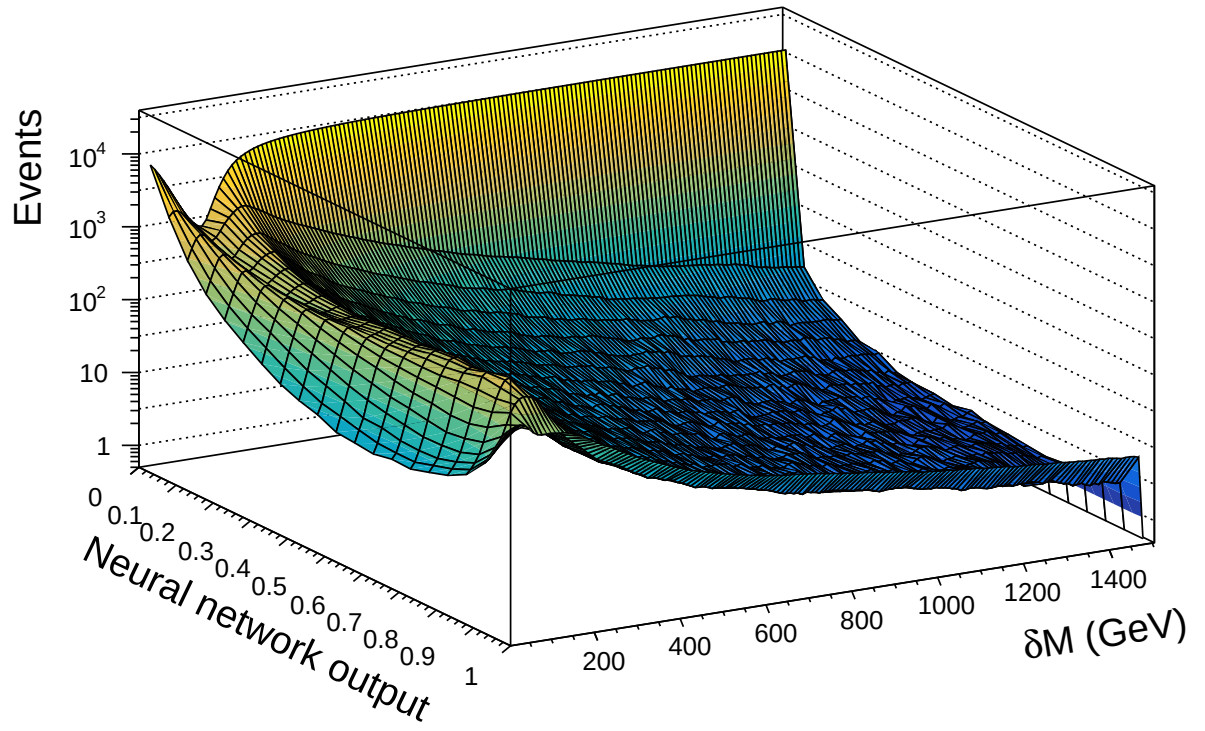


Figure 9.9: Expected yield of $3\ell A$ events from SM WZ production as a function of δm and the output of the parametric neural network targeting slepton-mediated superpartner decays at $x = 0.5$. Each slice at a constant δm value contains the same set of events.

each event requires seven evaluations of the neural network to account for systematic uncertainties that directly modify the event kinematics. We established that the neural network is evaluated at each δm value available in the signal simulation, in this case 34 distinct δm points. Putting all of this together results in a total of more than $2 \cdot 10^{18}$ floating point operations that have to be carried out. A full analysis evaluation takes a bit less than twenty-four hours on a few hundred CPU cores. The available number of $(m_{\tilde{\chi}_2^0}, m_{\tilde{\chi}_1^0})$ points in the same signal simulation is about 1000. The amount of floating point operations required to evaluate a neural network parametric in both $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$ would be approximately $8 \cdot 10^{19}$. This would require several weeks of computing time on the infrastructure that was available for this work. This is too long considering that using two parameters would only improve the results marginally compared to one parameter and that the neural network evaluation has to be carried out many times during the course of the analysis development.

Verifying the agreement between the observed and expected event yields and the robustness of the results also becomes considerably more challenging when using a parametric neural network. Presenting the results to a reader is an additional challenge since all event yield distributions used to obtain the analysis results can not be shown in a journal paper. There are simply too many of them. Plots such as those in Fig. 9.33 to 9.36, of a kind which has not been used in any earlier CMS analysis, are made to be able to concisely present the results. When using more than one parameter, the verification of the results and the subsequent visualization would become even more difficult. The aforementioned summary plots can not easily be extended to multiple parameters. In addition, this analysis was also the first supersymmetry search to use parametric machine learning in practice. For all the reasons listed above, we decided to use a single new physics parameter in the neural networks not to drive the complexity, both computationally and in terms of the interpretation of the results, too high.

Aside from the new physics parameter δm , the neural networks require a set reconstructed event quantities to determine whether an event is signal- or background-like. The following input variables are found to be sensitive to superpartner production in $3\ell A$ events and are used as inputs to each of the four neural networks:

- Invariant mass of the dilepton pair forming the best Z boson candidate ($M_{\ell\ell}$): The leptons forming the OSSF pair of closest invariant mass to m_Z are assigned to the Z boson decay. The dominant background contribution in $3\ell A$ events comes from SM WZ production in which $M_{\ell\ell}$ peaks around m_Z . For slepton-mediated superpartner decays $M_{\ell\ell}$ depends on δm and the slepton mass parameter x , and has no peak at m_Z . In events with WZ-mediated superpartner decays $M_{\ell\ell}$ can be used to separate signal events from SM processes without a $Z \rightarrow \ell\ell$ decay. The structure of $M_{\ell\ell}$ in SM WZ production and WZ-mediated superpartner decays also differs substantially at low $M_{\ell\ell}$ because of the interference between the WZ and $W\gamma^{(*)}$ amplitudes in the SM.
- p_T^{miss} : Signal events contain two LSPs that escape detection. For sufficiently high δm values this results in a signal p_T^{miss} spectrum that is much harder than that of the SM background.
- Transverse mass of the lepton not forming the best Z boson candidate and $\vec{p}_T^{\text{miss}}$ (M_T): In SM WZ events M_T is the transverse mass of the W boson if the leptons from the Z boson decay are identified correctly. Beyond the value of m_W , the WZ background distribution falls steeply as a function of M_T (Section 6.3.1). In signal events with WZ-mediated superpartner decays there is no clear cutoff for M_T because of the additional p_T^{miss} induced by the escaping LSPs. In events with slepton-mediated superpartner decays there is no intermediate W boson, resulting in a completely different M_T distribution from background events.
- Scalar sum of the transverse momenta of the three leptons in the event (L_T) and p_T^{miss} : This variable represents the scale of leptonic momentum in events and is generally high for signal events at high δm . Events with nonprompt leptons or photon conversions tend to have low L_T values.

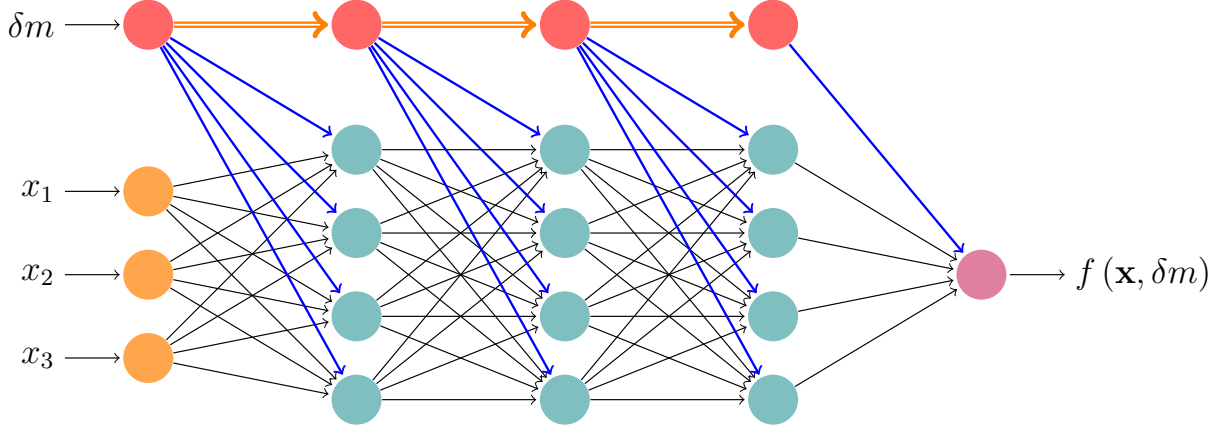


Figure 9.10: Novel architecture of the parametric neural networks used in this search. The new physics parameter δm is not only one of the input variables but is also directly injected into each hidden layer of the neural network. The thick orange lines depict the unity transforms that give each hidden layer direct access to δm . The blue and black arrows are normal weighted connections and at each node the incoming connections are summed after which a simple nonlinear function is applied. Reconstructed event quantities used as input variables are depicted by x_i . The actual neural networks used in the search use the seven input variables listed in the text (not counting δm) and have orders of magnitude more nodes in each hidden layer and have more hidden layers. This architecture differs from the original parametric neural network proposed in [346] which simply uses the new physics parameter as one of the input variables. Our architecture is found to converge faster during training and to yield models with superior final performance.

- Transverse mass of the three leptons in the event and p_T^{miss} ($M_T^{3\ell}$): This variable probes the total mass scale of an event. In signal events $M_T^{3\ell}$ is strongly correlated with δm and takes on much larger values than in SM events.
- Invariant mass of the three lepton system ($M_{3\ell}$): This variable is particularly useful for separating superpartner production from $Z\gamma^{(*)}$ events with photon conversions, in which $M_{3\ell}$ peaks strongly around m_Z .
- Scalar sum of the p_T of all jets in the event (H_T): This variable is sensitive to the amount of ISR in events, which generally differs between signal and background events due to the different initial states. Signal models of WZ-mediated superpartner decays that fall into the WZ corridor have highly similar lepton kinematics and p_T^{miss} distributions to SM WZ production. The H_T variable is therefore of great importance for identifying such signal events.

The neural networks used in this search are fully-connected feed-forward neural networks with a single output value. The parameter δm is used as one of the inputs in the first layer and is also directly injected into each hidden layer of the neural network with skip connections. Having direct access to the parameter at each neural network layer reinforces the learning of the parametrization by implicitly stressing its importance to the neural network. This design differs from the parametric neural network originally introduced in [346] and result in a significantly faster convergence of the neural network training in all of our experiments and better final results. The neural network architecture described here is illustrated in Fig. 9.10.

The neural networks are trained to discriminate signal events from the total SM background. Only simulated background events are used during training so the contribution from nonprompt leptons is also estimated from simulation. Events from each background source are weighted according to the cross section of the corresponding process. For the dominant WZ background alternative simulation samples, simulated with MADGRAPH5_aMC@NLO at NLO in QCD, are used in the training data set.

The WZ simulation samples used in the nominal analysis are generated with POWHEG v2, also at NLO in QCD. This makes the WZ and nonprompt lepton background predictions in the analysis immune to any potential overfitting bias in the neural network, while still retaining a good statistical precision. Of the available set of simulated events, 40% is used for training, 40% for validation, and 20% for testing. The architecture and hyperparameters of the neural networks are optimized by evaluating the results on the validation data set. The optimization and design strategies that went into the neural networks are discussed in detail in Section 9.4.1.3. After fixing the configuration of each neural network, the training and validation data sets are combined to retrain the final version of each neural network. The predictions in the extended training data set are then compared to those in the test data set and the differences are found to be negligible compared to the systematic uncertainties in the analysis. All training data set events, with the exception of those from the WZ and nonprompt lepton backgrounds, are therefore reused in the background prediction of the analysis.

It was anticipated earlier that in the ideal case a parametric neural network can achieve optimal performance at each δm point in the training data set, and is able to interpolate to unseen δm values without loss of performance. We check explicitly whether the four neural networks trained here have both of these properties. The question of whether a parametric neural network is optimal at a particular δm point can be answered by comparing the performance of the parametric neural network at this δm value to a nonparametric neural network trained only at this single δm value. The ability of a parametric neural network to operate at unseen δm points without loss of performance is verified for each δm point by training a parametric neural network with this particular δm point removed from the training data set. At the vetoed δm point the performance is compared to that of the nominal parametric neural network trained on all δm points. If at all available δm points the performance of the neural network remains unchanged after vetoing the δm point in question from the training data set, we can conclude that the neural network can optimally interpolate to any unseen δm value falling in between the minimum and maximum δm values in the training data set. To pass this check the neural network needs to be able to interpolate twice as far in δm as it will ever need to do in the actual search. The reason for this is that in reality it would only have to be able to interpolate in between the δm values in the training data set. The results of the two checks described above are shown in Fig. 9.11 to 9.13. From these studies we conclude that all four parametric neural network achieve optimal performance at each δm point and that their predictions generalize with optimal performance to δm points not seen during training.

Care has to be taken to compare the performance of neural networks trained with different δm distributions and signal event yields in the training data set, as is done in the studies explained above. To choose an adequate metric for the performance comparisons, we write out what the output the neural networks really means. Each neural network used here has a single output value given a set of event quantities $\mathbf{x}$ and a value for the parameter δm . This output is the estimated probability that the event with quantities $\mathbf{x}$ is a signal event given that we search at the mass splitting δm . We can write:

$$\text{neural network}(\mathbf{x}, \delta m) \approx p(\text{signal}|\mathbf{x}, \delta m) \quad (9.4)$$

$$= \frac{p(\mathbf{x}, \delta m|\text{signal}) p(\text{signal})}{p(\mathbf{x}, \delta m)} \quad (9.5)$$

$$= \frac{p(\mathbf{x}|\delta m, \text{signal}) p(\delta m|\text{signal}) p(\text{signal})}{p(\mathbf{x}, \delta m)} \quad (9.6)$$

$$= \frac{p(\mathbf{x}|\delta m, \text{signal})}{p(\mathbf{x}|\delta m)} \cdot \frac{p(\delta m|\text{signal}) p(\text{signal})}{p(\delta m)} \quad (9.7)$$

$$= \frac{p(\mathbf{x}|\delta m, \text{signal})}{p(\mathbf{x}|\delta m)} \cdot p(\text{signal}) \quad (9.8)$$

We use Bayes' theorem to obtain Equation 9.5. Applying the general product rule of probability theory to respectively the numerator and the denominator results in Equations 9.6 and 9.7. Equa-

tion 9.8 holds because we constructed the training data set of each neural network so that background events have an identical δm distribution to signal events. We suggestively split the final two forms in two terms. The first term is the ratio between the probability density for observing a signal event with quantities $\mathbf{x}$ at a parameter value δm and the general probability density for observing such an event given the parameter value δm . Considering that $p(\mathbf{x}|\delta m, \text{background}) = p(\mathbf{x}|\text{background})$ holds because each background event is assigned a random δm value in the training data set we can interpret this term as the degree to which signal events with a given δm can be separated from the SM background. This term can be compared between any two trained neural networks, regardless of the number of signal events and signal δm distribution in the training data set. The second term is arbitrary because it depends on the number of simulated signal events in the training data set. This term is also trivial for any neural networks to learn, so we could easily factor it out of a trained neural network. From this discussion we can conclude that we need to compare neural network performances in terms of a metric that is invariant under the relative number of signal events. The AUC introduced in Chapter 8 fulfills this criterion. The ROC curve of any classifier is invariant under the prior class distributions. Imagine training two binary classifiers c_1 and c_2 on the same problem but with different prior class probabilities. In the idealized case, if a threshold t on the output of c_1 has a signal efficiency s and a background rate b then a threshold t' exists on the output of c_2 that also results in an identical signal efficiency s and background rate b . This motivates the comparisons of the performance of trained neural networks in terms of their AUC in Fig. 9.11 to 9.13.

The performance of the parametric neural networks as a function of δm (Fig. 9.11 to 9.13) and their expected output distribution in SM WZ production events (Fig. 9.8 and 9.9) reveals interesting physics. For WZ-mediated superpartner decays a clear ridge in the neural network output distribution, with higher background yields at signal-like neural network values, is seen when δm approaches m_Z . At the same δm values a sharp drop in the classification performance is observed. This is the infamous WZ corridor where the similarity between SM and signal events limits the ability of the neural network to correctly classify them. A similar structure is also seen for slepton-mediated superpartner decays, but far less pronounced because no W and Z bosons are involved in the superpartner decays. At higher δm values a clear downward slope as a function of δm can be seen in the expected yields at fixed neural network output values. Simultaneously, the classification performance clearly increases at high δm . This matches the physical intuition that signal events with a larger mass splitting between the superpartners are easier to separate from the background.

To use the parametric neural networks to search for superpartner production we must define a way in which events are binned as a function of the neural network output. The output shapes of the neural networks change significantly as a function of δm , and are also different for each of the four neural networks. This is clear in Fig. 9.8 and 9.9. The variable output shape makes it impossible to define a single set of bins in the raw neural network output that is satisfactory at each δm value. In general a narrower binning results in better search results because more information is preserved. On the other hand, each bin should have a sufficient expected SM yield to ensure the background estimation remains reliable. When fitting a neural network the exact output shape of the neural network does not matter because it depends on the arbitrary prior signal probability as seen in Equation 9.8. Only the first term in the equation, the extent to which the signal and background shapes differ, matters. We exploit this and transform the output distribution of each neural network at every evaluated δm value. The transformation is specified so that when 3 ℓ A events are uniformly binned in the neural network output the last bin always contains one expected background event in the 2016 data set⁸. Each preceding bin is defined to contain twice the yield of the next bin, and the first bin contains the remainder. This binning scheme results in good expected exclusion limits at all δm values, while simultaneously guaranteeing that every bin at any δm value has a sufficient expected yield for a robust background estimation. The reason for requiring at least one expected event per bin in the 2016 data set is that it is the smallest of the three collision data sets. The background prediction is done separately for

⁸Because of the finite number of simulated and sideband data events used in the background prediction, the actual expected yields in each bin might deviate from the targeted values by a small margin.

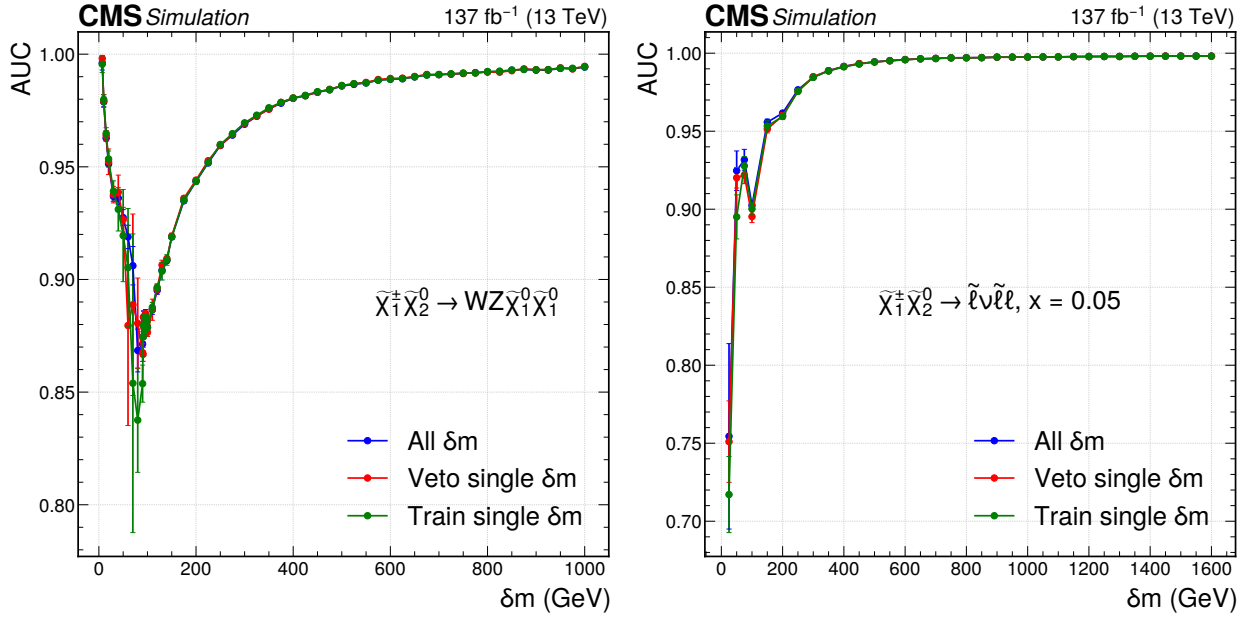


Figure 9.11: Studies of the parametrization of the neural network used to search for WZ-mediated superpartner decays (left) and the neural network used to search for slepton-mediated superpartner decays at $x = 0.05$ (right). The blue curve in each figure shows the AUC performance in the validation data set for the nominal parametric neural network trained at all δm values available in the signal simulation. Each point on the green curve shows the performance at a particular δm value of a parametric neural network trained at all δm values except for the one on the x-axis of that particular point. The points on the red curve show the performance of nonparametric neural networks trained while solely using signal events corresponding to the represented δm point. Each neural network is trained ten times to assess the performance variations due to the random initialization of its internal weights. Each green and red point thus correspond to ten separate neural network training iterations, while the entire blue curve is also obtained from ten training iterations. The drawn points show the mean across the ten iterations and the error bars the estimated standard deviation. From these figures we conclude that the parametric neural networks achieve optimal performance at each δm point in the training data set and that they generalize to any unseen δm point between the minimum and maximum δm values in the training data set without loss of performance.

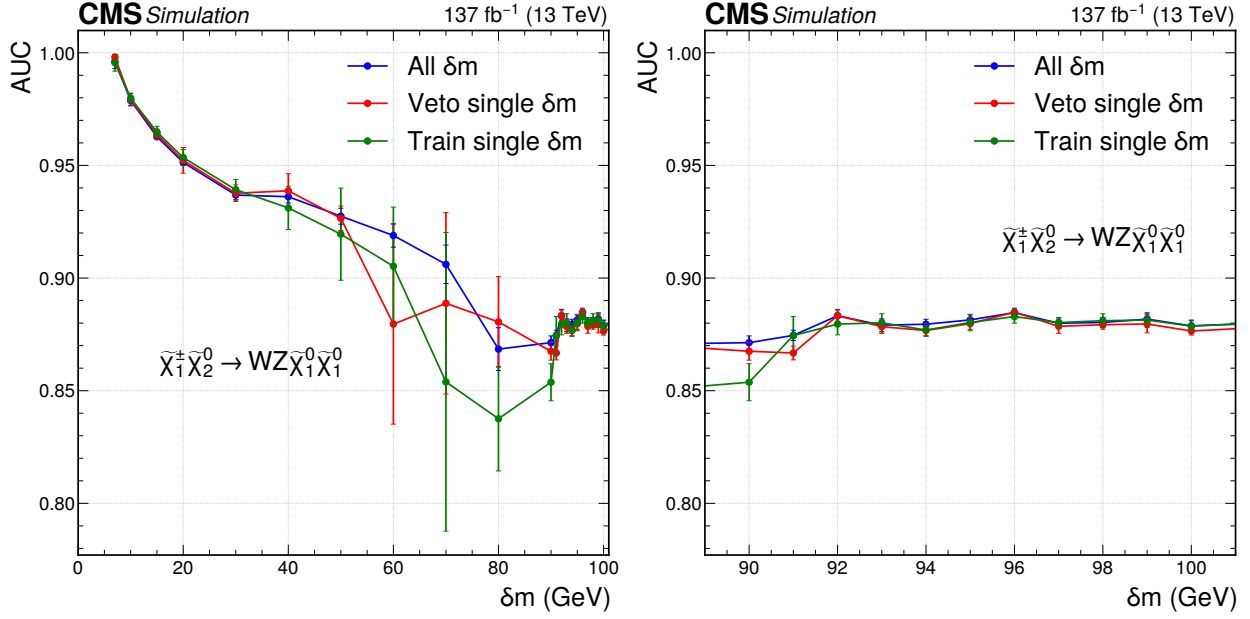


Figure 9.12: Studies of the parametrization of the neural network used to search for WZ-mediated superpartner decays at low δm values (left) and in the WZ corridor (right). These figures zoom in on parts of the left-hand side of Fig. 9.11 and the contents are explained in the corresponding caption.

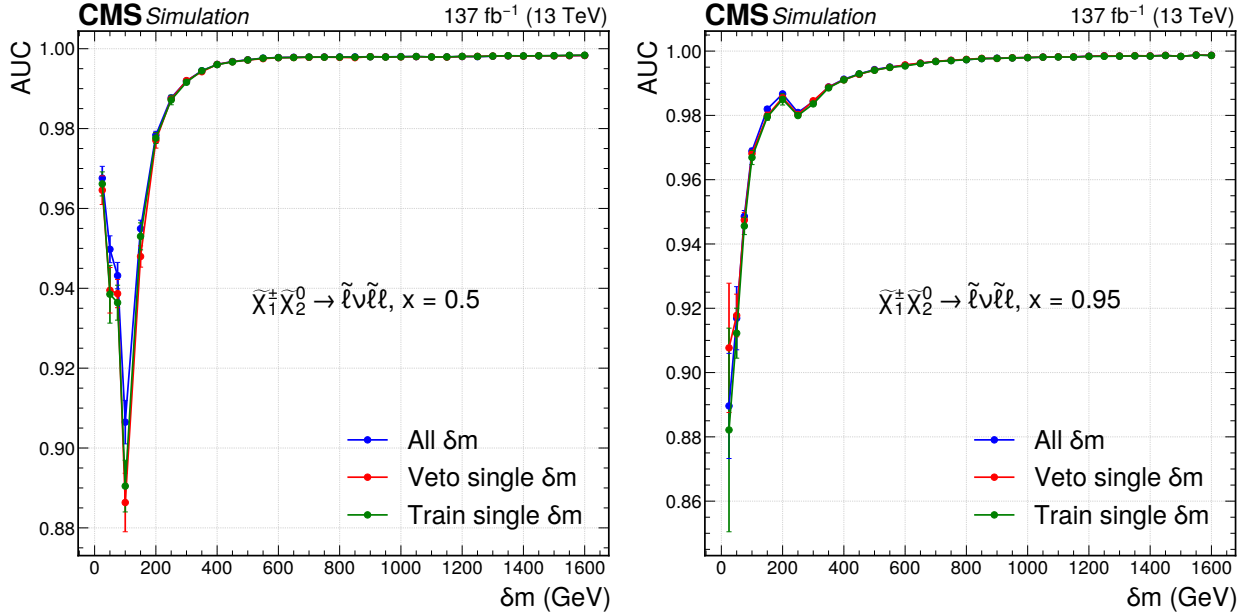


Figure 9.13: Studies of the parametrization of the neural network used to search for slepton-mediated superpartner decays at $x = 0.5$ (left) and at $x = 0.95$ (right). The contents of the figure are explained in the caption of Fig. 9.13.

each data-taking year and a high enough expected background yield is necessary in each bin. The results of the binning scheme can be seen in Fig. 9.19 to 9.22 when the analysis results are shown. In each histogram the expected number of SM events has approximately the same distribution although different events fall into different bins in each of these histograms. The expected signal event yield distributions are determined by the extent to which particular signal hypotheses can be separated from the background and vary strongly with δm .

It has been mentioned before that neural networks can only be assumed to interpolate in the manifold spanned by the training data set and can not safely extrapolate outside of this manifold. When given input variable combinations that fall outside of the training data set manifold, the output of the neural network could take on any arbitrary value allowed by its architecture. Pathological events in real collision data therefore present a possible danger to the neural networks. Mismeasurements or detector noise might result in large or otherwise unexpected values for one or more of the neural network input variables. Such an event might be neither signal- nor SM-like. No guarantees on the prediction of the neural networks can be given for such events, a situation which is to be avoided. For that reason we impose a cutoff on each of the input variables. When a variable exceeds its cutoff, the cutoff value is passed to the neural network instead of the real value of the variable. This ensures that no event in the collision data processed by the neural networks can fall in a regime where the neural networks have undefined behavior. The maximum allowed value for $L_T + p_T^{\text{miss}}$, H_T , $M_{3\ell}$, and $M_T^{3\ell}$ is 2 TeV, while p_T^{miss} and $M_{\ell\ell}$ are limited to 1 TeV, and M_T to 500 GeV. These values are chosen to guarantee safety in the neural network evaluation, while limiting the loss of analysis sensitivity.

9.4.1.3 Neural network optimization

In what follows the technical design choices that went into the neural network classifiers are laid out. Some of this discussion is relatively dense and requires a deep knowledge of the inner workings of neural networks. An introduction to all the necessary aspects of neural networks is beyond the scope of this dissertation and can be found in [359]. This section can be skipped without losing any insight in the physics behind the analysis, but is included to make the results reproducible by future analyzers.

Except for the parameter δm all the input variables to the neural networks are high-level kinematic quantities defined based on knowledge about the physics processes involved. Each of these is found to provide discrimination power between signal and background events. A well studied property of deep neural networks is that if they are sufficiently deep they can, given enough data, efficiently learn arbitrary nonlinear functions of the input variables. Neural networks often achieve this by hierarchically building higher level knowledge from the inputs at each layer. This means that a neural network can be trained to find new physics using only low-level inputs such as $\vec{p}_T^{\text{miss}}$, the Lorentz vectors of the leptons and jets in the event, and the lepton charges. No specialized knowledge in constructing discriminating variables is required. A neural network can implicitly learn everything it needs and might construct high-level variables similar to those designed by human physicists in its hidden layers. In [360] it is demonstrated that studying the internal states of neural networks might even lead to new physics insights. A general procedure for understanding what a neural network is learning internally is introduced in [361]. What is more, deep neural networks trained on low-level inputs are found to generally outperform those trained with human engineered high-level inputs. This is demonstrated in [357, 358]. The reason for this is that a set of low-level input variables generally contains more information about the event than precomputed high-level inputs.

Nonetheless, we proceeded to train neural networks with the list of high-level input variables given in the previous section. To a good approximation these variables contain all the event information that can be used to classify the signal and background events in the search at hand. This is verified for each neural network by training and optimizing an additional neural network that has access to only low-level inputs and another one that has access to both low- and high-level inputs. The performance of the nominal neural networks, trained with high-level inputs, is equal to those of the other neural networks. This shows that any feature the neural networks build from low-level inputs in order to separate signal and background events, can also be constructed from the set of high-level

input variables used here.

If performance is equal, using high-level input variables is more convenient than using low-level ones. The neural networks necessary to achieve optimal performance on preconstructed high-level inputs are generally smaller than in the case of low-level inputs, and less training data is needed. It is also easier to verify that the expected neural network output distribution in collision data can be accurately predicted by simulation. When feeding low-level inputs to a neural network one has no control over what variables are built internally and guaranteeing their accurate modeling in simulation is difficult. That being said, this difficulty persists to a large extent for high-level inputs because one still does not necessarily know what functions of the input variables the neural network uses. In general it is best to directly compare the simulation of the neural network output to real collision data to be confident in the robustness of simulated predictions. Such checks are done in several signal depleted data sets, as explained in Section 9.5.

The set of input variables used in the final neural networks is further studied to assess whether it is a minimal set. To verify that no input variable is superfluous, neural networks are trained with and without each of the input variables. Removing any of the input variables listed in Section 9.4.1.2 reduces the classification performance of all four of the neural networks. Trainings with additional high-level input variables, such as the number of jets in the events and invariant masses of alternative lepton pairings, are also performed. In all cases the performance of the neural networks is found not to increase with additional input variables. From these findings and the studies with low-level inputs we conclude that the set of input variables given in Section 9.4.1.2 is a minimal set of inputs that contains all necessary event information for an optimal classification of signal and background events.

Given an event and a parameter value δm the neural networks give a single output value, namely the probability that an event is from superpartner production. In each neural network a sigmoid function is applied to the output of the final neural network layer to ensure that the outputs can always be interpreted as probabilities. A sigmoid function is defined as:

$$\text{sigmoid}(x) = \frac{1}{1 + \exp(-x)} \quad (9.9)$$

and always has an output value between zero and one.

The objective function used to train the neural networks is the cross entropy between the predicted probabilities and the ground truth labels in the training data set. In other words, the neural networks attempt to minimize the negative log-likelihood of the classification predictions over the training data set. The optimization procedure used to carry out this fit is a variant of the Adam algorithm [362] that uses Nesterov momentum [363]. The basic algorithm for training neural networks is minibatch gradient descent. A small batch of training data examples is used to compute the gradient of the objective function with respect to all of the neural network weights. The weights of the neural network are updated with this gradient to lower the objective function⁹, so the neural network “learns” from the training data examples. The learning rate is a value used to scale the gradient when updating the weights. A large learning rate results in large weight updates from each data sample. During the initial training phases of a neural network large weight updates are desirable, but smaller and smaller weight updates become preferable as the training progresses. This is handled in modern neural network training setups by using adaptive optimizers that keep track of the updates that have been seen previously for each weight in the neural network. The learning rate for each weight can then be adjusted accordingly. Another important concept in the optimization procedure is “momentum”. At a particular training step the gradient at the previous time step can be added to the newly computed gradient, emulating physical momentum like that of a ball rolling down a hill. This makes the gradient descent procedure less likely to get stuck in a local, nonoptimal, minimum. The procedure can be further improved by using the current momentum to give a first estimate of

⁹Conventionally neural network training is always done by means of the minimization of an objective function. Maximization of a reward would be equivalent. An example is the equivalence between the minimization of a negative log-likelihood and the maximization of a likelihood.

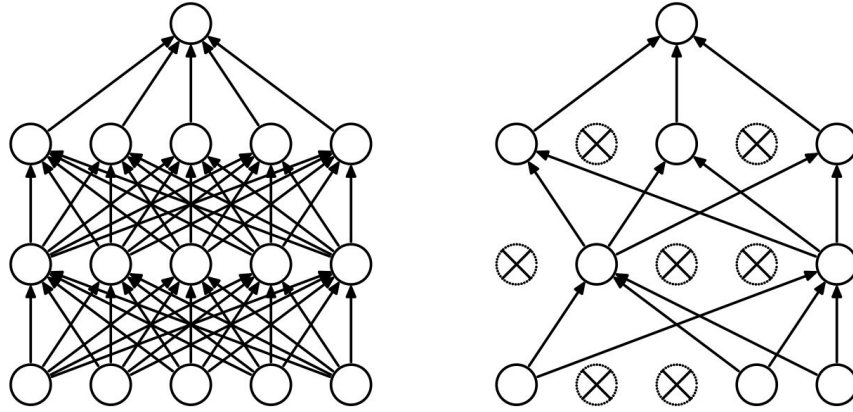


Figure 9.14: Dropout drops different random connections in a neural network for each event in the training data set. A small fully-connected neural network is drawn on the left and a random subnetwork sampled by applying dropout is shown on the right.

where the neural network weights will end up at the next time step and then computing the gradient there. This is effectively momentum with precognition and is known as Nesterov momentum. It makes the optimization procedure less likely to overshoot a minimum in the objective function. The Adam algorithm combines momentum with adaptive weights and is one of the most powerful and robust gradient descent algorithms at the time of writing. An overview of gradient descent procedures can be found in [364], and all gradient descent algorithms listed in this reference were tried during the optimization of the neural network training configuration.

Batch normalization [365] layers are used after each hidden layer. When training a neural network, the gradient of the objective function with respect to the weights is computed for a batch of samples at each training step. This gradient is propagated backwards through the network. Changes in the weights of the deepest layers have to be compensated for in the earlier layers, reducing the rate at which these layers can learn new useful information. This phenomenon is known as internal covariate shift and is detrimental to the training of deep neural networks. Batch normalization normalizes the output of each hidden layer over the training data batch so that the output distribution of the layer is centered at zero and has a unit variance. This eliminates internal covariate shift to first order. During training a running average of the per batch mean and standard deviation of the outputs of each hidden layer are maintained. In this running average the contribution from older training batches decays exponentially. The output mean and standard deviation values for each layer at the end of training are used during inference of the neural network on new data.

The neural networks are regularized by applying dropout [366] to the outputs of all hidden layers. For each single training data event dropout randomly drops connections between layers with a given probability. This can be seen as randomly sampling a smaller subnetwork to train on each event. A neural network trained with dropout is effectively an ensemble of smaller neural networks. Each connection inside the neural network is forced to become independently useful because it can not rely on the presence of any other connection during training. The application of dropout on a small neural network is illustrated in Fig. 9.14. The probability with which connections are dropped is fine tuned together with the other free hyperparameters in the network. The inclusion of dropout in neural networks is known to significantly reduce the bias from overfitting the training data set, thus boosting the performance of the neural network on real collision data.

A colossal amount of design and hyperparameter choices have to be made for each neural network. In this work the following properties of each neural network are all varied when optimizing their configuration: The number of hidden layers, the number of nodes per hidden layer, the nonlinear function used after each hidden layer, the dropout probability in each hidden layer, the gradient descent optimization algorithm, the learning rate, and the size of the mini batches of training data

used to compute each gradient step. There is no general prescription for how these parameters should be chosen. We constrain the number of nodes to be identical in each hidden layer and use the same nonlinear function and dropout probability for each hidden layer to reduce the freedom of choice. Even so we are left with a seven dimensional optimization problem to which no gradient based algorithm can be applied to find the optimum¹⁰.

To choose the optimal hyperparameters the neural networks are evaluated in terms of their AUC on the validation data set. The choice of this metric was motivated earlier on the basis of Equation 9.8. Two independent optimization procedures are used. The first method is a simple grid scan in which several thousand configurations are checked. This method is only able to probe a tiny fraction of the possible configuration space. The second method is a custom evolutionary algorithm, based on evolutionary biology, designed specifically for this analysis. At first, a group of neural networks with random configurations is trained and the performance of each neural network is evaluated. Neural network configurations belonging to the top half of the population in terms of performance are kept, while the rest is discarded. Subsequently, a new generation of neural networks is spawned by randomly mixing the configurations of the remaining 50%. In the mixing step individuals are randomly put into nonexclusive pairs and no configuration can be paired with itself. Given a hyperparameter h_1 of one neural network in a pair and a hyperparameter h_2 of the other neural network, a mixed hyperparameter is generated as:

$$h_{\text{mixed}} = \beta h_1 + (1 - \beta) h_2, \quad (9.10)$$

where β is a random number sampled from a uniform distribution between zero and one. This mixing procedure is repeated for all hyperparameters to create a new neural network configuration. For integer or boolean hyperparameters the result of Equation 9.10 is replaced by the closest allowed value. Nonnumerical hyperparameters are represented by integers spanning the range of possible hyperparameter options. With this mixing procedure all neural networks that were discarded earlier are replaced by new ones to spawn a new generation of neural network configurations.

The procedure of training, discarding the worst performing neural networks, and making new configurations through mixing, is repeated many times. After each generation the average performance of the neural network population increases. To make sure the exploration of the configuration space does not become stale because it is limited by the randomly chosen population of neural network configurations in the first generation, a mutation step is introduced. After spawning a new generation, a small number of neural network configurations is randomly chosen and a random number of their hyperparameters is set to random values. This ensures that each generation keeps trying to explore previously unseen regions of the configuration space. The mutation step in the evolutionary algorithm serves a similar purpose as it does in evolutionary biology.

The evolutionary algorithm is found to converge quickly in practice. It finds the same final performance as the grid scans but requires more than an order of magnitude fewer neural network trainings to do so. The fact that the grid scan and evolutionary algorithm give the same performance builds confidence that the neural network configurations are optimized well and also shows that the performance of the neural networks is relatively robust with respect to the chosen hyperparameters. Variations in most hyperparameters only result in very small variations in the performance of the neural networks.

As a final check parametric and nonparametric BDT classifiers are trained to solve the same task as the parametric neural networks. Their configuration is also optimized with the evolutionary algorithm described above. We find that BDT classifiers give unilaterally worse results than the optimized neural networks.

¹⁰The careful reader will observe that we ended up with a difficult optimization problem once again. This is precisely the situation we wanted to avoid in the first place by using neural networks rather than search regions. The essential difference is that the optimization of the neural network hyperparameters, after a reasonable base design has been chosen, is a matter of squeezing out the final metaphorical drops of performance from the analysis data set. The optimization of search regions, on the other hand, is absolutely crucial and changing the set of search regions in an analysis can easily shift the expected exclusion limits by several factors. Even without a dedicated optimization of the hyperparameters a neural network vastly outperforms an orthogonal search region strategy.

Table 9.3: Definitions of the search regions for events with three light leptons at least two of which form an OSSF pair, and which satisfy $75 < M_{\ell\ell} < 105$ GeV (labeled AXX).

3 ℓ A: three light leptons with at least one OSSF pair, $75 < M_{\ell\ell} < 105$ GeV				
M_T (GeV)	p_T^{miss} (GeV)	$H_T < 100$ GeV	$100 \leq H_T < 200$ GeV	$H_T \geq 200$ GeV
0–100	50–100	A23	A36	A49
	100–150	A24	A37	
	150–200	A25	A38	A50
	200–250	A26	A39	
	250–350	A27	A40	A51
	≥ 350			A52
100–160	50–100	A28	A41	A53
	100–150	A29	A42	A54
	150–200	A30	A43	A55
	200–250	A31	A44	A56
	250–300			A57
	≥ 300			A58
≥ 160	50–100	A32	A45	A59
	100–150	A33	A46	A60
	150–200	A34	A47	A61
	200–250	A35	A48	A62
	250–300			A63
	≥ 300			A64

9.4.1.4 Search regions

To probe WZ-mediated and flavor-democratic slepton-mediated superpartner decays the nominal search in 3 ℓ A events uses the neural networks introduced earlier. Here we define a classical set of search regions which facilitate easier reinterpretations of the results. In addition, 3 ℓ A events add sensitivity to the search for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with τ -enriched slepton-mediated superpartner decays, as well as for models of effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ pair production. For those models it is nontrivial to properly evaluate the neural networks, and their performance might be far from optimal. The search region bins are defined in terms of a subset of the neural network input variables introduced in Section 9.4.1.2: M_T , p_T^{miss} , $M_T^{3\ell}$, and H_T .

Events falling inside the window $75 < M_{\ell\ell} < 105$ GeV are binned in search regions optimized for finding signal events from WZ-mediated superpartner decays with $\delta m \geq m_Z$. The events are binned in M_T , p_T^{miss} , and H_T . Binning in M_T aims to reduce the SM WZ background which falls off rapidly as a function of M_T when $M_T > m_W$ (Section 6.3.1). Signal events often have high p_T^{miss} values because of the escaping LSPs and differences between the signal and background p_T^{miss} distributions get further enhanced at high H_T . The binning in H_T improves the sensitivity to signal models in the WZ corridor in particular. The definition of the search regions in this $M_{\ell\ell}$ window is given in Table 9.3.

In events with slepton-mediated superpartner decays the OSSF lepton pair is produced nonresonantly. Most of these events fall outside of the interval $75 < M_{\ell\ell} < 105$ GeV. Low $M_{\ell\ell}$ events provide sensitivity to low δm scenarios of both WZ-mediated and slepton-mediated superpartner decays, while high $M_{\ell\ell}$ events are used to probe high δm slepton-mediated superpartner decays. Events are divided into a number of $M_{\ell\ell}$ and $M_T^{3\ell}$ bins because of the strong relationship between these variables and δm . The M_T variable is used once again to reduce the SM WZ background. The search region bins are given in Table 9.4.

In the previous iteration of this analysis [5] most of the optimization effort was spent on tuning the search regions that target WZ-mediated superpartner decays. This was especially the case after the refinement of the analysis for the combination of the results with those of other searches in [356]. No

Table 9.4: Definitions of the search regions for events with three light leptons at least two of which form an OSSF pair, excluding those with $75 < M_{\ell\ell} < 105$ GeV (labeled AXX).

3ℓA: three light leptons with at least one OSSF pair, $M_{\ell\ell} \leq 75$ GeV or $M_{\ell\ell} \geq 105$ GeV					
M_T (GeV)	$M_T^{3\ell}$ (GeV)	$M_{\ell\ell} < 50$ GeV	$50 \leq M_{\ell\ell} < 75$ GeV	$105 \leq M_{\ell\ell} < 250$ GeV	$M_{\ell\ell} \geq 250$ GeV
0–100	0–50	A01	A06	A13	A19
	50–100	A02			
	100–400	A03	A07	A14	A20
	≥ 400		A08		
100–200	0–200	A04	A09	A15	A21
	≥ 200		A10	A16	
≥ 200	0–400	A05	A11	A17	A22
	≥ 400		A12	A18	

binning strategy is found in the new analysis that significantly improves the search regions previously defined in the $75 < M_{\ell\ell} < 105$ GeV region, at least without increasing the number of bins. We decide not to increase the bin count to avoid raising the complexity of the search. The search region bins in the $75 < M_{\ell\ell} < 105$ GeV region are therefore left unchanged with respect to [356].

Outside of the $75 < M_{\ell\ell} < 105$ GeV window the new search region strategy is a major improvement compared to what came before. Studies of the input variables to the neural networks revealed $M_{\ell\ell}$ and $M_T^{3\ell}$ to be extraordinarily powerful variables to search for slepton-mediated superpartner decays. Previously, the search regions were mainly defined in terms of M_T and p_T^{miss} , while only three $M_{\ell\ell}$ bins were used. The $M_T^{3\ell}$ variable was not used at all. Using more $M_{\ell\ell}$ bins and replacing the binning in p_T^{miss} by a binning in $M_T^{3\ell}$ results in expected exclusion limits that are up to two times lower compared to those achieved with the old search regions. This is a relative improvement of up to 100%. At the same time, the number of search region bins is decreased compared to the previous analysis iteration.

However large the improvements to the search regions for slepton-mediated superpartner decays are, the results still fall way short of those achieved with the parametric neural networks. For many mass points the neural networks improve the exclusion limits by 100% compared to the new search regions and by 300% (or expected exclusion limits on the signal cross sections that are four times lower) compared to the old ones. More quantitative details on this are given in Section 9.8 where the interpretation of the analysis results is discussed. It should be noted that the search regions can in the theoretical best case scenario only achieve a performance as good as that of a neural network giving a single output distribution over the data set. In practice, even this is not achievable because there is presently no gradient based optimization procedure for search regions. To theoretically be able to achieve results on par with parametric neural networks, the search regions would have to be parametric too. This means that the search region bin thresholds would have to vary as a function of δm .

9.4.2 Three light leptons without an OSSF pair

Events with three light leptons in which none of the leptons form an OSSF pair (3ℓB) do not occur frequently in the SM. Most SM events in this region involve a $Z \rightarrow \tau\tau$ decay with subsequent $\tau \rightarrow \ell\nu$ decays or contain one or more nonprompt leptons. Because of the low SM background yields, 3ℓB events provide sensitivity to WH- and HH-mediated superpartner decays. Production of an H boson often results in two leptons via the $H \rightarrow WW \rightarrow \ell^+\nu\ell^-\nu$ decay chain. The zero spin of the H boson tends collimate the two resulting leptons because of the parity violating nature of W boson decays. Conservation of angular momentum forces the W bosons to have opposite spin. The left-handed lepton ℓ^- is almost always emitted opposite to the W^- boson spin axis, while the right-handed lepton ℓ^+ favors to be aligned to the W^+ boson spin direction. These spin correlations are illustrated in Fig. 9.15. To exploit this feature of H boson decays, 3ℓB events are binned in the minimum angular separation between any two leptons in the event ($\min \Delta R(\ell, \ell')$). Supersymmetric events with WH-

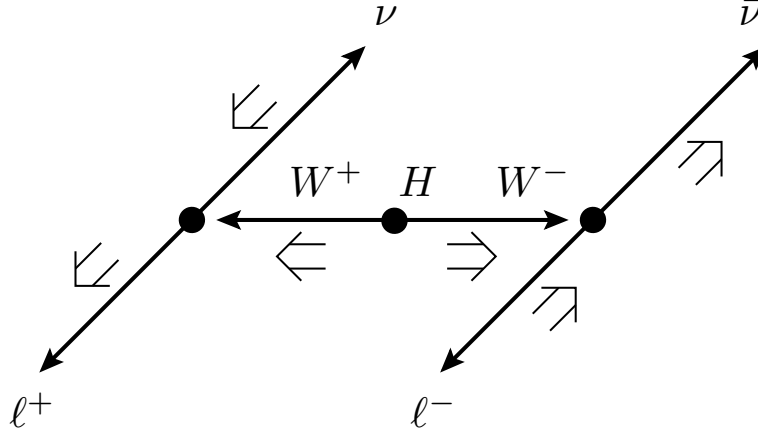


Figure 9.15: Spin correlations in an $H \rightarrow WW \rightarrow \ell^+\nu\ell^-\nu$ decay. The full black lines indicate the directions of the particle momenta while the short doubly lined arrows show the orientations of the particle spins [367].

Table 9.5: Definitions of the search regions for events with three light leptons, none of which form an OSSF pair (labeled BXX).

3ℓB: three light leptons without an OSSF pair		
$\min \Delta R(\ell, \ell') < 0.4$	$0.4 \leq \min \Delta R(\ell, \ell') < 1$	$\min \Delta R(\ell, \ell') \geq 1$
B01	B02	B03

mediated superpartner decays have a $\min \Delta R(\ell, \ell')$ distribution that peaks at low values. Background events, which mainly come from SM WZ production or involve one or more nonprompt leptons, lack strong spin correlations. These events tend to have larger $\min \Delta R(\ell, \ell')$ values than signal events that include one or more H bosons. The background from SM WH production has essentially the same $\min \Delta R(\ell, \ell')$ distribution as the targeted signal events, but is minuscule. The definition of the search region bins in 3ℓB are given in Table 9.5.

The previous version of this search [5] used twice as many 3ℓB search regions, and they were binned in p_T^{miss} , $M_{\ell\ell}$, and M_T . This provides no particular sensitivity to H boson decays. The new approach of targeting collimated leptons results in expected exclusion limits that are more than 50% better for WH- and HH-mediated superpartner decays.

9.4.3 Two same-sign light leptons

Events with two light leptons of the same sign (2ℓSS) are used to recover signal events with compressed WZ-, WH-, or slepton-mediated superpartner decays in which one of the leptons fails to be reconstructed or identified. Some additional event selection criteria are used in category 2ℓSS on top of those listed in Section 9.3 to further reduce the background. Events are vetoed if they contain more than two jets with a p_T above 40 GeV. This requirement reduces the contribution of background events with nonprompt leptons, while still leaving room for some ISR in signal events. Events with a third loose lepton are vetoed, ensuring that category 2ℓSS remains orthogonal to those with three or more leptons. Events with a third lepton passing only the baseline and not the loose lepton selection are retained because we target signal events in which three leptons are produced in the superpartner decay chain. Events in which a third baseline lepton forms an invariant mass within 15 GeV of m_Z with one of the other leptons in the event are removed. This veto reduces the yield from SM WZ production significantly, while having an efficiency of over 99% for the most compressed signal events.

The transverse mass (M_{T2}) [368] of the two leptons and p_T^{miss} is the primary variable used to separate signal and background events. In events with two semi-invisible decays, such as those from

superpartner pair production, M_{T2} is constructed to have a distribution that falls off sharply beyond the mass of the decaying particle. Consider a $WW \rightarrow \ell^+ \nu \ell^- \nu$ event. In the idealized case, the observed $\vec{p}_T^{\text{miss}}$ is the sum of the neutrino transverse momenta:

$$\vec{p}_T^{\text{miss}} = \vec{p}_{T,\nu 1} + \vec{p}_{T,\nu 2} \quad (9.11)$$

The transverse mass of the decay products of a particle, defined in Equation 6.4, is smaller or equal to the mass of the decayed particle. So we have:

$$m_W \geq \max [M_T(\vec{p}_{T,\ell 1}, \vec{p}_{T,\nu 1}), M_T(\vec{p}_{T,\ell 2}, \vec{p}_{T,\nu 2})]. \quad (9.12)$$

The momenta of the two neutrinos in Equation 9.11 are unknown in collision data. One can split up $\vec{p}_T^{\text{miss}}$ into every possible combination of two neutrino momenta. Minimization of the right hand side of Equation 9.12 over every $\vec{p}_T^{\text{miss}}$ split hypothesis gives a result that must also be smaller than m_W . This gives us the definition of M_{T2} :

$$M_{T2} = \min_{\vec{p}_T^{\text{miss}} = \vec{p}_{T,\nu 1} + \vec{p}_{T,\nu 2}} \{ \max [M_T(\vec{p}_{T,\ell 1}, \vec{p}_{T,\nu 1}), M_T(\vec{p}_{T,\ell 2}, \vec{p}_{T,\nu 2})] \} \leq m_W \quad (9.13)$$

The production of two W bosons is used to illustrate the M_{T2} variable, but the properties discussed above hold for any pair of particles of the same mass undergoing semi-invisible decays. For $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production to result in two leptons of the same sign both leptons have to come from the decay of a different superpartner. The M_{T2} of these two leptons and the $\vec{p}_T^{\text{miss}}$ ($M_{T2}(\ell\ell)$) should therefore not exceed the superpartner mass. This makes $M_{T2}(\ell\ell)$ an ideal variable for isolating different signal mass points from each other and to separate a signal from background events which tend to have low $M_{T2}(\ell\ell)$ values.

The $M_{T2}(\ell\ell)$ bins are further split in the p_T of the dilepton system ($p_T^{\ell\ell}$) and p_T^{miss} , both of which are usually large for signal events at sizable δm values. The pp nature of LHC collisions makes events with lepton pairs of positive sign more common than pairs of negative sign. The ratio of positively to negatively signed lepton pairs in a particular process depends heavily on the underlying production mechanism. This results in different sign distributions for signal and background events. All search regions with a sufficiently high expected background yield are divided in terms of the sign of the lepton pair, either positive or negative. The definitions of the search regions in 2ℓSS are given in Table 9.6.

9.4.4 Three leptons with one or more τ_h candidates

Slepton-mediated $\tilde{\chi}_1^\pm$ decays could strongly favor final states with τ leptons if left-handed sleptons are too heavy to participate in the chargino decay. This is the τ -enriched scenario introduced in Section 9.2. Sleptons of the first and second generations might also be much heavier than those of the third generation. In such a scenario, called τ -dominated, both $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ exclusively decay to τ leptons. Retaining events with one or more τ_h candidates markedly increases the sensitivity of the search to the aforementioned scenarios.

Category 3ℓC consists of events with an OSSF pair of light leptons and a τ_h candidate. This event category primarily provides sensitivity to τ -enriched superpartner decay scenarios. The $\tilde{\chi}_2^0$ decay does not favor τ leptons and always results in an OSSF lepton pair. The dominant background contribution in the 3ℓC event category comes from Drell-Yan production with an additional nonprompt τ_h candidate. The majority of the Drell-Yan background events are removed by requiring the mass of the light lepton pair to satisfy $|M_{\ell\ell} - m_Z| > 15 \text{ GeV}$. Another major background source is $t\bar{t}$ production in which both top quarks decay to a light lepton and one of the jets from the top quark gives rise to a nonprompt τ_h candidate. To separate signal events from this background source, events with low to moderate p_T^{miss} are binned in the transverse mass of the light leptons and $\vec{p}_T^{\text{miss}}$ ($M_{T2}(\ell\ell)$). In $t\bar{t}$ production the $M_{T2}(\ell\ell)$ distribution has an endpoint at m_W . At higher p_T^{miss} values the transverse mass of the same objects used in the $M_{T2}(\ell\ell)$ computation ($M_T^{2\ell}$) is found to be a more effective

Table 9.6: Definition of the search regions for events with two same-sign light leptons (labeled SSXX). The symbols $(++)$ and $(--)$ represent the sign of the lepton pair. Three bins of $M_{T2}(\ell\ell)$ are used. The first bin contains all events where $M_{T2}(\ell\ell)$ is exactly zero and the second bin contains those events in which $M_{T2}(\ell\ell)$ is greater than zero but less than 80 GeV. All other events fall into the third $M_{T2}(\ell\ell)$ bin. If $\vec{p}_T^{\text{miss}}$ lies in between the lepton transverse momentum vectors it can be split into two components, each of which is aligned with one of the transverse lepton directions. As a result, the minimization in Equation 9.13 invariably results in zero for such events [369].

2ℓSS: two light leptons of the same sign				
$M_{T2}(\ell\ell)$ (GeV)	$p_T^{\ell\ell}$ (GeV)	$60 < p_T^{\text{miss}} < 100 \text{ GeV}$	$100 \leq p_T^{\text{miss}} < 200 \text{ GeV}$	$p_T^{\text{miss}} \geq 200 \text{ GeV}$
0	< 70	SS01		
	≥ 70	SS02	SS03 (++)	SS05 (++)
			SS04 (--)	SS06 (--)
0–80	< 30	SS07 (++)		SS09
		SS08 (--)		
	≥ 30	SS10		
≥ 80	< 200	SS11	SS12 (++)	SS14 (++)
			SS13 (--)	SS15 (--)
	≥ 200	SS16	SS17 (++)	SS19 (++)
			SS18 (--)	SS20 (--)

Table 9.7: Definitions of the search regions for events with a $\mu^+\mu^-$ or e^+e^- pair and a τ_h candidate (labeled CXX).

3ℓC: $\mu^+\mu^-$ or $e^+e^- + \tau_h$				
p_T^{miss} (GeV)	$M_T^{2\ell}$ (GeV)	$M_{T2}(\ell\ell) < 80$ GeV	$80 \leq M_{T2}(\ell\ell) < 120$ GeV	$M_{T2}(\ell\ell) \geq 120$ GeV
50–200	≥ 0	C01	C02	C03
200–300	≥ 0	C04	C05	C06
≥ 300	0–250	C07		
	250–500	C08		
	≥ 500	C09		

discriminant because of its strong correlation to δm . The search region definitions used in 3ℓC are listed in Table 9.7.

Events with an $e^\pm\mu^\mp$ pair and a τ_h candidate form category 3ℓD. Category 3ℓE is made up of events with a same-sign light lepton pair and a τ_h candidate. These event categories add sensitivity to both τ -enriched and τ -dominated superpartner decay scenarios. In category 3ℓD the transverse mass of the light lepton pair ($M_{T2}(\ell\ell)$) forms a powerful discriminator against the $t\bar{t}$ background that is dominant in this category. In 3ℓE the transverse mass is computed with the leading light lepton and the τ_h candidate ($M_{T2}(\ell\tau)$), and aids in separating signal events from any background source involving W bosons. Background events with a $Z \rightarrow \tau\tau$ decay are separated from the signal by binning events in the opposite-sign lepton pair mass closest to what is expected for the Z boson. We call this variable $M_{\ell\ell}$ in 3ℓD and $M_{\ell\tau}$ in 3ℓE because any opposite-sign lepton pair must include one of the τ_h candidates in 3ℓE events. For a $Z \rightarrow \tau\tau \rightarrow e\mu$ ($Z \rightarrow \tau\tau \rightarrow \tau_h\ell$) decay the reconstructed Z boson mass distribution peaks around 50 (60) GeV, and this value is taken to be the expectation. In 3ℓE events where all three of the leptons have the same sign, $M_{\ell\tau}$ is set to zero. In both category 3ℓD and 3ℓE, p_T^{miss} is used as an additional search variable. The definitions of the search region bins in 3ℓD (3ℓE) are given in Table 9.8 (9.9).

The last three lepton event category contains events with two τ_h candidates and a single light lepton (3ℓF). This category primarily provides sensitivity to τ -dominated signal models and less so to τ -enriched scenarios because of the large background from nonprompt τ_h candidates. The principal

Table 9.8: Definitions of the search regions for events with a $e^\pm\mu^\mp$ pair and a τ_h candidate (labeled DXX).

3ℓD: $e^\pm\mu^\mp + \tau_h$				
$M_{T2}(\ell\ell)$ (GeV)	p_T^{miss} (GeV)	$M_{\ell\ell} < 60$ GeV	$60 \leq M_{\ell\ell} < 100$ GeV	$M_{\ell\ell} \geq 100$ GeV
0–100	50–100	D01	D06	D11
	100–150	D02	D07	D12
	150–200	D03	D08	D13
	200–250	D04	D09	D14
	≥ 250	D05	D10	
≥ 100	50–200	D15		
	≥ 200	D16		

Table 9.9: Definitions of the search regions for events with a pair of light leptons of the same sign and a τ_h candidate (labeled EXX).

3ℓE: same-sign light lepton pair + τ_h				
$M_{T2}(\ell\tau)$ (GeV)	p_T^{miss} (GeV)	$M_{\ell\tau} \leq 50$ GeV	$50 < M_{\ell\tau} \leq 100$ GeV	$M_{\ell\tau} > 100$ GeV
0–80	50–100	E01	E04	
	100–250	E02	E05	
	≥ 250	E03		
≥ 80	50–150	E06		E08
	150–200	E07		
	≥ 200			

source of these nonprompt τ_h candidates is Drell-Yan production. The invariant mass of the light lepton and the leading τ candidate ($M_{\ell\tau}$) is found often to be large for signal events, and is therefore used in the search region binning. The transverse mass of the leading τ_h candidate, the light lepton, and $\vec{p}_T^{\text{miss}}$ ($M_{T2}(\ell\tau)$) is another variable that is found to separate different signal mass points from each other and the SM background. The bins in $M_{\ell\tau}$ and $M_{T2}(\ell\tau)$ are further split in p_T^{miss} . The set of 3ℓF search regions is defined in Table 9.10. Events with three τ_h candidates are not selected because of an overwhelming background to signal ratio.

Table 9.10: Definitions of the search regions for events with two τ_h candidates and one light lepton (labeled FXX).

3ℓF: two τ_h candidates and a light lepton			
$M_{T2}(\ell\tau)$ (GeV)	p_T^{miss} (GeV)	$M_{\ell\tau} < 100$ GeV	$M_{\ell\tau} \geq 100$ GeV
0–100	50–100	F01	F07
	100–150	F02	F08
	150–200	F03	F09
	200–250	F04	F10
	250–300	F05	
	≥ 300	F06	
≥ 100	50–200	F11	
	≥ 200	F12	

Table 9.11: Definitions of the search regions for events with four light leptons forming two exclusive OSSF pairs (labeled GXX).

4ℓG: four light leptons forming two exclusive OSSF pairs		
$M_{T2}(Z, Z)$ (GeV)	$M_{Z2} \geq 60$ GeV	$M_{Z2} < 60$ GeV
0–150	G01	
150–250	G02	G03
250–400	G04	
≥ 400	G05	

9.4.5 Four or more leptons

The final state of effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production models, introduced in Section 9.2, can contain four leptons. Despite the limited branching fraction of the superpartner decays to four leptons, this final state yields a lot of search sensitivity because of the very small SM background. The event selection described in Section 9.3 includes events with four or more leptons. Events with more than four leptons are not vetoed but are extremely rare. In events where more than four leptons are selected, the four highest- p_T leptons are used to divide the event into the appropriate search category.

Events with four light leptons forming two exclusive OSSF pairs (4ℓG) provide the bulk of the sensitivity to effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production models in which both neutralinos emit a Z boson in their decay. The OSSF lepton pair forming the mass closest to m_Z is considered to be the primary Z boson candidate Z_1 . The remaining two leptons form the second Z boson candidate Z_2 . The distribution of the transverse mass of the two Z boson candidates and $p_T^{\text{miss}}(M_{T2}(Z, Z))$ extends up to $m_{\tilde{\chi}_1^0}$ in signal events before briskly diminishing. This makes $M_{T2}(Z, Z)$ an ideal variable for isolating signal models with different $m_{\tilde{\chi}_1^0}$ values and for separating the signal from the SM background which leans towards low $M_{T2}(Z, Z)$ values. To enhance the search for models in which the $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ pair decays via one Z and one H boson, one of the $M_{T2}(Z, Z)$ bins is further split in M_{Z2} . The 4ℓG search regions are defined in Table 9.11.

An event with four light leptons forming a single or no exclusive OSSF pair, belongs in category 4ℓH. Category 4ℓI accommodates events with a single τ_h candidate and three light leptons. If two τ_h candidates of opposite sign and an OSSF light lepton pair are present in an event, it falls into category 4ℓJ. All other events with two τ_h candidates and two light leptons form category 4ℓK. Events with more than two τ_h candidates are rejected. Categories 4ℓH to 4ℓK all provide sensitivity to $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production events in which one or both neutralinos emit an H boson in their decay. In all of these event categories the same search region bins are used. A potential Z boson candidate (Z_1) is reconstructed by finding the OSSF lepton pair of closest invariant mass to the expectation for a Z boson. If no OSSF lepton pair is present in the event, other combinations of opposite sign leptons are considered when forming the Z boson candidate. The mass of the Z boson candidate (M_{Z1}) is used to distinguish events with a Z boson decay from those including only H boson decays. The two leptons not forming Z_1 are taken to originate from an H boson decay. The angular separation in terms of ΔR of these two leptons (ΔR^H) is used as an additional discriminating variable between signal and background. As in category 3ℓB, the usage of ΔR^H is motivated by the spin correlations in $H \rightarrow WW \rightarrow \ell^+ \nu \ell^- \nu$ decays, illustrated in Fig. 9.15. The set of search region bin definitions used in event categories 4ℓH to 4ℓK is given in Table 9.12.

9.5 Background estimation

An overview of SM processes that form the background in the final states of this analysis is given in Section 6.3. As in earlier chapters of this dissertation, the background yields of processes that solely involve prompt leptons are estimated from simulation. The yields of events with one or more nonprompt leptons are estimated directly from data. The techniques and control regions used for the

Table 9.12: Definitions of the search regions for events with four leptons with one or more τ_h candidates, or without two exclusive light lepton OSSF pairs (labeled XYY).

4ℓH-K: four leptons with one or two τ_h candidates or without two exclusive light lepton OSSF pairs		
M_{Z1} (GeV)	$\Delta R^H < 0.8$	$\Delta R^H \geq 0.8$
0–60	X03	X02
> 60		X01

background estimation in search categories that contain events with three or more light leptons and no τ_h candidates are highly similar to those presented in chapters 7 and 8. This search also covers events with two same-sign leptons and events with one or more τ_h candidates, for which the background estimation procedure differs somewhat from that in events with three light leptons.

9.5.1 Nonprompt leptons

The background from nonprompt light leptons is estimated with the fake rate method following the recipe outlined in Section 6.4. Section 9.3 mentioned that two definitions of tight lepton selection are used in this analysis, one for events with two light leptons of the same sign with or without an additional τ_h candidate, and one for all the other event categories. The former selection definition is more stringent than the latter. This introduces an additional difficulty in applying the fake rate method compared to the other two analyses in this dissertation. Namely that the loose lepton selection criteria must be tuned to simultaneously result in a reliable background estimate for two different tight selection definitions. Different definitions of the loose lepton selection across categories are not used because this would lead to ambiguity in the event categorization. By scanning a large grid of cuts on p_T^{ratio} , p_T -dependent cuts on $\text{DEEPJET}_{\text{closest}}$, and cuts on the electron MVA for electrons, a satisfactory loose selection definition is found. Simulated closure tests for this analysis were shown earlier in Fig. 6.7 when introducing the fake rate method. A 30% normalization uncertainty is assigned to the background prediction for nonprompt light leptons, enough to cover any nonclosure of the method seen in simulation. The background prediction is verified in collision data in a region enriched in nonprompt light leptons. Events with three light leptons that contain either no OSSF pair, or in which all OSSF pairs satisfy $|M_{\ell\ell} - m_Z| > 15 \text{ GeV}$, are selected. The presence of one or more b jets is required to ensure orthogonality with the search regions. The expected and observed yields in this control region are compared as a function of all the search variables and the output of the parametric neural networks at each δm value evaluated in the analysis. The background predictions are found to consistently agree with the data within the 30% uncertainty assigned to the fake rate method. Some of the neural network output distributions in this control region are shown in Fig. 9.16.

The misidentification rate of jets as τ_h candidates is far greater than their misidentification rate as light leptons. Events with one or more nonprompt τ_h candidates are in fact the most important background source in the search categories that include τ_h candidates. Luckily, performing a data-driven estimation of the background yields from nonprompt τ_h candidates is technically much simpler than it is for light leptons. When applying the fake rate method to light leptons we have to perform the fake rate measurement in multijet events, and apply the measurement results to estimate the background in the search regions where nonprompt lepton events mainly come from $t\bar{t}$ and Drell-Yan production. This results in the need to heavily fine tune a loose lepton selection because jets from light quarks and gluons have a different misidentification rate compared to those containing b or c jets. The measurement of the light lepton fake rate in multijet data is further complicated by a large prompt lepton contamination, chiefly from W boson production. None of these complications arise for the τ_h fake rate measurement.

The background contribution with nonprompt τ_h candidates is also dominated by the Drell-Yan and $t\bar{t}$ production processes. The τ_h fake rate in Drell-Yan events is measured by selecting events with a

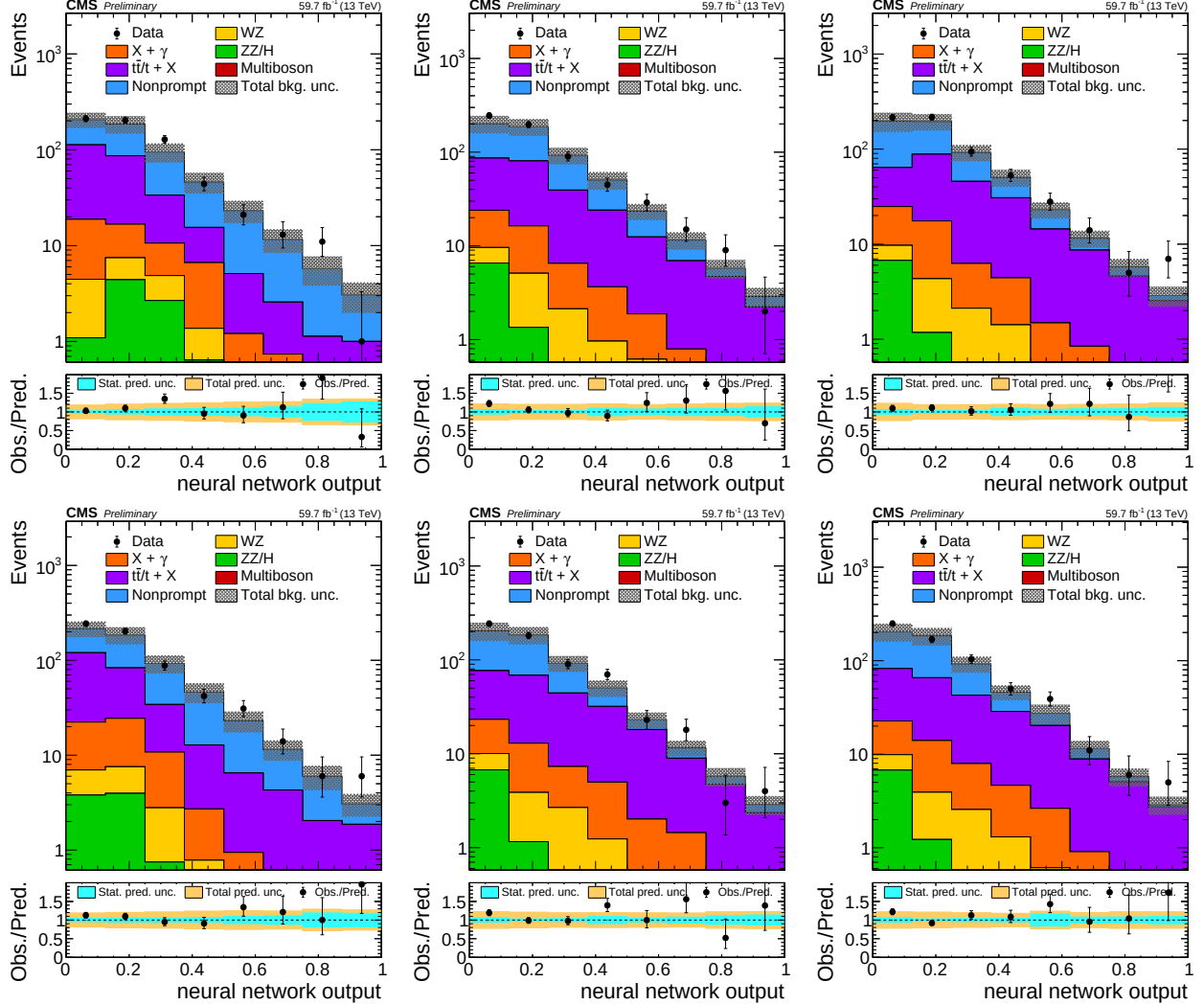


Figure 9.16: Observed and expected number of events in the nonprompt enriched region in 2018 data as a function of the output of two of the parametric neural networks used in the search, evaluated at a number of δm values. The top row shows the yields as a function of the neural network used to search for WZ-mediated superpartner decays, evaluated at δm values of 50 (left), 200 (middle), and 800 GeV (right). The bottom row shows the yields as a function of the neural network that targets slepton-mediated superpartner decays with $x = 0.5$, evaluated at δm values of 100 (left), 300 (middle), and 1000 GeV (right). In each case the neural network output distribution is transformed so that the last bin contains approximately two expected SM events in the 2016 collision data set. The expected yield in each preceding bin is constructed to be twice that in the following bin, and the first bin contains the remainder of the events.

$\mu^+\mu^-$ or e^+e^- pair and a τ_h candidate. The light lepton pair is required to satisfy $|M_{\ell\ell} - m_Z| < 15$ GeV and p_T^{miss} must be below 50 GeV. The fake rate is given by the ratio of the number of events with a loose τ_h candidate to the number of events with a tight τ_h candidate. The contamination from prompt τ_h candidates is estimated from simulation and found to be as good as negligible. The estimated contamination is subtracted from the numerator and denominator of the fake rate computation. The measurement of the fake rate is carried out in different bins of τ_h candidate p_T and $|\eta|$. A similar measurement for the τ_h fake rate in $t\bar{t}$ events is obtained by selecting $e^\pm\mu^\mp$ events with an additional b-tagged jet and a τ_h candidate. The same measurement procedure described above is repeated in this data set. In this region the prompt contamination is as good as negligible as well.

Studies in simulation indicate that Drell-Yan production is the dominant source of events with nonprompt τ_h candidates in $3\ell C$ and $3\ell F$. The nonprompt τ_h candidate background prediction in these event categories therefore uses the fake rate measured in Drell-Yan enriched events. In $3\ell D$ and $3\ell E$, on the other hand, $t\bar{t}$ production is found to make up the bulk of events with nonprompt τ_h candidates, and the corresponding fake rate measurement is used in the background prediction. As for light leptons, a 30% normalization uncertainty is assigned to the prediction of the background from nonprompt τ_h candidates with the fake rate method. This uncertainty is large enough to cover any discrepancy seen in the fake rate method for simulated τ_h candidates. The uncertainty is taken to be uncorrelated between the $t\bar{t}$ and Drell-Yan fake rate measurements. The fraction of the nonprompt τ_h candidate background in $3\ell C$ and $3\ell F$ that stems from $t\bar{t}$ production is estimated in simulation, and an additional uncorrelated uncertainty source that is 30% times the $t\bar{t}$ fraction is introduced. A similar uncertainty is propagated to account for the Drell-Yan contribution to the nonprompt τ_h candidate background in $3\ell D$ and $3\ell E$.

9.5.2 Electron sign misidentification

Electron sign misidentification is a considerable background source in $2\ell SS$ and $3\ell E$. The sign misidentification rate for light leptons is reduced significantly in these event categories by the additional requirements imposed in the lepton identification (Section 9.3). For electrons the sign misidentification rate remains high enough to require a precise estimation of the resulting background. In $2\ell SS$ and $3\ell E$ this background is estimated from collision data by applying the probability that an electron sign is misidentified to a sideband of events with two leptons of opposite sign. Aside from the lepton sign combination, the sideband events must pass the same selection requirements as imposed on search region events. For muons that pass the imposed selection requirements the sign misidentification rate is negligible, and the tiny background contribution that remains is estimated from simulation.

The sign misidentification probability of electrons as a function of their p_T and $|\eta|$ is a priori determined in a large cocktail of simulated events from Drell-Yan, $t\bar{t}$, and diboson production. All of these simulated samples are combined to reduce the statistical uncertainty in the prediction of the very small electron sign misidentification rate. An integral data to simulation scale factor is subsequently measured for the sign misidentification predictions in each data-taking year. The scale factor measurements are based on a data sample of events with two tight electrons that satisfy $|M_{ee} - m_Z| < 10$ GeV. The number of events with electrons of the same sign is predicted by applying the sign misidentification probabilities determined in simulation to collision data events with two opposite sign electrons. The ratio of the number of observed and predicted same-sign electron events gives the global scale factor for the sign misidentification predictions¹¹. The scale factors for each data-taking year are given in Table 9.13.

After scaling up the data-driven sign misidentification predictions with the measured scale factors, the number of predicted same-sign electron events as a function of each of the search variables used in $2\ell SS$ and $3\ell E$ is compared to collision data. These comparisons are made in the same $Z \rightarrow ee$ enriched data sample used to measure the prediction scale factors. Some of the closure tests in 2018 data

¹¹Note that this is not a scale factor for the sign misidentification probabilities. Given that both electrons in a $Z \rightarrow ee$ decay could have a misidentified sign, scaling up the charge misidentification probability by a c will increase the predicted background rate by a value $c' < c$.

Table 9.13: Ratio of the predicted event yield for the electron sign mismeasurement background to the observation in a sample of same-sign electron events enriched in $Z \rightarrow ee$ decays. The prediction is made by applying sign mismeasurement probabilities from simulation to collision data events with two opposite-sign electrons. The listed uncertainties are statistical only.

data taking year	2016	2017	2018
data/prediction	0.89 ± 0.03	1.26 ± 0.04	1.17 ± 0.03

are shown in Fig. 9.17. A robust prediction of all the analysis variables is found in each data-taking year. A 20% normalization uncertainty is assigned to the data-driven prediction of the electron sign misidentification background to cover any observed discrepancy.

9.5.3 Prompt leptons

Special care is taken in the estimation of the WZ background because it is the dominant source of SM events in $3\ell A$. A control region enriched in SM WZ production is defined by selecting $3\ell A$ events with the additional requirement $|M_{\ell\ell} - m_Z| < 15$ GeV to enrich the sample in Z boson decays. The contribution from photon conversions is reduced by imposing $|M_{3\ell} - m_Z| > 15$ GeV. The control region should contain the bulk of SM WZ production events while not including potentially sensitive superpartner production events, in particular from WZ-mediated superpartner decays. This is ensured by requiring $50 < p_T^{\text{miss}} < 100$ GeV and $50 < M_T < 100$ GeV. In order to constrain the WZ normalization the control region is included in the fits used to evaluate the analysis results. The control region overlaps with some of the $3\ell A$ search regions, namely A23, A36, and A49 defined in Table 9.3. These search regions are removed when evaluating the analysis in $3\ell A$ with the orthogonal search region approach. When using the parametric neural networks, the overlap between $3\ell A$ and the WZ control region can not easily be removed, so the control region is not included in the corresponding analysis fits. Instead, the bins at low neural network output values, containing events the neural network considers very background-like, implicitly constrain the WZ normalization. Note that the neural network and search region interpretations of $3\ell A$ are always performed separately because they use the same events. The WZ control region also provides utility for checking simulated predictions as a function of the output of the parametric neural networks. The neural network outputs are found to be well modeled by the simulation in the control region. Because of the control region definition the predictions can however only be checked at low to moderate p_T^{miss} and M_T values.

Both the search regions and the neural networks in $3\ell A$ rely on an accurate modeling of the SM WZ background at high p_T^{miss} and M_T values. Very high p_T^{miss} events can occur if the p_T of the W boson is sufficiently high. The value of M_T should however not exceed m_W in a W boson decay. Simulated studies show that WZ events with an M_T value larger than m_W mainly originate from the association of the wrong lepton to the W boson decay. In $e^\pm e^\mp e^\pm$ or $\mu^\pm \mu^\mp \mu^\pm$ events it is ambiguous which leptons originate from the Z boson decay and which lepton originates from the W boson decay. We assign the OSSF lepton pair forming the closest invariant mass to m_Z to the Z boson decay. In rare cases the lepton from the W boson decay accidentally forms a mass closer to m_Z with one of the other leptons, and the lepton assignment becomes incorrect. If leptons are assigned to the wrong decay, M_T no longer has a physical boundary at m_W , resulting in SM WZ production events with high M_T values. Background events with high M_T values due to lepton mispairing are well modeled by simulation because there is no experimental peculiarity to these events.

Significant mismeasurements of $\vec{p}_T^{\text{miss}}$ contribute to the tails of the M_T distribution to a smaller extent than the mispairing of leptons. Such mismeasurements are the most important source of high M_T $e^\pm e^\mp \mu^\pm$ and $\mu^\pm \mu^\mp e^\pm$ events, in which there is no ambiguity in the origin of the leptons. This contribution to the high M_T tails of WZ production is harder to model, so we substantiate the predictions with collision data. To do this a data set enriched in $W\gamma$ events is selected. The WZ and $W\gamma$ processes share a large degree of similarity due to their similar production mechanisms [370].

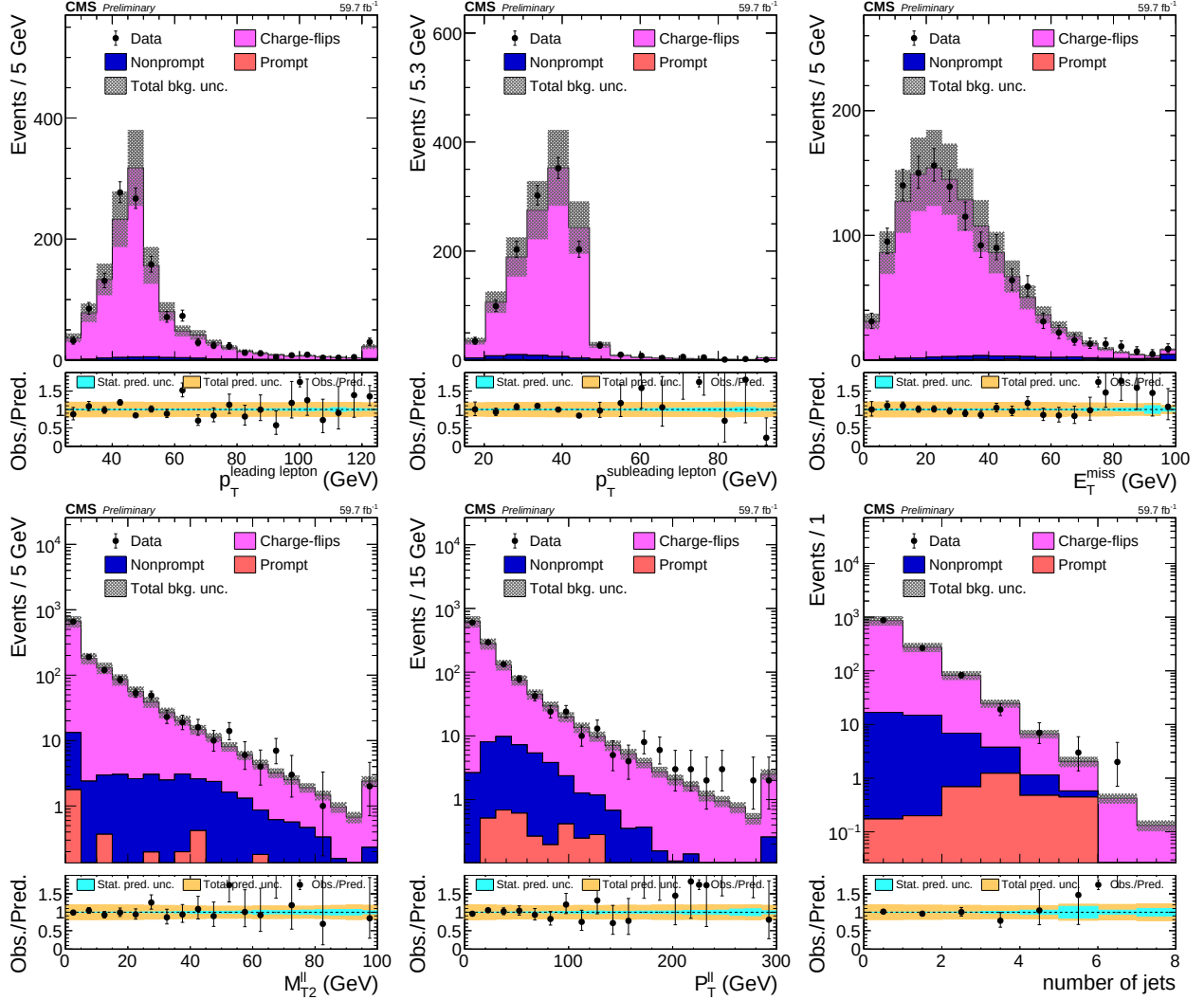


Figure 9.17: Comparison of the number of observed same-sign electron events with $|M_{ee} - m_Z| < 10 \text{ GeV}$ in 2018 collision data to the prediction obtained by applying electron sign misidentification probabilities from simulation to collision data events with two opposite-sign electrons. The predictions are scaled by the integral scale factor given in Table 9.13. The top row shows the comparison as a function of the p_T of the leading (left) and subleading (middle) leptons, and p_T^{miss} (right). The bottom row shows the comparison as a function of $M_{T2}(\ell\ell)$ (left), the p_T of the dilepton system (middle), and the number of jets in the event (right). The total systematic uncertainty includes the 20% normalization uncertainty applied to the data-driven electron sign misidentification background prediction.

The $W\gamma$ enriched sample consists of events with a single muon and a photon in the final state. The muon must satisfy $p_T > 25(28)$ GeV in 2016 (2017 and 2018) data to pass triggers selecting a single muon. As opposed to the Z boson in WZ production the photon in $W\gamma$ events is often the result of FSR. The contribution of FSR photons in the $W\gamma$ enriched sample is reduced by not selecting events with electrons in the final state and requiring $p_T(\gamma) > 40$ GeV, a minimum angular separation of $\Delta R > 0.3$ between the photon and the muon, and $M_T > 40$ GeV. The criterion $p_T^{\text{miss}} > 50$ GeV is introduced for the $W\gamma$ enriched sample to be compatible with the $3\ell A$ selection. Simulated predictions in the $\mu\gamma$ region are compared to data in bins of M_T and p_T^{miss} . The level of agreement between data and simulation in each bin is used to derive an uncertainty in the WZ background prediction at the corresponding M_T and p_T^{miss} values. The uncertainty ranges from 5% at low M_T and p_T^{miss} , up to slightly less than 30% at high M_T and p_T^{miss} . The uncertainty is applied to both the search region and neural network analyses of the $3\ell A$ region. It should be noted that these uncertainties constitute an upper limit on the uncertainty in the prediction of the WZ background. The main source of high M_T events from WZ production, assigning leptons to the wrong decay, is not present in $W\gamma$ production. Events with normal p_T^{miss} measurements but wrongly assigned leptons are far easier to model than those with large p_T^{miss} mismeasurements which dominate the tail of the M_T distribution $W\gamma$ events. Additionally, the presence of a photon makes $W\gamma$ production generally harder to model than WZ production.

Studies of the assignment of leptons to different particle decays reveal an interesting property of the parametric neural networks. The fraction of WZ production events that have leptons assigned to the wrong boson decays in the computation of M_T and $M_{\ell\ell}$ is found not to significantly increase as a function of how signal-like the neural networks consider an event to be. Meanwhile, higher M_T values generally do result in the neural network considering an event more likely to originate from superpartner production, assuming δm is not tiny. These results indicate that the neural networks implicitly learn to recognize events with wrongly assigned leptons by using the interplay between all their input variables. The neural networks are not sensitive to wrong lepton assignments in the computation of their input variables.

The modeling of the neural network predictions without constraints on M_T and p_T^{miss} is directly verified in two control regions which have no requirements on either variable: A region enriched in nonprompt leptons and one enriched in $t\bar{t}Z$ production. The nonprompt lepton enriched region was already introduced when discussing the background prediction for nonprompt light leptons. The $t\bar{t}Z$ enriched region consists of events passing the same selection criteria as those in $3\ell A$, but inverting the b jet veto and adding the criterion $|M_{\ell\ell} - m_Z| < 15$ GeV. Because of the b jet requirement in both control regions there is no overlap with any of the search regions. This means that the simulated predictions of the neural network outputs can be verified without any limitations on M_T or p_T^{miss} . Some of the results in the nonprompt lepton ($t\bar{t}Z$) enriched region in the 2018 collision data set are shown in Fig. 9.16 (Fig. 9.18). Similar closure tests are done in the 2016 and 2017 collision data sets, and at all δm values evaluated in the analysis. Good agreement between the expected and observed yields as a function of the neural network outputs is found.

Control regions enriched in ZZ and $Z\gamma$ production are defined in a similar way to chapters 7 and 8. The ZZ ($Z\gamma$) enriched region consists of $4\ell G$ ($3\ell A$) events in which the p_T^{miss} veto is inverted to $p_T^{\text{miss}} < 50$ GeV. In the ZZ enriched region one of the OSSF lepton pairs must have an invariant mass within 15 GeV of m_Z , while in the $Z\gamma$ enriched region $M_{3\ell}$ must satisfy the same requirement. Both control regions are included in the analysis fits to constrain the normalization of the ZZ and $Z\gamma$ processes. In addition, the predicted neural network output values are compared to the observation in both control regions, and no significant discrepancies are seen.

9.6 Systematic uncertainties

The experimental and theoretical uncertainties in the background prediction are very similar to those in the analyses presented earlier in this dissertation. The general outline of the uncertainty estimation,

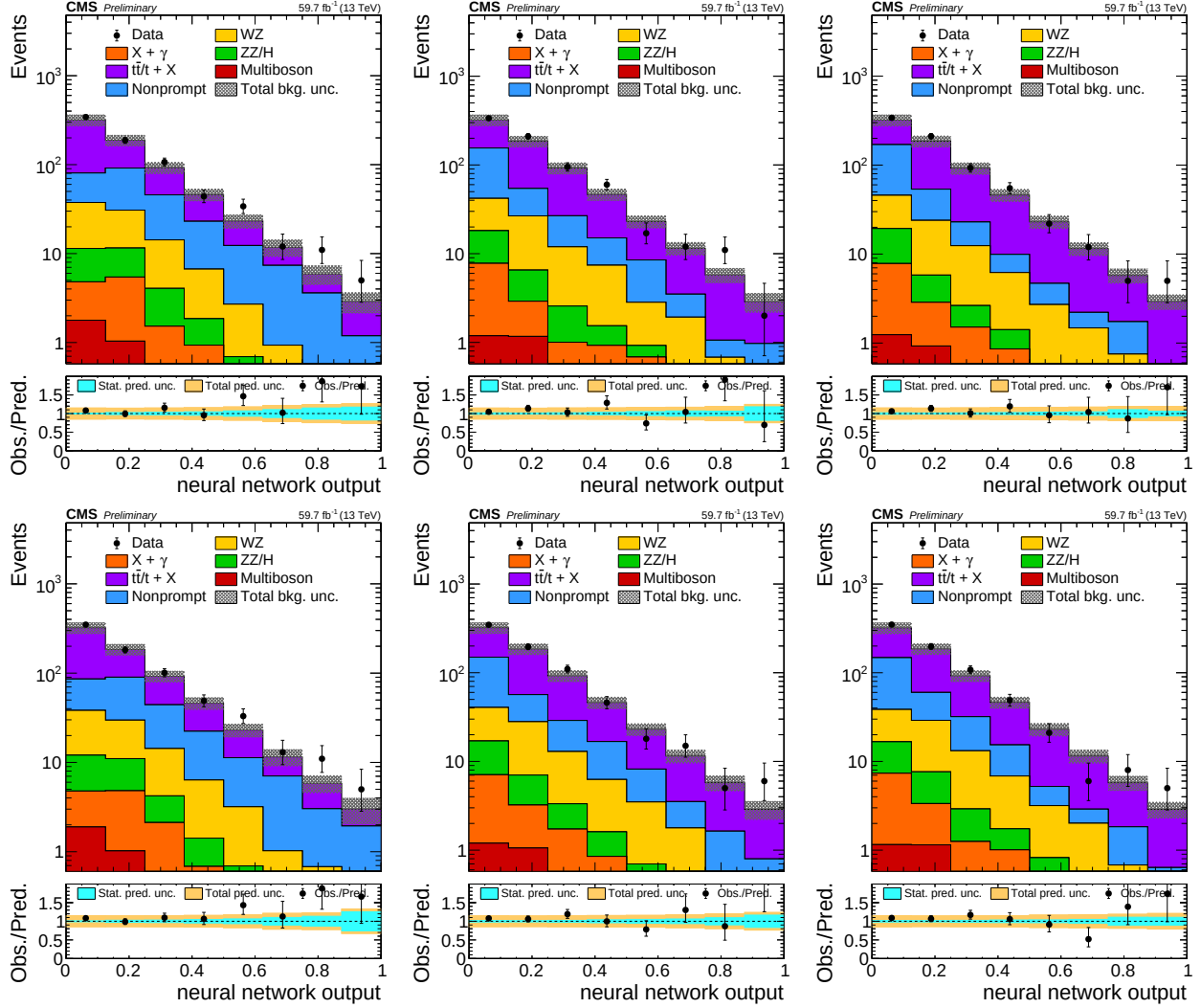


Figure 9.18: Observed and expected number of events in the $t\bar{t}Z$ enriched region in 2018 collision data as a function of the output of two of the parametric neural networks used in the search. The top row shows the yields as a function of the output of the neural network used to search for WZ-mediated superpartner decays, evaluated at δm values of 50 (left), 200 (middle), and 800 GeV (right). The bottom row shows the evaluation of the neural network used to search for slepton-mediated superpartner decays with $x = 0.5$ at δm values of 100 (left), 300 (middle), and 800 GeV (right). In each case the neural network output distribution is transformed so that the last bin contains approximately two expected SM events in the 2016 collision data set. The expected yield in each preceding bin is constructed to be twice that in the following bin, and the first bin contains the remainder of the events.

Table 9.14: Overview of systematic uncertainty sources affecting the search, the typical variations they induce in the expected yields across the analysis bins, and their correlation across the data-taking years. The values listed for general uncertainty sources represent variations in the total expected yields, while those shown for process specific sources are the variations in the yields of the corresponding processes.

Source	Typical uncertainty (%)	Correlation across data-taking years
General uncertainty sources		
e/μ identification efficiency	1–2 per lepton	Correlated
τ_h efficiency	1–3 per lepton	Correlated
Pileup	1–2	Correlated
Integrated luminosity	1.8	Partially correlated
Trigger efficiency	1.4–5	Partially correlated
Jet energy scale	1	Partially correlated
Jet energy scale (fast simulation)	1	Correlated
b-veto efficiency	1–3	Correlated
b-veto efficiency (fast simulation)	1–3	Correlated
PDF	1–10	Correlated
Renormalization and factorization scales	1–10	Correlated
Process specific uncertainty sources		
Signal ISR	1–5	Correlated
Signal p_T^{miss}	1–2	Correlated
WZ shape	5–30	Uncorrelated
WZ normalization	10	Correlated
ZZ normalization	10	Correlated
$X\gamma^{(*)}$ normalization	10	Correlated
Nonprompt normalization ($e/\mu/\tau_h$)	30	Correlated
Charge misidentification normalization	20	Correlated
$t\bar{t}X$ normalization	15	Correlated
Multiboson normalization	50	Correlated

common among analyses, is given in Section 6.5. Additional analysis specific details are given below. An overview of all sources of systematic uncertainty considered in the analysis and the size of their effect on the expected yields across analysis bins is given in Table 9.14.

As in the other analyses, the identification efficiencies for light leptons are measured in a Drell-Yan enriched data set using the tag-and-probe technique (Section 6.1.3). The search is expected to be sensitive to the production of superpartners with masses more than an order of magnitude larger than those of the electroweak bosons. To estimate the impact high superpartner masses might have on the lepton identification efficiencies, the lepton efficiencies are measured twice: Once in events where the p_T of the reconstructed Z boson candidate is below 80 GeV and a second time in events where the p_T of the reconstructed Z boson candidate is above 80 GeV. The effect of a high Z boson p_T on the measured lepton identification efficiencies in a given lepton p_T and η bin gives an idea of the impact the high superpartner masses might have on these efficiencies. The mean in each p_T and η bin between the two efficiency measurements is used as the nominal lepton identification efficiency by which all simulated yields are scaled. The difference between the two measurements is taken as a systematic uncertainty. This uncertainty in the extrapolation to highly energetic events is around 0.5% per lepton for the bulk of leptons from SM processes, but ranges up to 3% for very low and high p_T leptons which are often present in respectively compressed and noncompressed signal models.

Measurements of the efficiency of the τ_h identification algorithms used in the analysis were readily available [232]. The efficiency for τ_h candidates of p_T below about 60 GeV is measured in $\mu\tau_h$ events

enriched in $Z \rightarrow \tau\tau$ decays. Efficiencies for τ_h candidates with intermediate p_T values, up to 100 GeV, are measured in $t\bar{t}$ enriched $\mu\tau_h$ events. Single τ_h events from high mass virtual W bosons are used for measuring the efficiencies in the high p_T regime. The resulting uncertainties range between 1 and 3% per τ_h candidate in an event.

Integrated trigger efficiencies in the different analysis final states are measured in a data set selected with trigger paths independent of those used in the analysis selection. More details were given earlier in Section 9.3. The statistical uncertainties in these efficiency measurements are propagated to simulated events as normalization uncertainties that are independent between data-taking years. Variations between the data-taking years in the data to simulation scale factors of the trigger efficiencies are taken as an additional independent source of systematic uncertainty. This uncertainty source is considered correlated among the data-taking years. The values of the uncertainties in the trigger efficiency scale factors can be found in Table 9.1.

Theoretical uncertainties from the limited knowledge of the proton PDFs and from the choice of the QCD renormalization and factorization scales are estimated in accordance with the recipe presented in Section 6.5. Both sources of theoretical uncertainty result in variations of the integral cross sections and kinematic distributions of simulated processes. The kinematic variations are taken into account for all processes. The cross section uncertainties are only considered for processes without a dedicated data control region in which the normalization is measured. For WZ, ZZ, and $X\gamma$ production, prior normalization uncertainties of 10% are used instead. These priors are constrained based on the data in the respective control regions. The input values of the prior uncertainties does not affect the analysis results.

The prediction of the WZ background as a function of M_T and p_T^{miss} is validated in $W\gamma$ enriched $\mu\gamma$ events, a procedure explained in more detail in Section 9.5. The maximum of the statistical uncertainty in the data and the discrepancy between data and simulation in the $W\gamma$ enriched region as a function of M_T and p_T^{miss} is taken as a systematic uncertainty in the WZ prediction. This uncertainty ranges from 5% at low to moderate M_T and p_T^{miss} to values 30% at high M_T and p_T^{miss} values. This uncertainty largely stems from the limited statistical precision in the $W\gamma$ enriched data sample at high M_T and p_T^{miss} values, rather than a systematic discrepancy between the expectation and the observation.

The simulation of superpartner production events under 2016 data-taking conditions does not adequately model QCD ISR. This problem is addressed by correcting the distribution of the total ISR transverse momentum (p_T^{ISR}) in signal events. The corrections are derived from the observed Z boson p_T spectrum in dilepton events, which is compared to Z boson production events simulated under the conditions as the signal simulations. The entire size of the correction is taken to be a systematic uncertainty in the signal prediction. The signal simulations for 2017 and 2018 data-taking conditions have a more modern configuration (Section 9.2.4), and studies of the Z boson p_T spectrum in 2017 and 2018 collision data reveal that p_T^{ISR} is well modeled in the corresponding signal simulations. A small residual correction of the spectrum of the number of jets per event is applied to these simulations. This correction is also derived from data enriched in Z boson decays. The size of this much smaller correction is also taken as a systematic uncertainty in the signal predictions.

The majority of kinematic quantities used in the search depend on $\vec{p}_T^{\text{miss}}$. Accurate modeling of $\vec{p}_T^{\text{miss}}$ is therefore essential. Simulated signal events use a fast simulation of the CMS detector that is less accurate than the GEANT4 based simulation used for the simulation of background events (Section 9.2.4). The modeling of $\vec{p}_T^{\text{miss}}$ depends on the simulation of all particles and detector components and is more sensitive to mismodeling than other event variables. For several signal mass points the shapes of the p_T^{miss} distributions and those of all analysis variables derived from $\vec{p}_T^{\text{miss}}$ are compared between fast simulation events and GEANT4 based simulation events. The differences are generally small, but nonetheless an additional uncertainty is applied to signal events. The difference between the $\vec{p}_T^{\text{miss}}$ before and after simulating detector and reconstruction effects is taken to be a systematic uncertainty. The resulting uncertainty is significantly larger than the differences observed between the $\vec{p}_T^{\text{miss}}$ in a GEANT4 based and fast simulation ¹².

¹²This overly conservative uncertainty estimate was prescribed to be used in all CMS supersymmetry searches. The

A number of uncertainty sources are taken into account in the nonprompt background prediction. A 30% normalization uncertainty is applied to the background prediction of both nonprompt light leptons and τ_h candidates. This uncertainty in the nonprompt lepton background prediction is considered to be correlated between muons and electrons and uncorrelated between τ_h candidates and light leptons. For τ_h candidates the normalization uncertainty is also considered uncorrelated between the nonprompt background predictions using the fake rate measurements from Drell-Yan and $t\bar{t}$ enriched data. We also account for an additional uncertainty source that addresses the remaining fraction of nonprompt τ_h candidates from Drell-Yan ($t\bar{t}$) when the fake rate measurement in $t\bar{t}$ (Drell-Yan) enriched data is used. This uncertainty is set to 30% multiplied by the contamination fraction, so its size varies from bin to bin. For light leptons an extrapolation uncertainty is applied to account for uncertainties in the fake rate measurement at high lepton p_T values. At high lepton p_T the light lepton fake rate can no longer be reliably measured in QCD multijet enriched data because of the large prompt lepton background from W boson production (Section 6.4.3). In practice, the fake rate values measured for lepton p_T values above 45 GeV are unreliable. For leptons with higher transverse momenta the fake rate for 45 GeV leptons is used. The difference between the fake rate in the last p_T bin with reliable measurements and the fake rate at higher p_T is studied in simulated events. The difference is applied as a fractional uncertainty to the data-driven prediction of events with high p_T nonprompt light leptons. This extrapolation uncertainty is taken to be correlated between data-taking years and among electrons and muons.

A 20% normalization uncertainty is applied to the data-driven prediction of the electron sign misidentification background. The size of this uncertainty is set based on closure tests in events with two same-sign electrons enriched in $Z \rightarrow ee$ decays (Section 9.5).

9.7 Results

The observed and expected event yields are evaluated in each of the analysis bins. To search for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with WZ-mediated and flavor-democratic slepton-mediated superpartner decays we use the four parametric neural network introduced in Section 9.4.1.2. The neural networks are evaluated at each δm value for which signal events are simulated, and some of these evaluations are shown in Fig. 9.19 to 9.22. The list of evaluated δm values can be found in Fig. 9.33 to 9.36. The observed and expected event yields in the search region bins are shown in Fig. 9.24 to 9.31 for each of the event categories considered. The estimation of the expected background yields in all of these figures follows the recipe of Section 9.5, and the systematic uncertainties are those listed in Section 9.6.

Given the large multiplicity of search region bins, we study the agreement between the SM expectation and the observation in terms of test statistic values. The test statistic values quantify the goodness-of-fit of a background-only fit to the observed data. The expected test statistic distributions under the background-only hypothesis are compared to the observed test statistic values in each event category, and for each parametric neural network evaluation. The test statistic results for fits to the search region bins in each event category are shown in Fig. 9.32. The test statistic distributions for the parametric neural network evaluations in $3\ell A$ events are shown in Fig. 9.33 for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with WZ-mediated superpartner decays and in Fig. 9.34 to 9.36 for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with flavor-democratic slepton-mediated superpartner decays. The test statistic computation follows the procedure in [371]. These figures are a novel way to summarize the compatibility between the observed and expected yields in a search for new physics. They are introduced to enable the simultaneous visualization of the results of all δm evaluations of the parametric neural networks.

recommendation was introduced at a time when the modeling of $\vec{p}_T^{\text{miss}}$ in fast simulation events was much worse than it was at the time this analysis was performed. The effect of this uncertainty on the expected exclusion limits of the analysis is however minute.

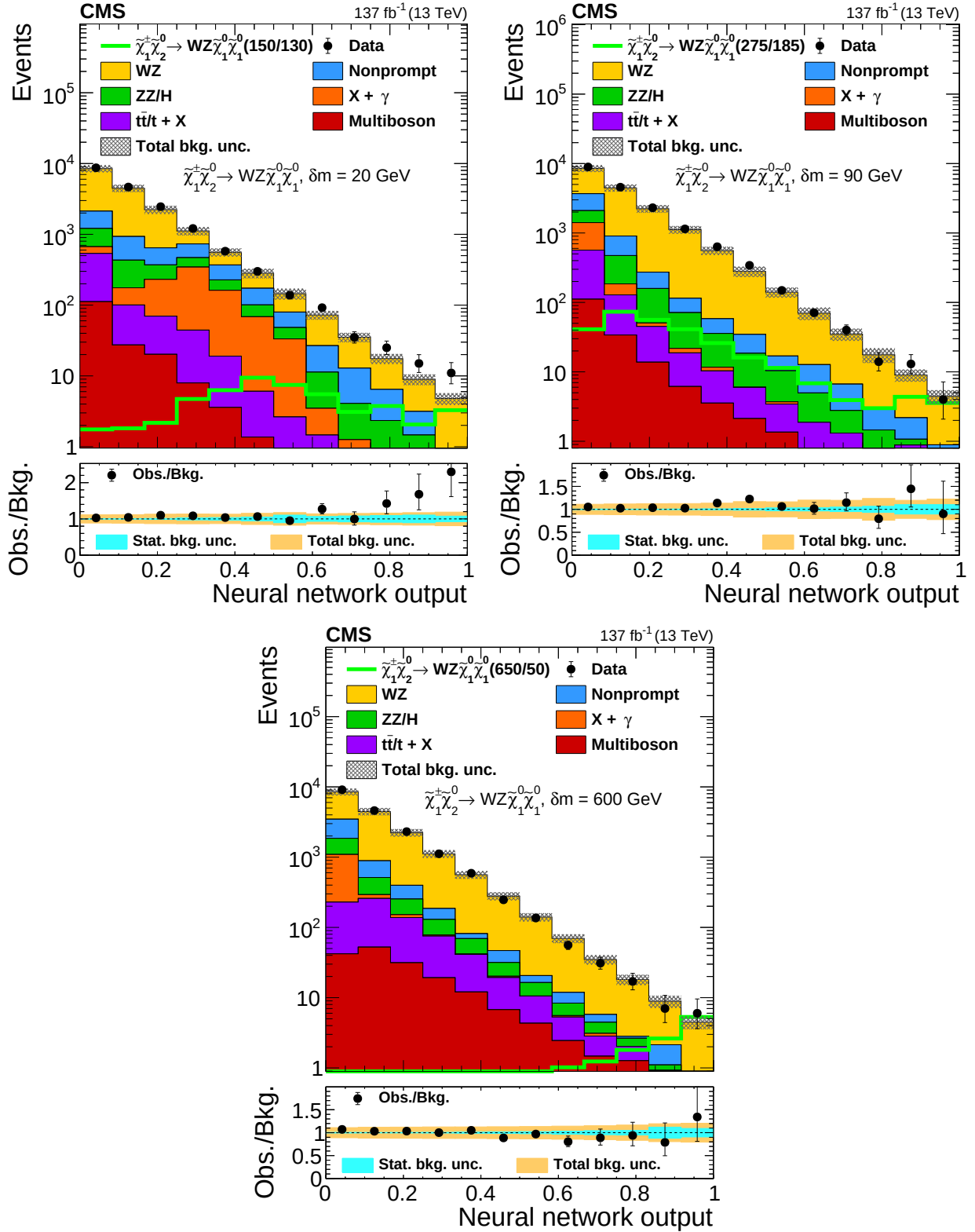


Figure 9.19: Observed and expected number of events as a function of the output of the parametric neural network used to search for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with WZ-mediated superpartner decays. Evaluations of the neural network are shown at $\delta m = 20$ GeV (top left), 90 GeV (top right), and 600 GeV (bottom). The expected yields for signal mass hypotheses close to the expected mass exclusion limits of the analysis are shown in green, and the legend lists the masses of the superpartners in these signal hypotheses as $(m_{\tilde{\chi}_2^0}/m_{\tilde{\chi}_1^0})$ in units of GeV.

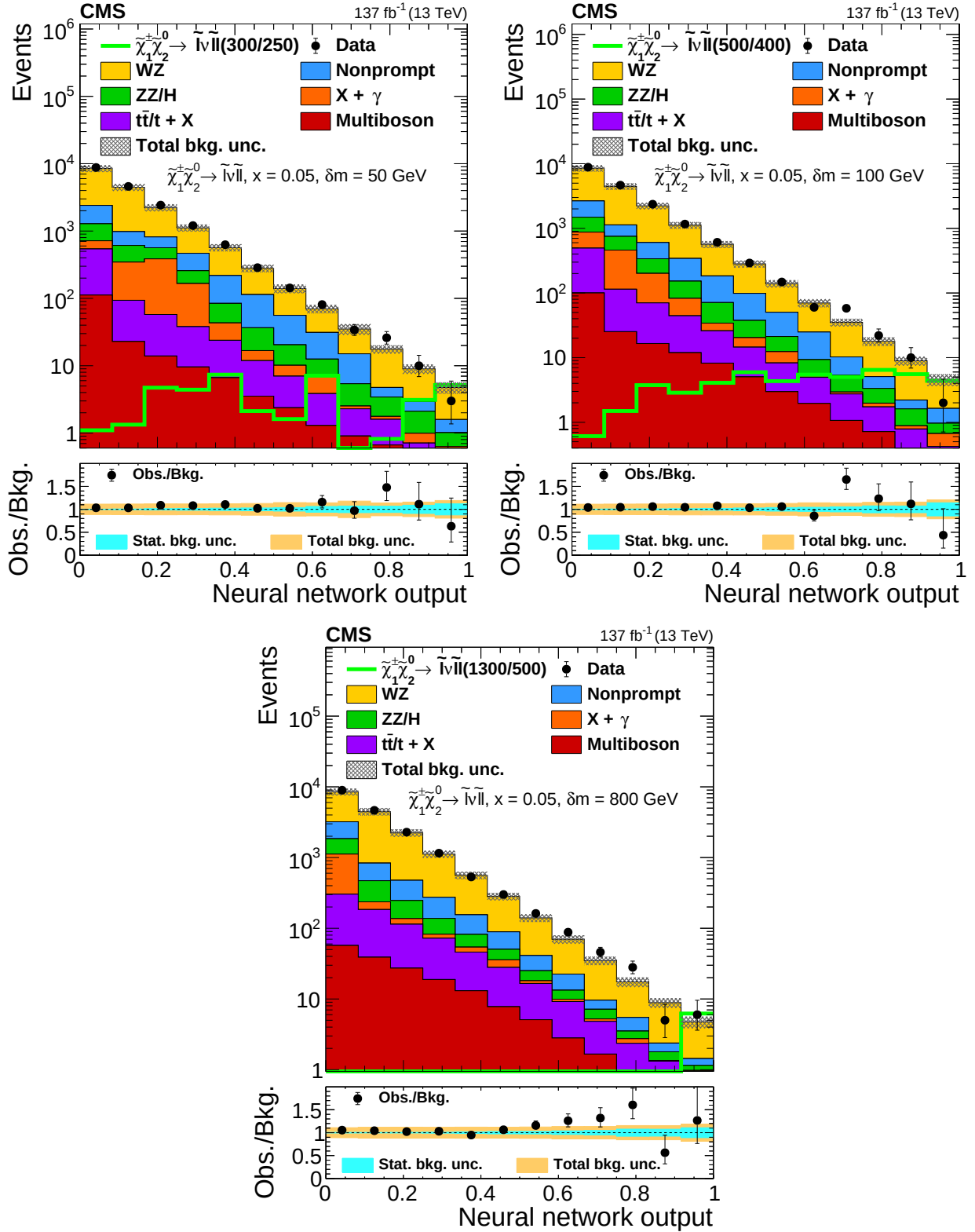


Figure 9.20: Observed and expected number of events as a function of the output of the parametric neural network used to search for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with flavor-democratic slepton-mediated superpartner decays at $x = 0.05$. Evaluations of the neural network are shown at $\delta m = 50$ GeV (top left), 100 GeV (top right), and 800 GeV (bottom). The expected yields for signal mass hypotheses close to the expected mass exclusion limits of the analysis are shown in green, and the legend lists the masses of the superpartners in these signal hypotheses as $(m_{\tilde{\chi}_2^0}/m_{\tilde{\chi}_1^0})$ in units of GeV.

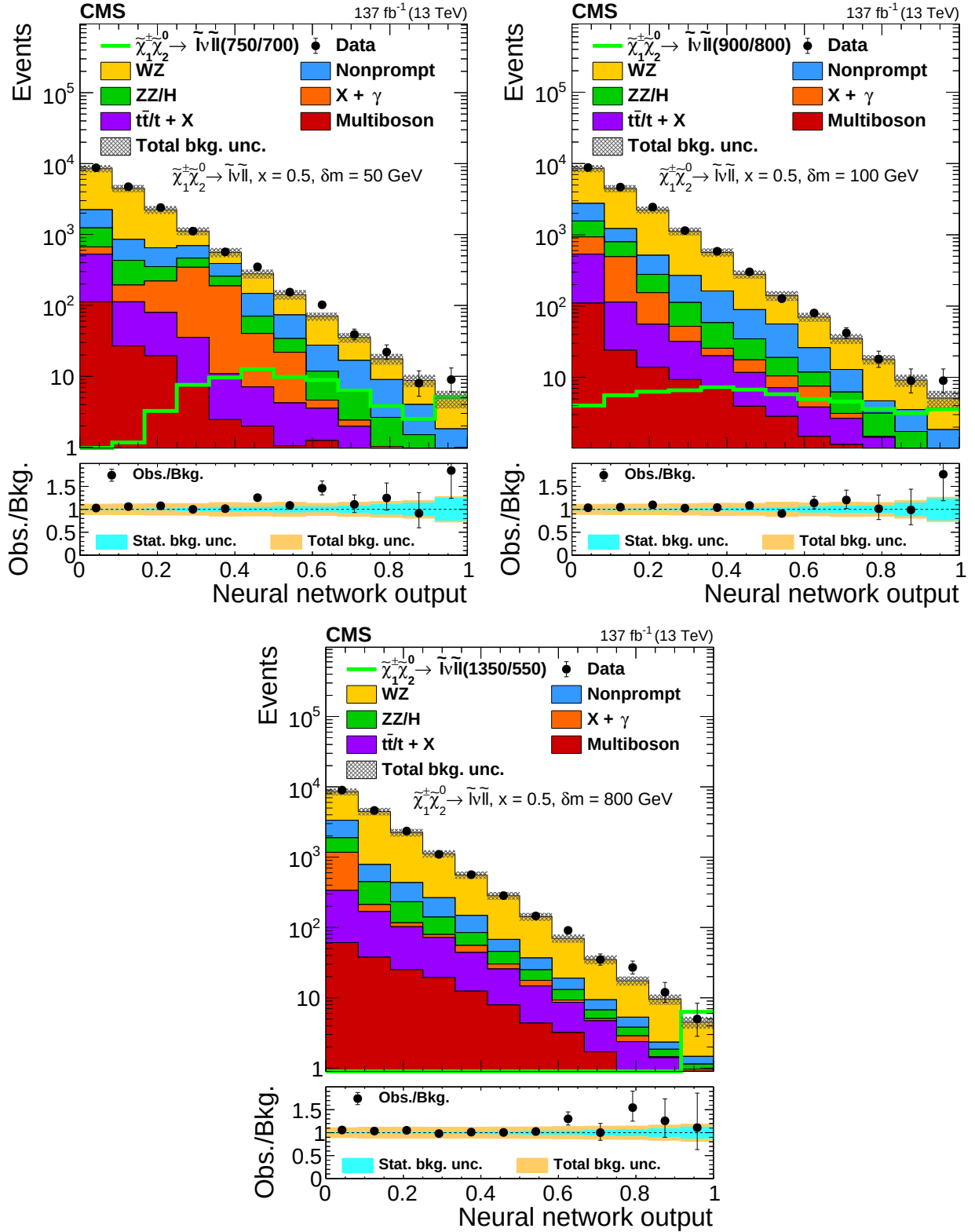


Figure 9.21: Observed and expected number of events as a function of the output of the parametric neural network used to search for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with flavor-democratic slepton-mediated superpartner decays at $x = 0.5$. Evaluations of the neural network are shown at $\delta m = 50$ GeV (top left), 100 GeV (top right), and 800 GeV (bottom). The expected yields for signal mass hypotheses close to the expected mass exclusion limits of the analysis are shown in green, and the legend lists the masses of the superpartners in these signal hypotheses as $(m_{\tilde{\chi}_2^0}/m_{\tilde{\chi}_1^0})$ in units of GeV.

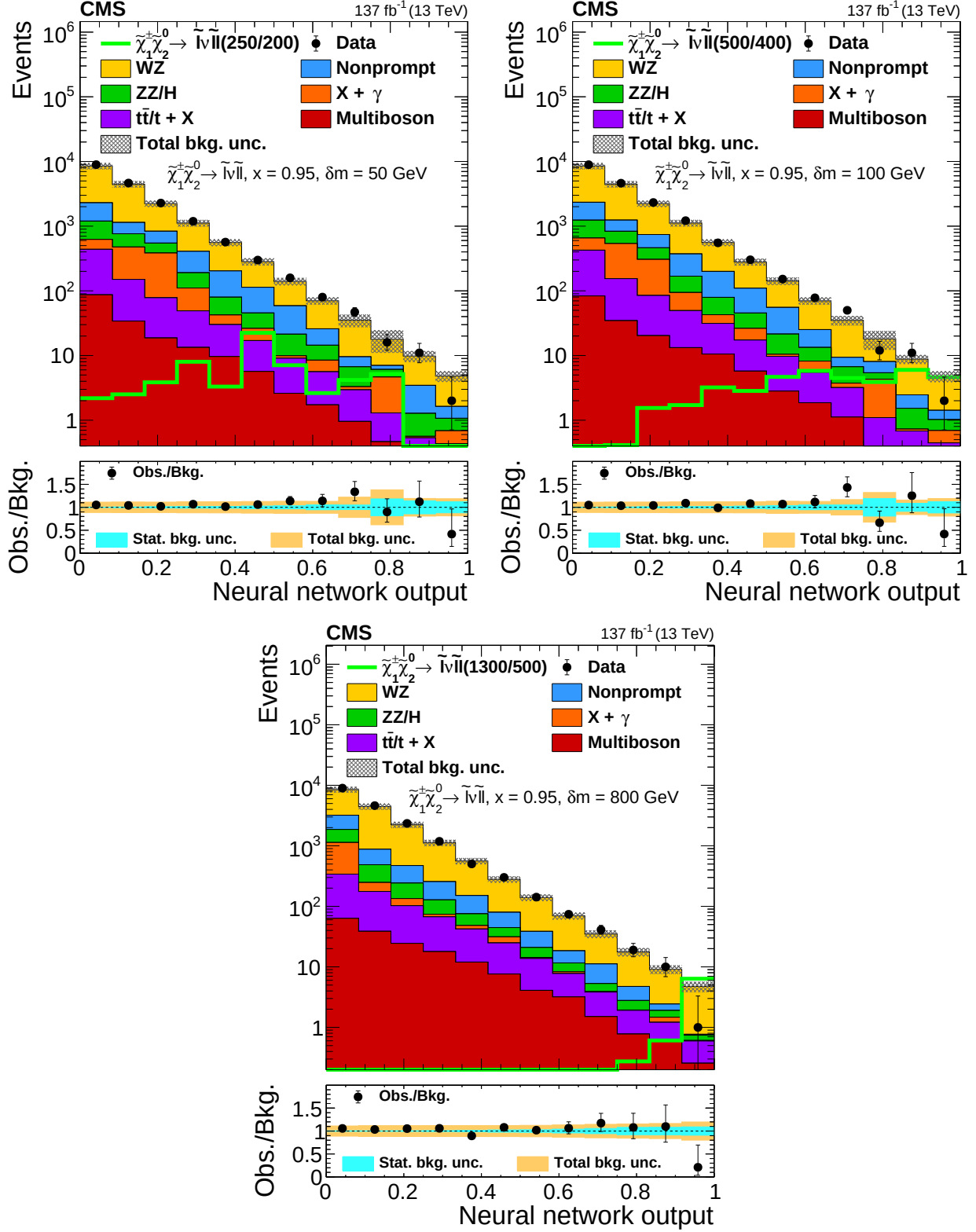


Figure 9.22: Observed and expected number of events as a function of the output of the parametric neural network used to search for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with flavor-democratic slepton-mediated superpartner decays at $x = 0.95$. Evaluations of the neural network are shown at $\delta m = 50$ GeV (top left), 100 GeV (top right), and 800 GeV (bottom). The expected yields for signal mass hypotheses close to the expected mass exclusion limits of the analysis are shown in green, and the legend lists the masses the superpartners in these signal hypotheses as $(m_{\tilde{\chi}_2^0}/m_{\tilde{\chi}_1^0})$ in units of GeV.

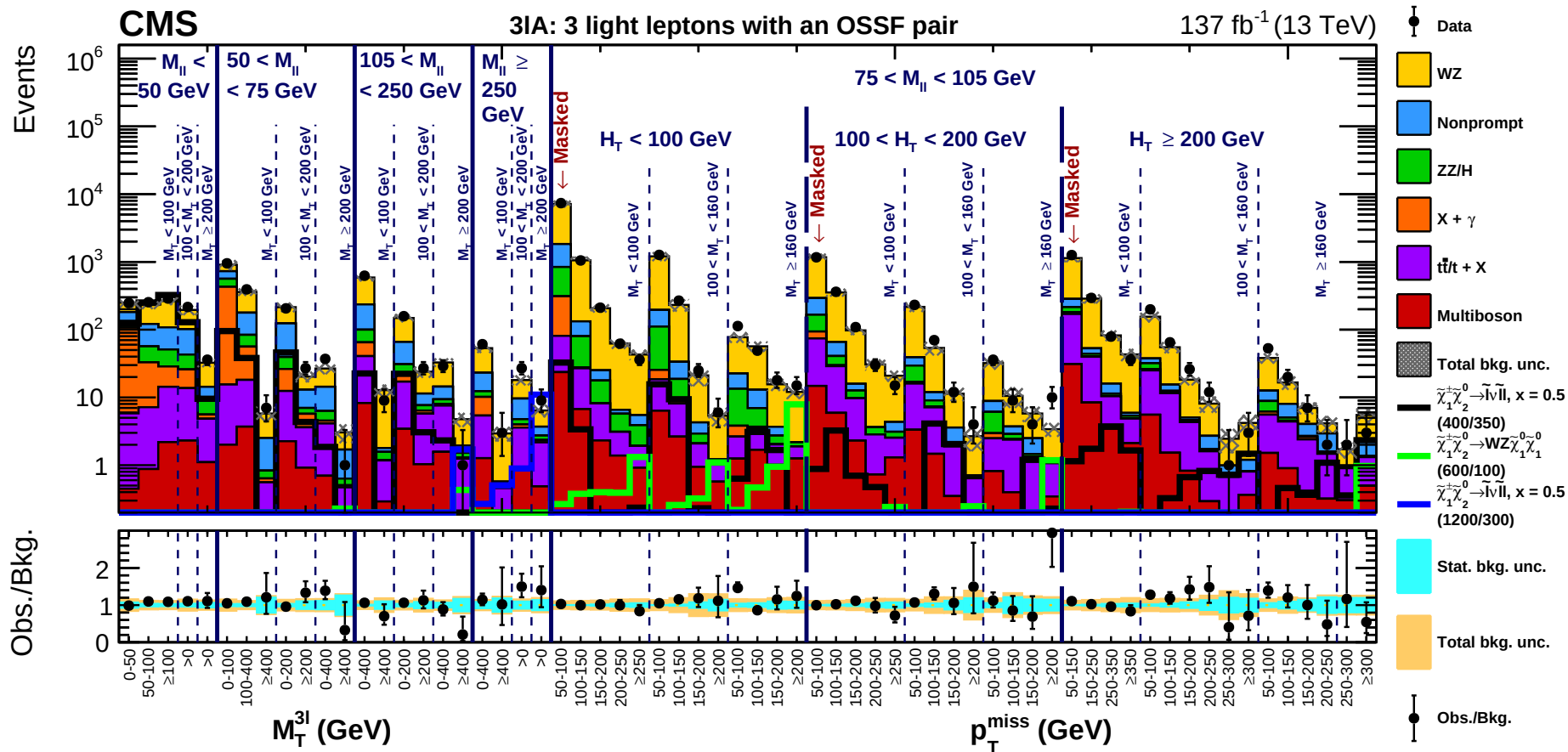


Figure 9.23: Observed and expected event yields in the search region bins for events with three light leptons out of which at least two form an OSSF pair (3lA). Three bins are labeled “Masked” and are excluded from the analysis fits to the 3lA search region distribution because they overlap with the WZ control region. Expected yields for three hypotheses of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production are shown. A compressed (noncompressed) model of flavor-democratic slepton-mediated superpartner decays with $\delta m = 50$ (800) GeV at $x = 0.5$ is drawn in black (blue). A noncompressed model of WZ-mediated superpartner decays with $\delta m = 500$ GeV is shown in green. The masses of the superpartners in the shown signal hypotheses are given in the legend as $(m_{\tilde{\chi}_2^0}/m_{\tilde{\chi}_1^0})$ in units of GeV.

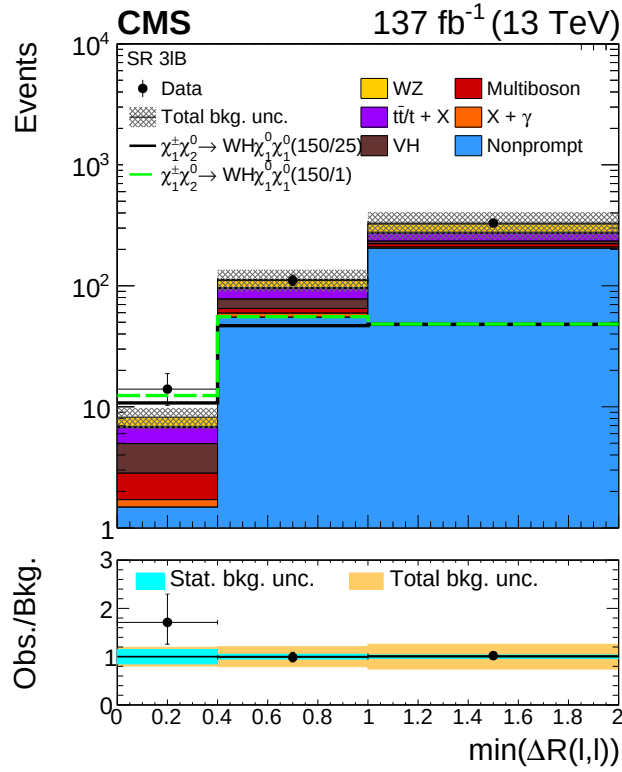


Figure 9.25: Observed and expected event yields in the search region bins for events with three light leptons, none of which form an OSSF pair (3ℓB). The expected yields for an hypothesis of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with WH-mediated superpartner decays at $\delta m = 125$ (150) GeV are drawn as a solid black (dashed green) line. The masses of the superpartners in the shown signal hypotheses are given in the legend as $(m_{\tilde{\chi}_2^0}/m_{\tilde{\chi}_1^0})$ in units of GeV.

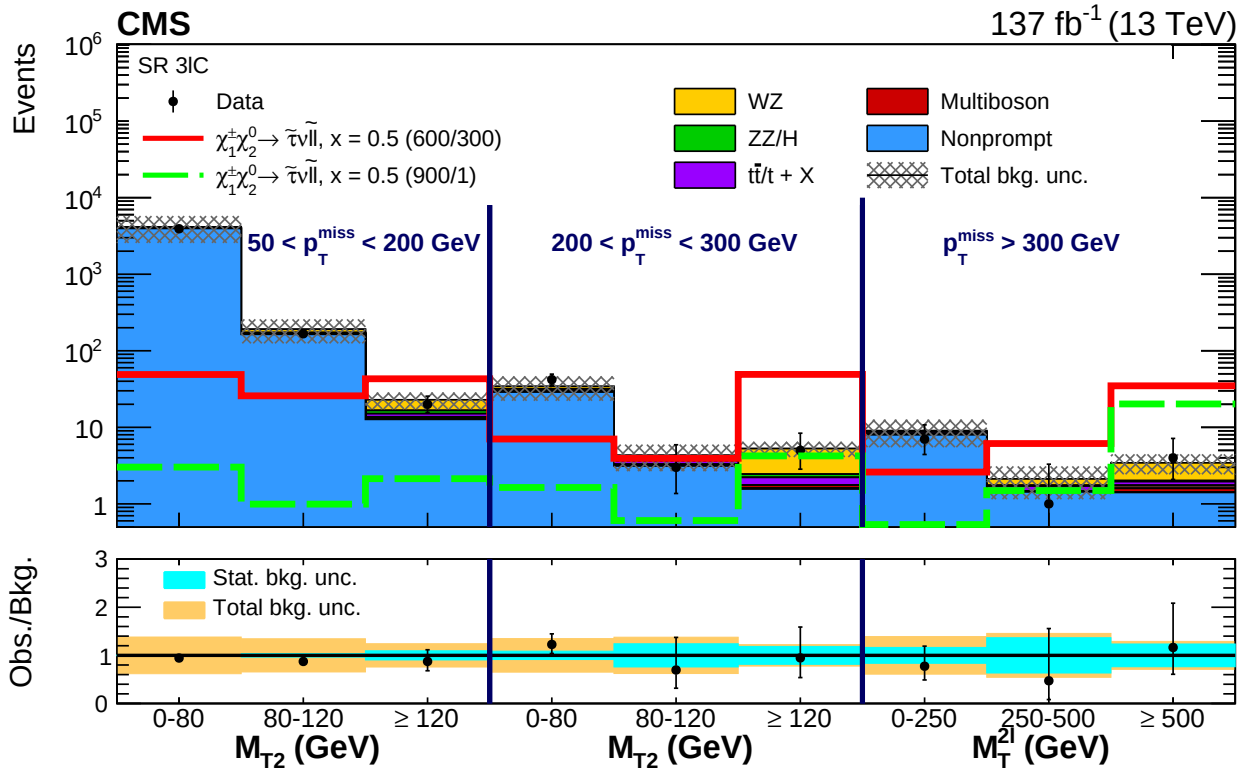


Figure 9.26: Observed and expected event yields in the search region bins for events with a $\mu^\pm\mu^\mp$ or $e^\pm e^\mp$ pair and a single τ_h candidate (3lC). The expected yields for an hypothesis of $\tilde{\chi}_1^\pm\tilde{\chi}_2^0$ production with τ -enriched slepton-mediated superpartner decays with $\delta m = 300$ (900) GeV at $x = 0.5$ are drawn as a solid red (dashed green) line. The masses of the superpartners in the shown signal hypotheses are given in the legend as $(m_{\tilde{\chi}_2^0}/m_{\tilde{\chi}_1^0})$ in units of GeV.

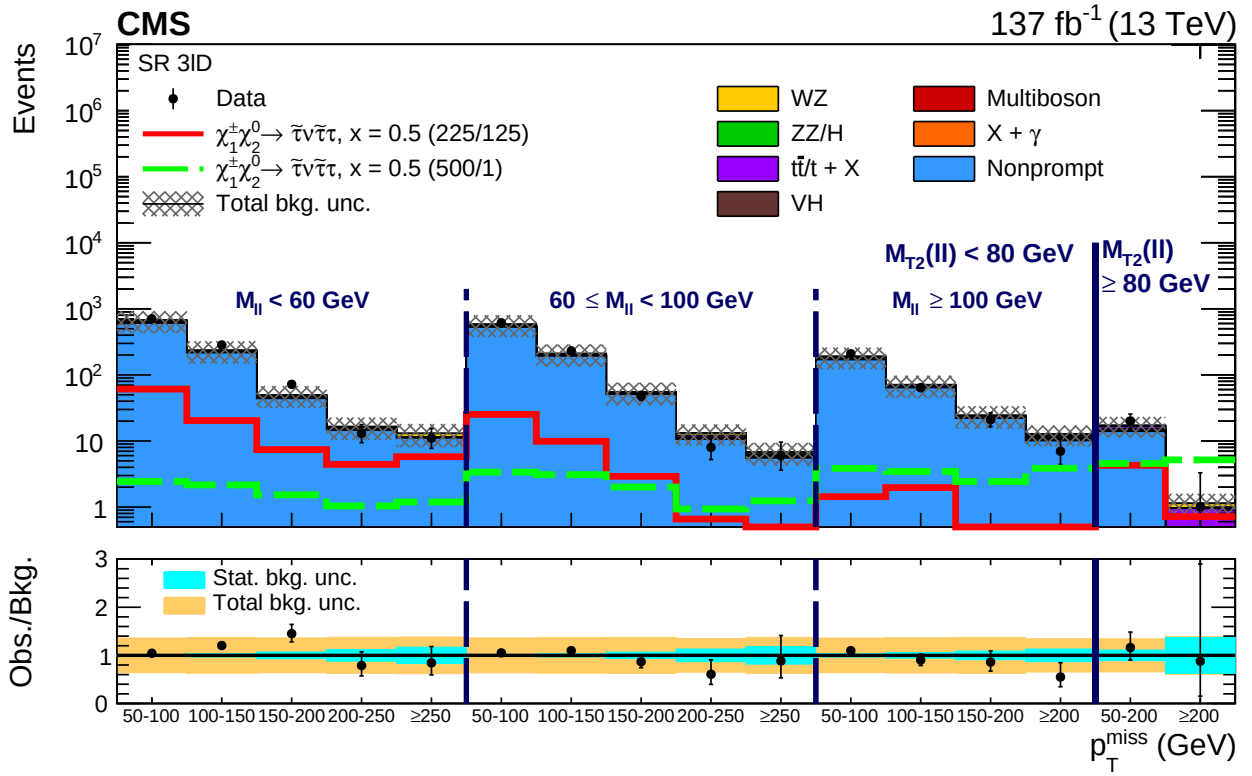


Figure 9.27: Observed and expected event yields in the search region bins for events with an $e^\pm \mu^\mp$ pair and a single τ_h candidate (3ℓD). The expected yields for an hypothesis of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with τ -dominated slepton-mediated superpartner decays with $\delta m = 100$ (500) GeV at $x = 0.5$ are drawn as a solid red (dashed green) line. The masses of the superpartners in the shown signal hypotheses are given in the legend as $(m_{\tilde{\chi}_2^0}/m_{\tilde{\chi}_1^0})$ in units of GeV.

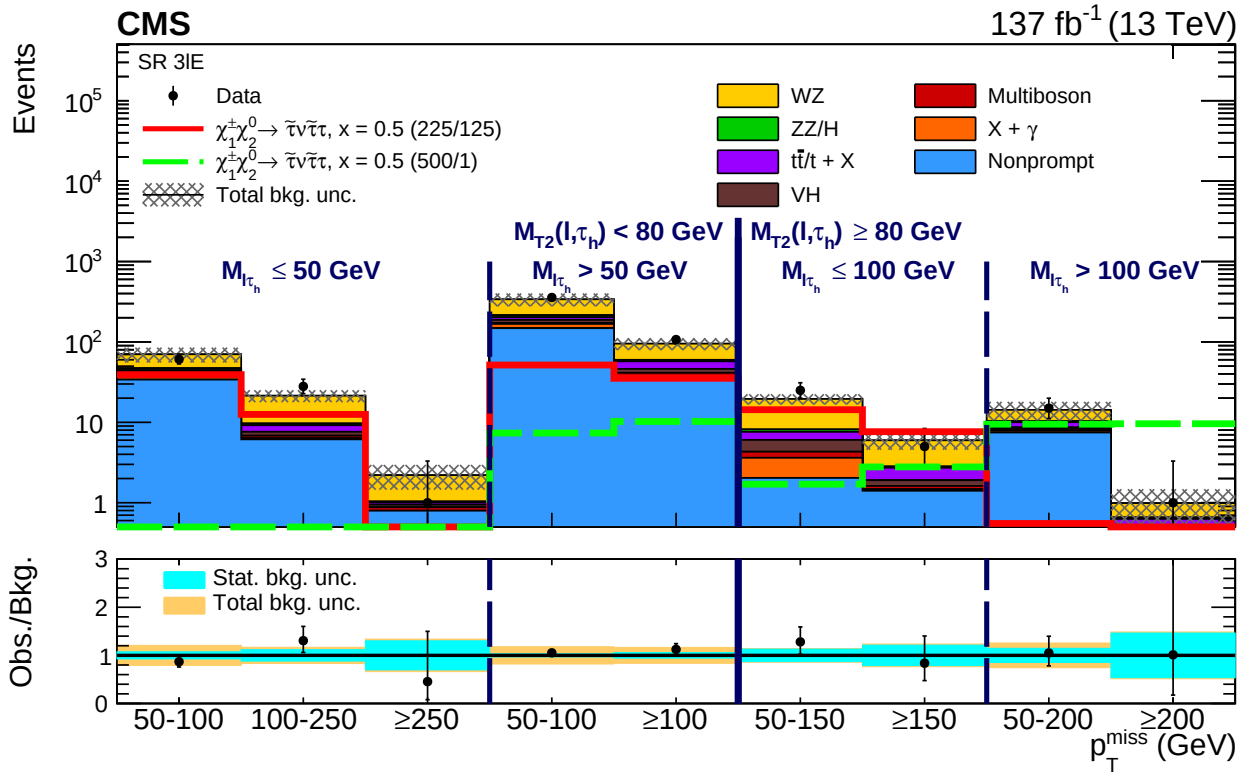


Figure 9.28: Observed and expected event yields in the search region bins for events with a same-sign light lepton pair and a single τ_h candidate (3 ℓ E). The expected yields for an hypothesis of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with τ -dominated slepton-mediated superpartner decays with $\delta m = 100$ (500) GeV at $x = 0.5$ are drawn as a solid red (dashed green) line. The masses of the superpartners in the shown signal hypotheses are given in the legend as $(m_{\tilde{\chi}_2^0}/m_{\tilde{\chi}_1^0})$ in units of GeV.

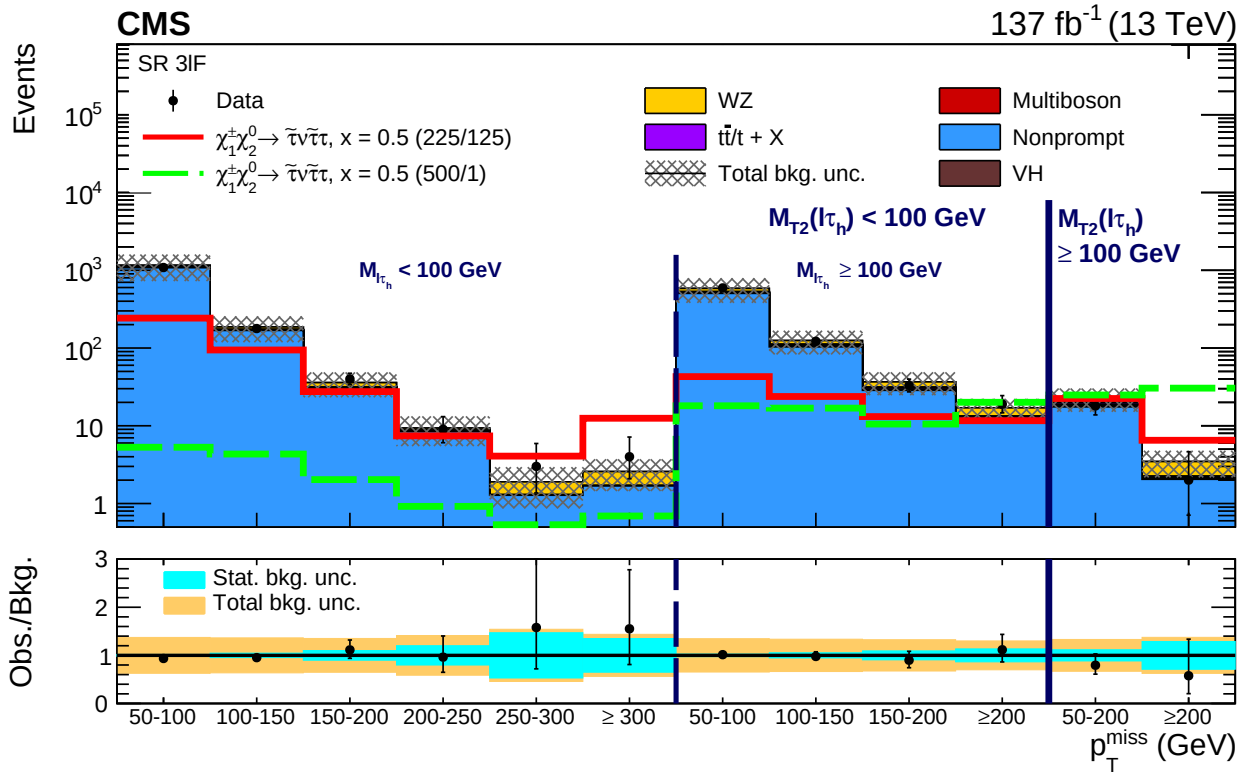


Figure 9.29: Observed and expected event yields in the search region bins for events with two τ_h candidates and a single light lepton ($3\ell F$). The expected yields for an hypothesis of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with τ -dominated slepton-mediated superpartner decays with $\delta m = 100$ (500) GeV at $x = 0.5$ are drawn as a solid red (dashed green) line. The masses of the superpartners in the shown signal hypotheses are given in the legend as $(m_{\tilde{\chi}_2^0}/m_{\tilde{\chi}_1^0})$ in units of GeV.

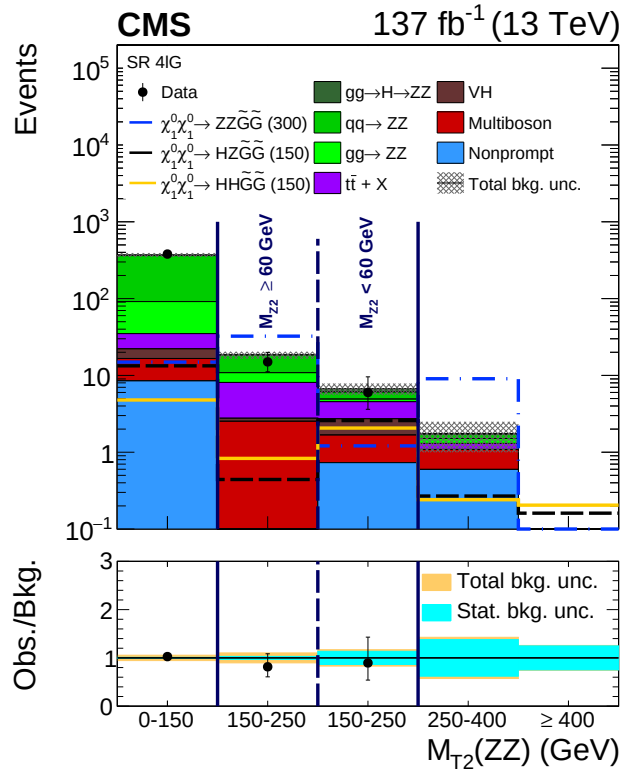


Figure 9.30: Observed and expected event yields in the search region bins for events with four light leptons forming two exclusive OSSF pairs ($4\ell\text{G}$). The expected yields for three hypotheses of effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production are shown. A model of ZZ-mediated superpartner decays with $m_{\tilde{\chi}_1^0} = 300 \text{ GeV}$ is drawn in blue. A model of HZ- (HH)-mediated superpartner decays with $m_{\tilde{\chi}_1^0} = 150 \text{ GeV}$ is drawn in black (orange). The masses of the mass-degenerate charginos and neutralinos in the shown signal hypotheses are given in the legend as $(m_{\tilde{\chi}_1^0})$ in units of GeV.

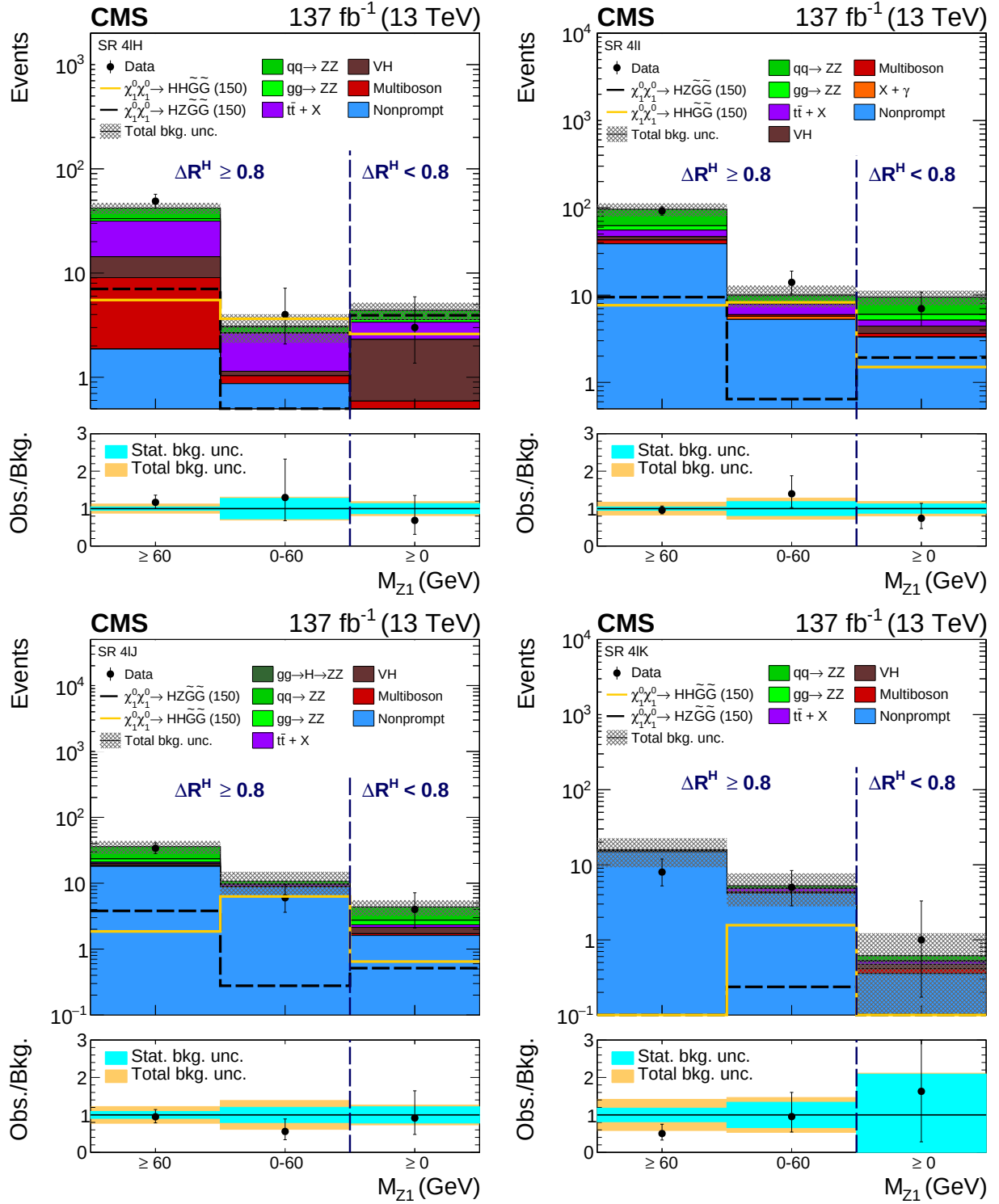


Figure 9.31: Observed and expected event yields in the search region bins for events with four light leptons forming one or no exclusive OSSF pairs (4ℓH, top left), three light leptons and a single τ_h candidate (4ℓI, top right), two light leptons and two τ_h candidates forming two OSSF pairs (4ℓJ, bottom left), and two light leptons and two τ_h candidates forming one or no OSSF pairs (4ℓK, bottom right). The expected yields for an hypothesis of effective $\tilde{\chi}_1^0\tilde{\chi}_1^0$ production with HH- (HZ)-mediated superpartner decays at $\delta m = 150$ GeV are drawn for each event category as a solid orange (black dashed) line. The masses of the mass-degenerate charginos and neutralinos in the shown signal hypotheses are given in the legend as $(m_{\tilde{\chi}_1^0})$ in units of GeV.

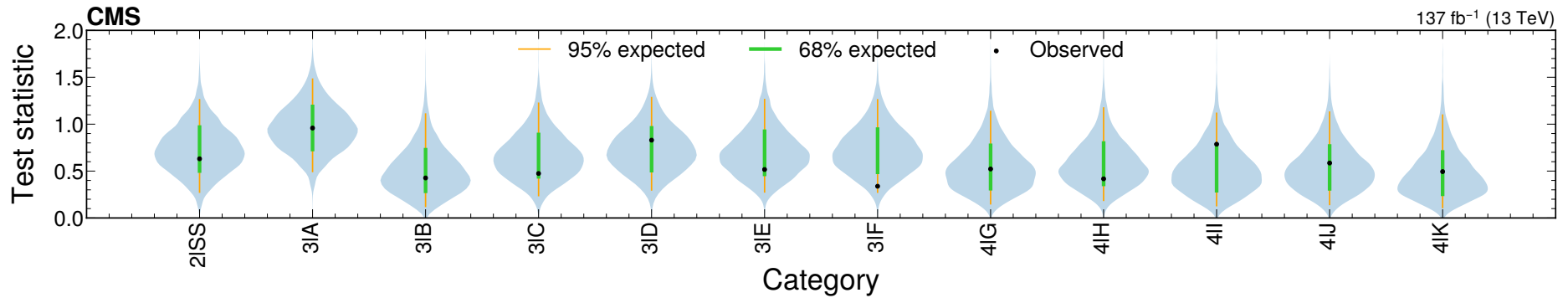


Figure 9.32: Observed test statistic values compared to the expected test statistic distributions for background-only fits to the search region bins in each event category. The computation of the test statistic values follows the procedure in [371]. The symmetrized expected probability density functions of the test statistics are drawn as gray shaded areas. The green (orange) bars represent the 68 (95)% expected test statistic ranges, and the black dots indicate the observed test statistic values.

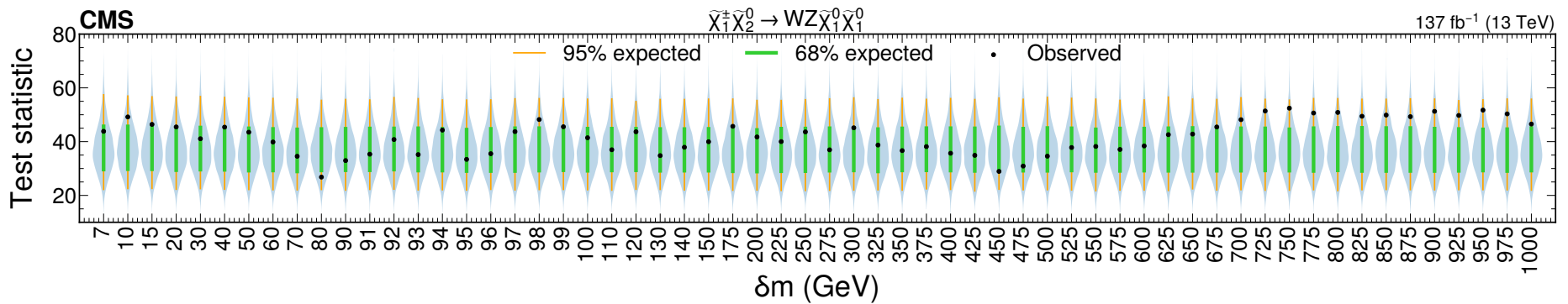


Figure 9.33: Observed test statistic values compared to the expected test statistic distributions for background-only fits to the output distributions of the parametric neural network used to search for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with WZ-mediated superpartner decays in $3\ell A$ events. The computation of the test statistic values follows the procedure in [371]. Test statistic distributions are shown separately for each δm evaluation of the neural network used in the analysis. The symmetrized expected probability density functions of the test statistics are drawn as gray shaded areas. The green (orange) bars represent the 68 (95)% expected test statistic ranges, and the black dots indicate the observed test statistic values.

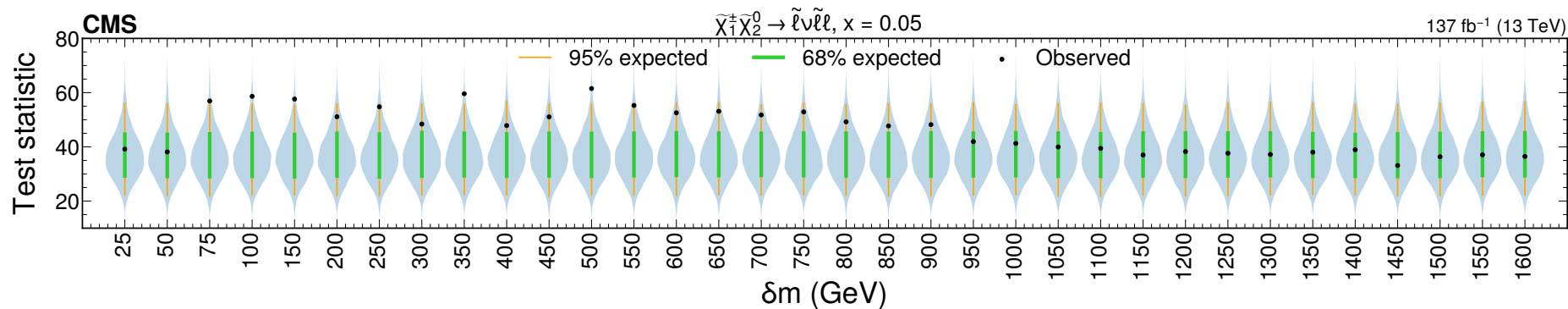


Figure 9.34: Observed test statistic values compared to the expected test statistic distributions for background-only fits to the output distributions of the parametric neural network used to search for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with slepton-mediated superpartner decays at $x = 0.05$ in $3\ell A$ events. The figure contents are described in the caption of Fig. 9.33.

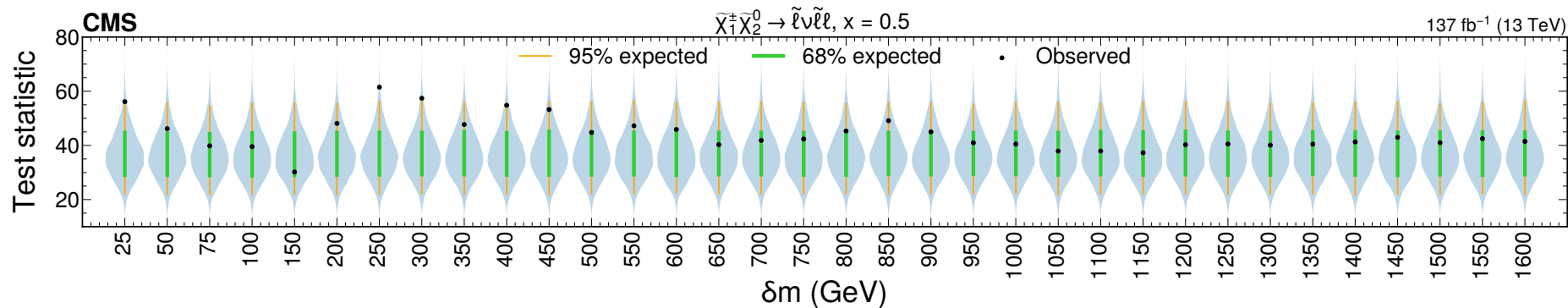


Figure 9.35: Observed test statistic values compared to the expected test statistic distributions for background-only fits to the output distributions of the parametric neural network used to search for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with slepton-mediated superpartner decays at $x = 0.5$ in $3\ell A$ events. The figure contents are described in the caption of Fig. 9.33.

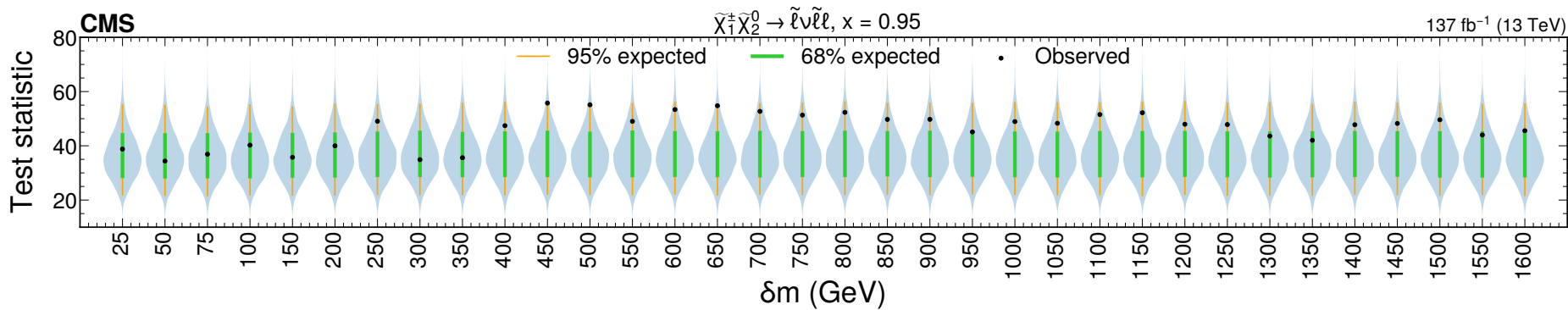


Figure 9.36: Observed test statistic values compared to the expected test statistic distributions for background-only fits to the output distributions of the parametric neural network used to search for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with slepton-mediated superpartner decays at $x = 0.95$ in $3\ell A$ events. The figure contents are described in the caption of Fig. 9.33.

The observed test statistic values match the expected distributions in the search regions as well as in the majority of the parametric neural network evaluations. For $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with slepton-mediated superpartner decays there are a number of evaluations where the observed test statistic value falls outside of the 95% expected test statistic range. In none of these evaluations a significant excess of events over the SM expectation is seen, but rather small mismatches between the predicted neural network output shape and the distribution observed in collision data that are not signal-like. Given the large number of evaluations some deviations are expected. Each neural network evaluation is neither fully uncorrelated from the others nor fully correlated, and evaluations of the data at close by δm values are more correlated than those at far away δm values.

Table 9.15: Summary of the event categories used for the interpretation of the results in terms of different signal models, and references to the associated figures summarizing the observed and expected 95% confidence level upper limits on the superpartner masses and production cross sections.

Signal model	Categories included in fits	Figures
$\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production, slepton-mediated decays		
Flavor-democratic	2 ℓ SS, 3 ℓ A (search regions and neural networks fit separately in 3 ℓ A)	9.37–9.39
τ -enriched	3 ℓ A–3 ℓ F	9.40–9.42
τ -dominated	3 ℓ B–3 ℓ F	9.43–9.43
$\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production, WZ-mediated decays	2 ℓ SS, 3 ℓ A (search regions and neural network fit separately in 3 ℓ A)	9.46
$\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production, WH-mediated decays	2 ℓ SS, 3 ℓ A–3 ℓ F	9.47
Effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production	3 ℓ A–4 ℓ K	9.48–9.50

9.8 Interpretations

No significant excess of events over the SM expectation is observed in any of the event categories. We proceed to use the observed event yields to put constraints on $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production and effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production. Each of the search categories provides sensitivity to one or more of the models introduced in Section 9.2. The interpretation of the results in terms of a particular model is done by means of a maximum likelihood fit to the analysis bins in the search categories providing sensitivity to this model. An overview of which event categories are included in the fit for each signal model, as well as a reference to the figures showing the exclusion limits for each signal model are given in Table 9.15.

For each signal mass hypothesis a 95% confidence level upper limit on the cross section is computed with CL_s method [287, 288] following a similar approach to that described in Section 7.8 where upper limits on sterile neutrino production were determined. The asymptotic approximation of the distribution of the profile likelihood test statistic [291, 292] is used, and the systematic uncertainties introduced in Section 9.6 are included as nuisance parameters in the likelihood function. The template morphing technique [289] with Gaussian probability density functions is employed to account for uncertainty sources that affect the shape of the expected signal or background distributions. To model uncertainties that only affect the normalization of processes, nuisance parameters with log-normal probability density functions are used. The effects of the statistical uncertainties in the simulated samples and the sideband data used to predict the sign misidentification and nonprompt lepton backgrounds, are included in the fit with the Barlow-Beeston method [290].

The expected cross sections for superpartner production decrease rapidly as a function of the masses of the produced superpartners, as illustrated in Fig. 9.1. By comparing the upper limits set on the production cross section at each combination of NLSP and LSP masses we can exclude the range of superpartner masses for which the excluded cross section value is smaller than the expected one. Note here that the expected signal production cross sections depend only on the NLSP mass while the exclusion limits depend on both the NLSP and LSP masses because the latter affects the kinematics of the final state particles. The theoretical reference cross sections used in the calculations are those described in Section 9.2.4, computed at NLO precision in perturbative QCD with a next-to-leading logarithmic precision in the resummation of soft gluon emission.

The exclusion limits for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with flavor-democratic slepton-mediated superpartner decays are shown in Fig. 9.37 to 9.39. The sensitivity of the analysis to this scenario is dominated by 3 ℓ A events. For compressed signal models the 2 ℓ SS search regions significantly boost the analysis reach by recovering events in which one or more leptons fails to be selected. For signal models with $x = 0.05$ or $x = 0.95$, and δm values below 30 GeV, 2 ℓ SS events form the most important contribution to the

search results. The exclusion limits for models with flavor-democratic slepton-mediated superpartner decays are evaluated twice: Once by fitting the neural network output distributions at each δm value, and once with the search region approach. The search region and neural network analyses use the same set of events, so they can not be used simultaneously.

The neural network analysis provides more stringent exclusion limits than the search regions for every signal mass hypothesis. The difference between the two approaches is largest when $x = 0.95$, i.e. $m_{\tilde{\ell}}$ is close to $m_{\tilde{\chi}_2^0}$. These are most compressed signal hypotheses. The expected neural network exclusion limits are higher by up to 250 GeV in $m_{\tilde{\chi}_2^0}$ and 150 GeV in $m_{\tilde{\chi}_1^0}$ when compared to the 3 ℓ A search region results. This corresponds to excluded cross sections that are two times lower, or in other words a relative analysis improvement of 100%. Here it must be noted that the new search regions themselves already improve the results by about 100% in the same data set when compared to the search regions of the previous analysis iteration. In addition, the new search regions are inspired by what the neural networks learned. Compared to the old analysis strategy the neural networks thus reduce the excluded cross sections by up to a factor four in the same data set. The full Run 2 collision data set is about four times larger than the 2016 data set used previously, corresponding to an expected analysis improvement by approximately a factor two. One can then conclude that the new neural network analysis run on the 2016 data set would be roughly twice as powerful as the old analysis run on the full Run 2 data set. When the slepton mass gets smaller ($x = 0.5$ and $x = 0.05$) signal events are easier to distinguish from the background because of the harder lepton p_T spectra, and the improvements given by the neural network analysis become smaller. When compared to the new search region analysis the neural networks extend the expected exclusion limits by up to 100 GeV in $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$, a relative improvement of 50% in the excluded cross sections. The improvements in the excluded cross sections at very low δm values are often greater than 100% at $x = 0.5$, and smaller than 50% for models with $x = 0.05$ and $x = 0.95$. All in all, NLSP and LSP values up to respectively 1450 GeV and 1000 GeV can be excluded by the neural network analysis under certain model hypotheses. These are by far the highest direct exclusion limits on the masses of electroweak superpartners achieved in any analysis, as shown in Fig. 9.3, and in any experiment. Compared to the previous analysis iteration the exclusion limits are extended by up to 400 GeV in the NLSP mass, corresponding to production cross sections that are a factor 9 lower.

The sensitivity to models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with τ -enriched and τ -dominated slepton-mediated superpartner decays is driven by search categories 3 ℓ B to 3 ℓ F. For τ -enriched models 3 ℓ A and 2 ℓ SS events increase the reach of the search so these categories are also evaluated. For τ -dominated scenarios these categories are not included in the fits. The exclusion limits for τ -enriched (τ -dominated) scenarios are shown in Fig. 9.40 to 9.42 (Fig. 9.43 to 9.45). In the τ -enriched scenario $m_{\tilde{\chi}_2^0}$ ($m_{\tilde{\chi}_1^0}$) values up to 1150 (1000) GeV are excluded. Compared to the previous analysis, the limits are extended by 150 GeV in the NLSP mass, equating excluded cross sections that are two times lower. For τ -dominated scenarios the exclusion limits reach up to $m_{\tilde{\chi}_2^0}$ ($m_{\tilde{\chi}_1^0}$) values of 970 (450) GeV. The analysis is notably less sensitive to compressed τ -dominated decay scenarios because of the large p_T threshold of 20 GeV imposed in the τ_h candidate selection. In the previous analysis iteration relatively little effort was spent on the optimization of the search for τ -dominated scenarios, so the improvements from the updated search region strategy and τ_h selection criteria are large. The excluded mass range is increased by 350 GeV in $m_{\tilde{\chi}_2^0}$, equivalent to excluded cross sections that are a factor 10 lower, and 200 GeV in $m_{\tilde{\chi}_1^0}$.

The reach of the search for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with WZ-mediated superpartner decays is largely determined by 3 ℓ A events. The 2 ℓ SS event category adds a bit of sensitivity, mostly at low δm values. As for the models with flavor-democratic slepton-mediated superpartner decays, the computation of the excluded cross section for each mass point is performed twice: Once with the parametric neural network analysis of 3 ℓ A and once with the search region approach. The exclusion limits for this decay scenario are shown in Fig. 9.46. The parametric neural network has a greater sensitivity than the search region bins for all signal mass hypotheses, and results in excluded $m_{\tilde{\chi}_2^0}$ ($m_{\tilde{\chi}_1^0}$) values that are up to 130 (50) GeV larger. The relative improvements in the excluded cross sections that result from using the neural network range from 40 to 30% at medium to high δm values,

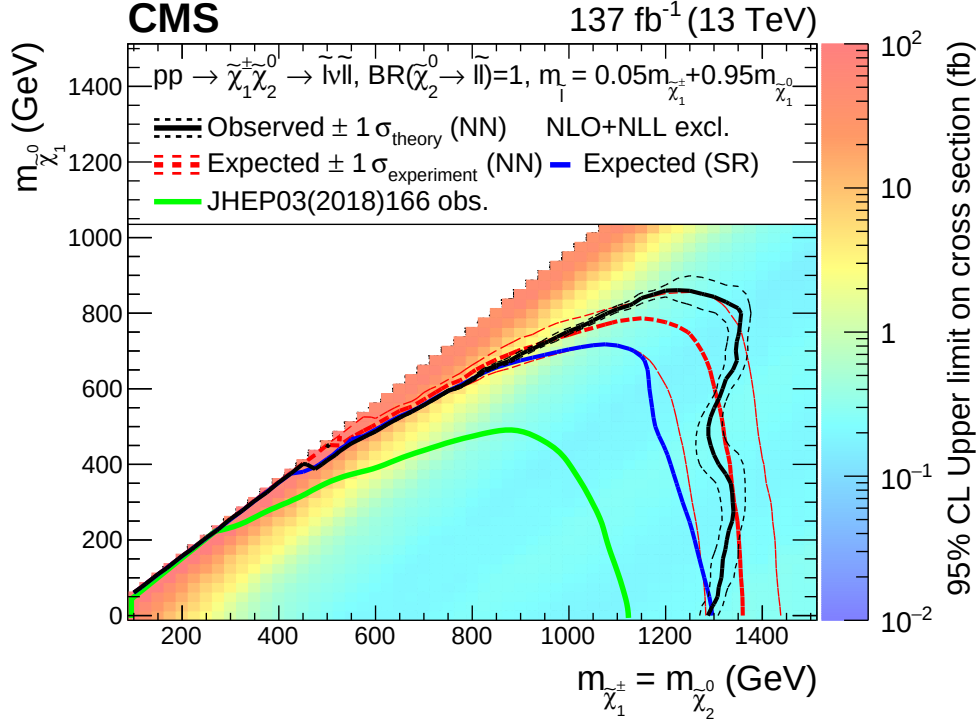


Figure 9.37: Interpretation of the search results for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with flavor-democratic slepton-mediated superpartner decays at $x = 0.05$. The colored shading, the red, and the black lines in the figure are determined with the parametric neural network analysis of $3\ell A$ events. The colors of the shading represent the 95% confidence level upper limits on the $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production cross sections, where the theoretical reference cross sections are those described in Section 9.2.4. The lines superimposed on the shaded area mark superpartner masses below which the excluded cross sections are lower than the predicted ones. This means that $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production is excluded in scenarios below these lines. The thick and thin red dashed lines respectively show the expected exclusion limits in the absence of signal, and the one standard deviation bounds from the experimental uncertainties in the signal and background predictions as well as the theoretical uncertainties in the background prediction. The thick and thin solid black lines respectively delineate the observed exclusion limit, and the one standard deviation bounds due to the uncertainty in the signal production cross sections. The solid blue line represents the expected exclusion curve obtained by using the search region based analysis in $3\ell A$ events. The solid green curve contains the results from the previous iteration of this analysis that analyzed 35.9 fb^{-1} of 13 TeV pp collision data [5].

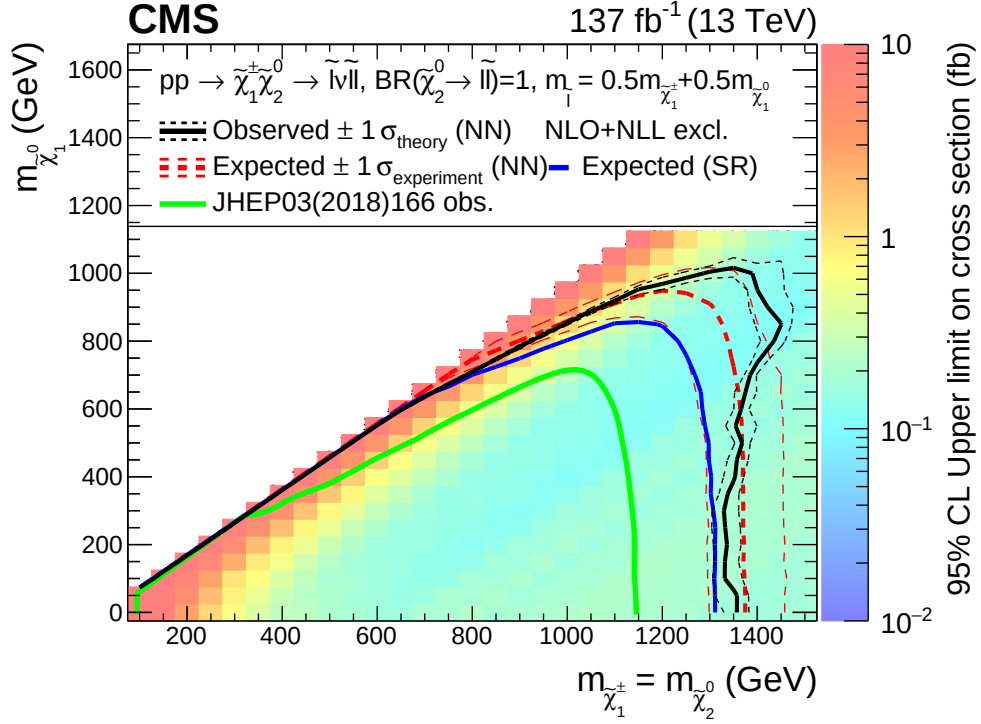


Figure 9.38: Interpretation of the search results for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with flavor-democratic slepton-mediated superpartner decays at $x = 0.5$. The figure contents are explained in the caption of Fig. 9.37.

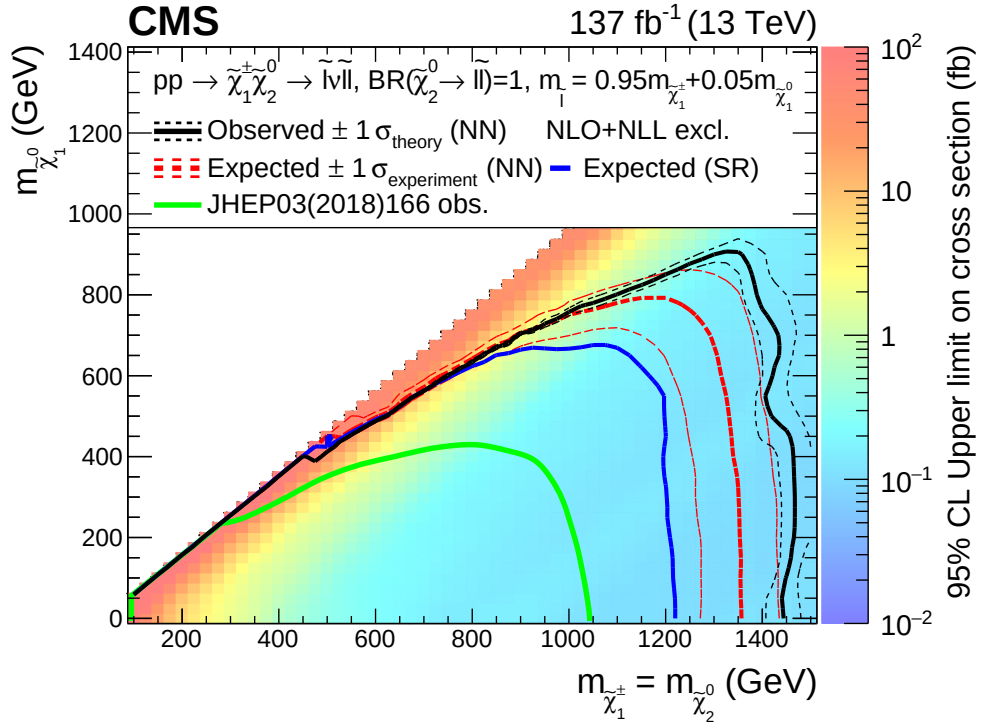


Figure 9.39: Interpretation of the search results for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with flavor-democratic slepton-mediated superpartner decays at $x = 0.95$. The figure contents are explained in the caption of Fig. 9.37.

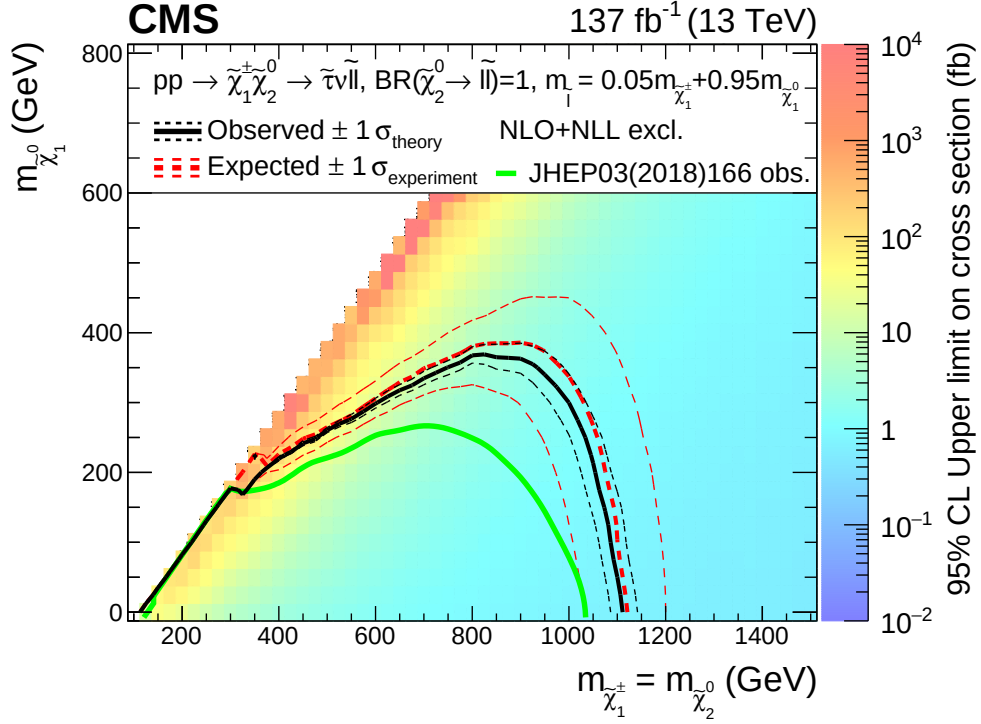


Figure 9.40: Interpretation of the search results for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with τ -enriched slepton-mediated superpartner decays at $x = 0.05$. The figure contents are explained in the caption of Fig. 9.37, with the difference that only a search region based analysis is used for this model.

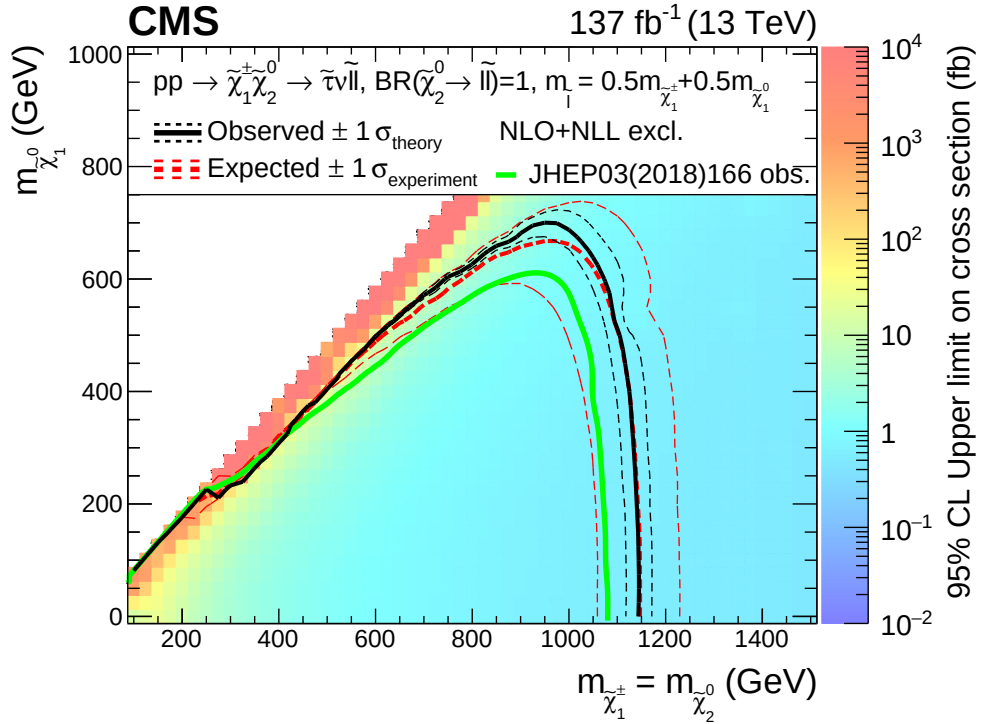


Figure 9.41: Interpretation of the search results for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with τ -enriched slepton-mediated superpartner decays at $x = 0.5$. The figure contents are explained in the caption of Fig. 9.37, with the difference that only a search region based analysis is used for this model.

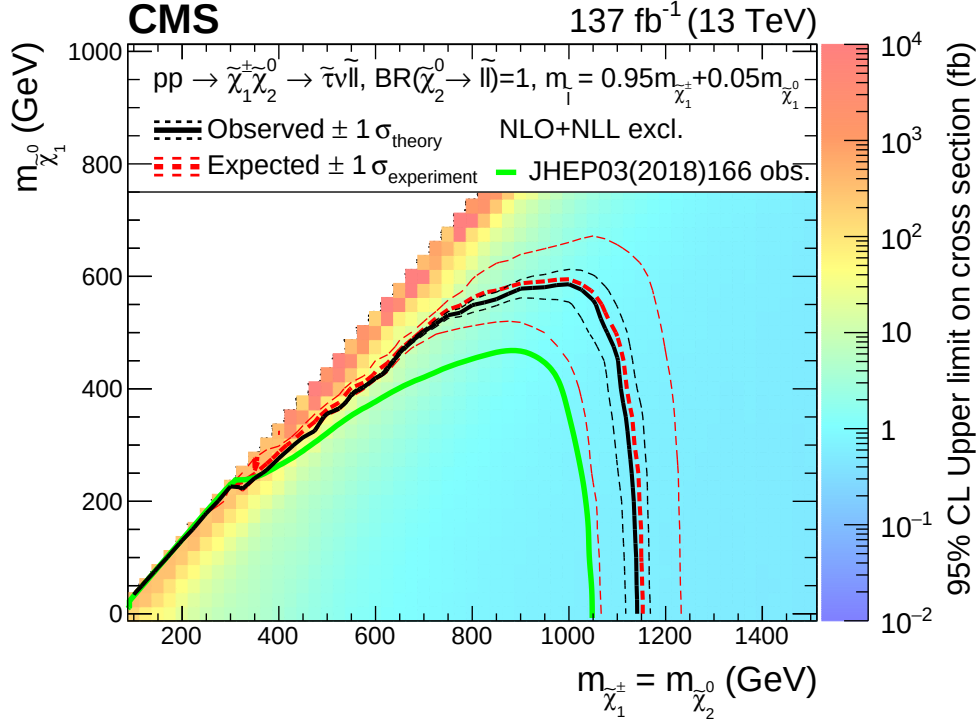


Figure 9.42: Interpretation of the search results for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with τ -enriched slepton-mediated superpartner decays at $x = 0.95$. The figure contents are explained in the caption of Fig. 9.37, with the difference that only a search region based analysis is used for this model.

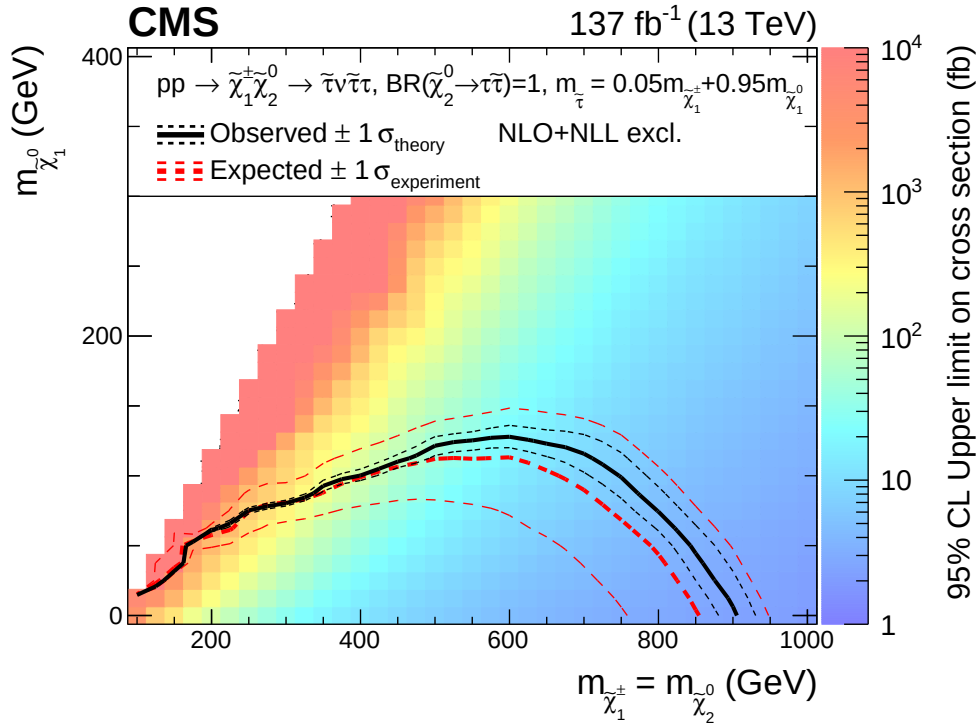


Figure 9.43: Interpretation of the search results for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with τ -dominated slepton-mediated superpartner decays at $x = 0.05$. The figure contents are explained in the caption of Fig. 9.37, with the difference that only a search region based analysis is used for this model. The results of the previous iteration of this search were not evaluated in terms of this $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ decay model, so no corresponding exclusion curve is available.

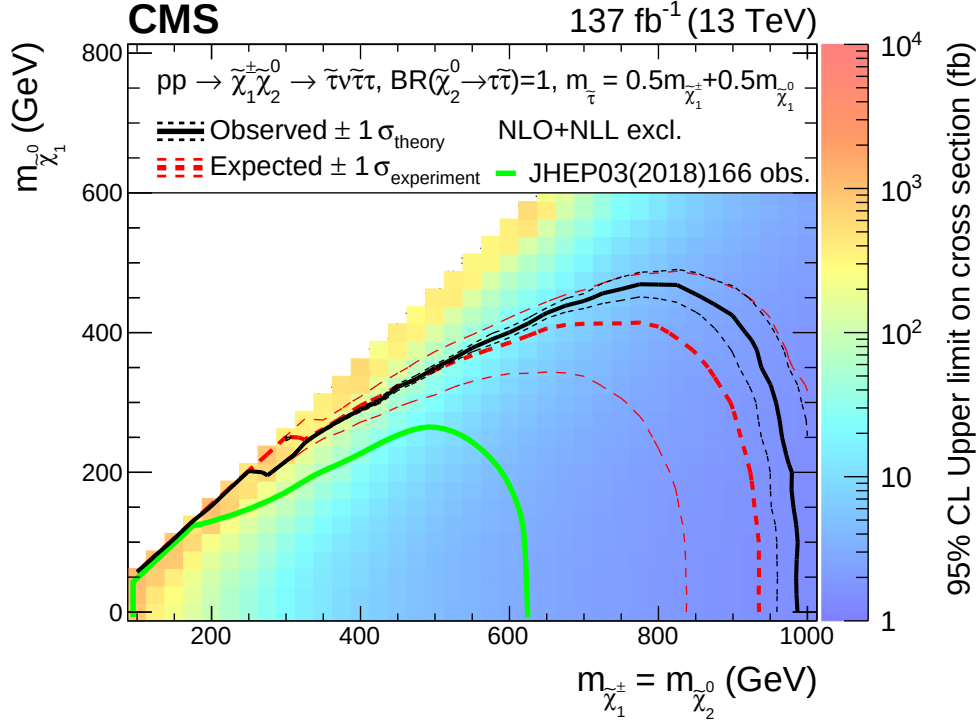


Figure 9.44: Interpretation of the search results for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with τ -dominated slepton-mediated superpartner decays at $x = 0.5$. The figure contents are explained in the caption of Fig. 9.37, with the difference that only a search region based analysis is used for this model.

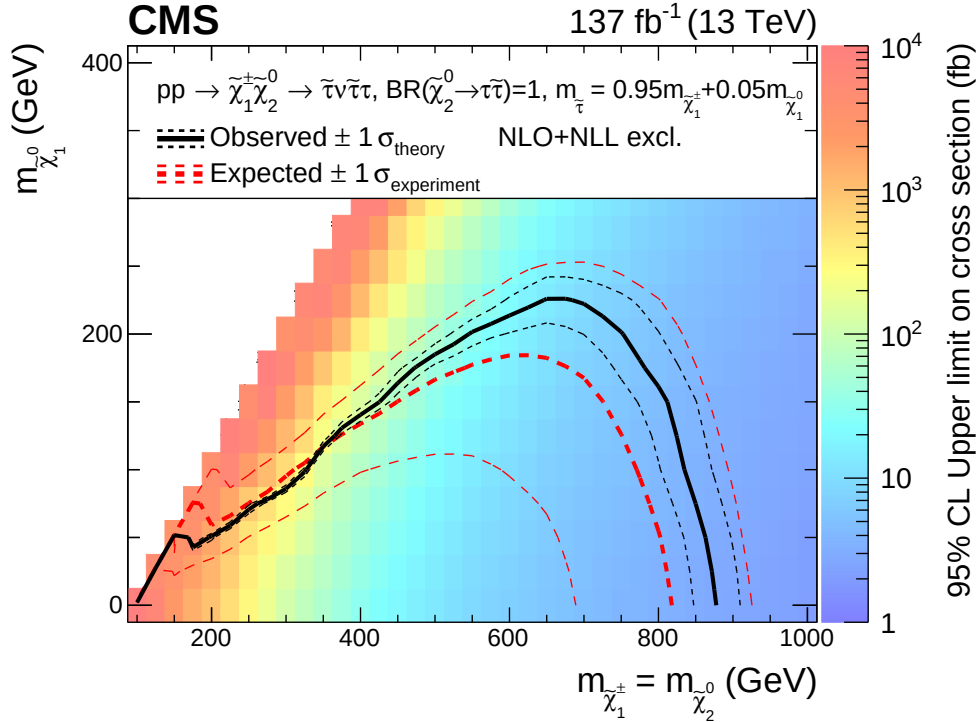


Figure 9.45: Interpretation of the search results for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with τ -dominated slepton-mediated superpartner decays at $x = 0.95$. The figure contents are explained in the caption of Fig. 9.37, with the difference that only a search region based analysis is used for this model. The results of the previous iteration of this search were not evaluated in terms of this $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ decay model, so no corresponding exclusion curve is available.

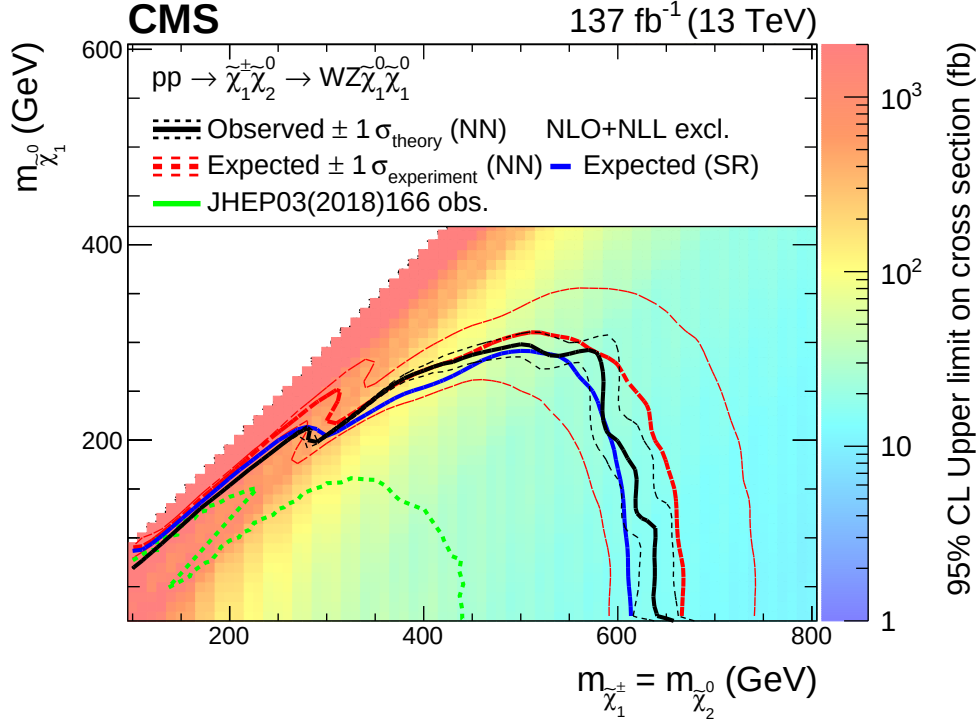


Figure 9.46: Interpretation of the search results for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with WZ-mediated superpartner decays at $x = 0.95$. The figure contents are explained in the caption of Fig. 9.37.

and are usually more than 200% at low δm values. The excluded mass range increases by 200 GeV in $m_{\tilde{\chi}_2^0}$ compared to the previous analysis iteration, meaning cross sections a factor 6 lower are excluded, and 130 GeV in $m_{\tilde{\chi}_1^0}$. The reach of the previous analysis was extremely limited in the WZ corridor (Section 9.4.1), resulting in a gap in the exclusion limits. This gap is almost completely closed by the new search. The presented results are the most stringent exclusion limits available in the WZ corridor at the time of writing.

The H boson has a highly diverse set of decay modes. All search categories are therefore combined to search for $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with WH-mediated superpartner decays. Of particular importance is category 3ℓB, in which the search regions are designed to exploit the spin correlations in $H \rightarrow WW \rightarrow \ell^+ \nu \ell^- \nu$ decays (Section 9.4.2). The exclusion limits for scenarios with WH-mediated superpartner decays can be found in Fig. 9.47. The exclusion limits extend up to 260 GeV in $m_{\tilde{\chi}_2^0}$ and 70 GeV in $m_{\tilde{\chi}_1^0}$. Compared to the previous analysis iteration the excluded $m_{\tilde{\chi}_2^0}$ range is extended by about 80 GeV, meaning the excluded cross sections are lowered by a factor 4. Meanwhile, the excluded $m_{\tilde{\chi}_1^0}$ range is increased by 40 GeV.

The search for models of effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production also makes use of all analyzed event categories. For ZZ-mediated superpartner decays 4ℓG events are the main source of sensitivity. If the decay goes via the emission of an H and a Z boson, all of the four lepton categories drive the sensitivity at low $m_{\tilde{\chi}_1^0}$ values, while events with three leptons rise to prominence at high $m_{\tilde{\chi}_1^0}$ values. When both neutralinos decay via the emission of an H boson the three lepton event categories are the most important ones. The exclusion limits for all effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production models are shown in Fig. 9.48 to 9.50. Under the hypothesis that both neutralinos decay to a Z boson, neutralino masses below 600 GeV are excluded. For decays via an H and a Z boson, neutralino masses below 400 GeV are excluded, while the exclusion limits only extend up to 200 GeV if both neutralinos decay by emitting an H boson.

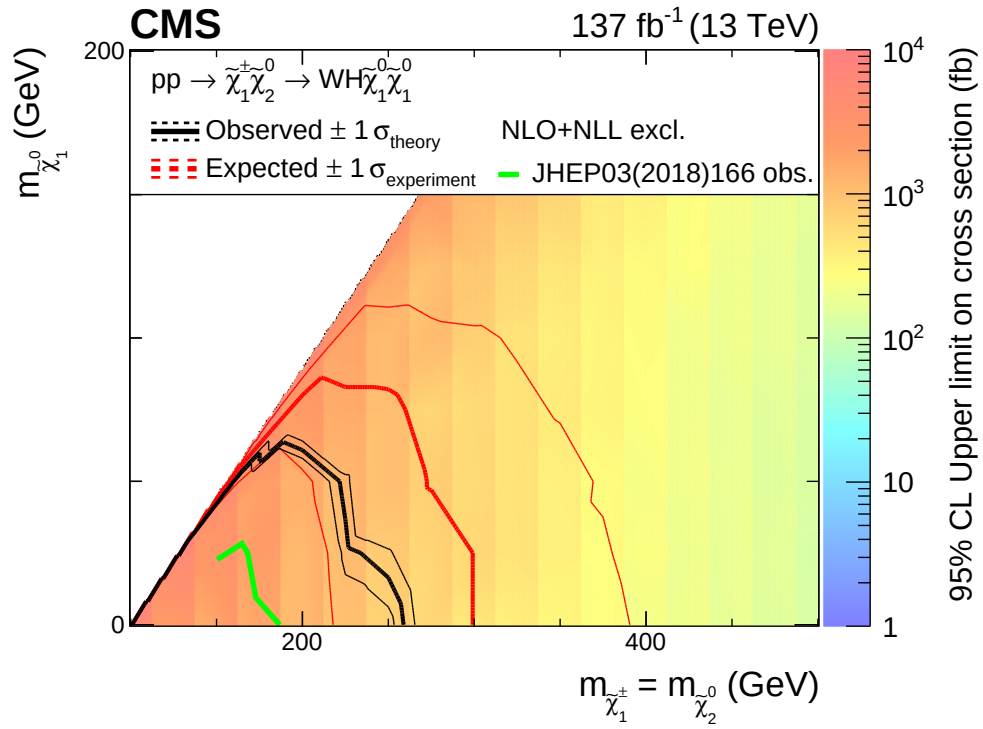


Figure 9.47: Interpretation of the search results for models of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with WH-mediated superpartner decays at $x = 0.5$. The figure contents are explained in the caption of Fig. 9.37, with the difference that only a search region based analysis is used for this model.

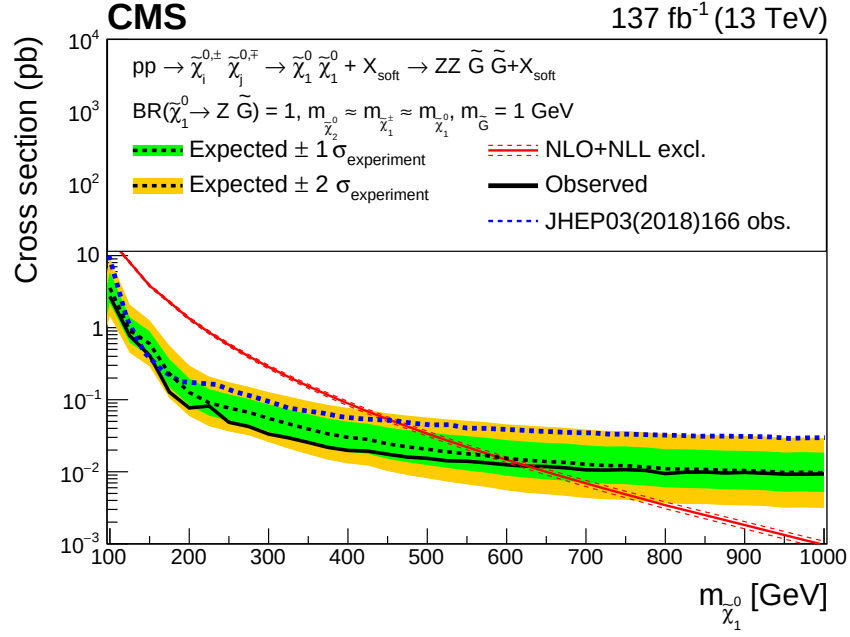


Figure 9.48: Interpretation of the search results for models of effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production with ZZ-mediated superpartner decays. The expected exclusion limits on the $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production cross section are shown as a black dashed line, with the corresponding one and two standard deviation contours shown in green and orange. These deviations come from experimental and theoretical uncertainties in the background prediction as well as experimental uncertainties in the prediction of the expected signal yields. The expected $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ pair production cross sections with their associated uncertainties are drawn in red. The solid black line represents the observed exclusion limits on the cross sections, and any $m_{\tilde{\chi}_1^0}$ value to the left of the crossing of the red and black lines is excluded. The observed exclusion limit for the previous iteration of the search [5] is drawn as a blue dashed line.

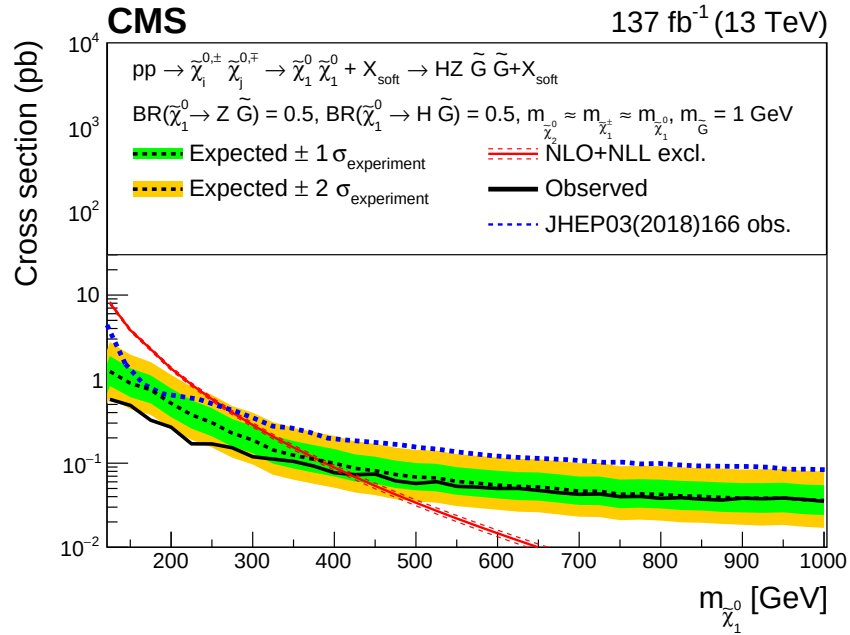


Figure 9.49: Interpretation of the search results for models of effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production with HZ-mediated superpartner decays. The figure contents are explained in the caption of Fig. 9.48.

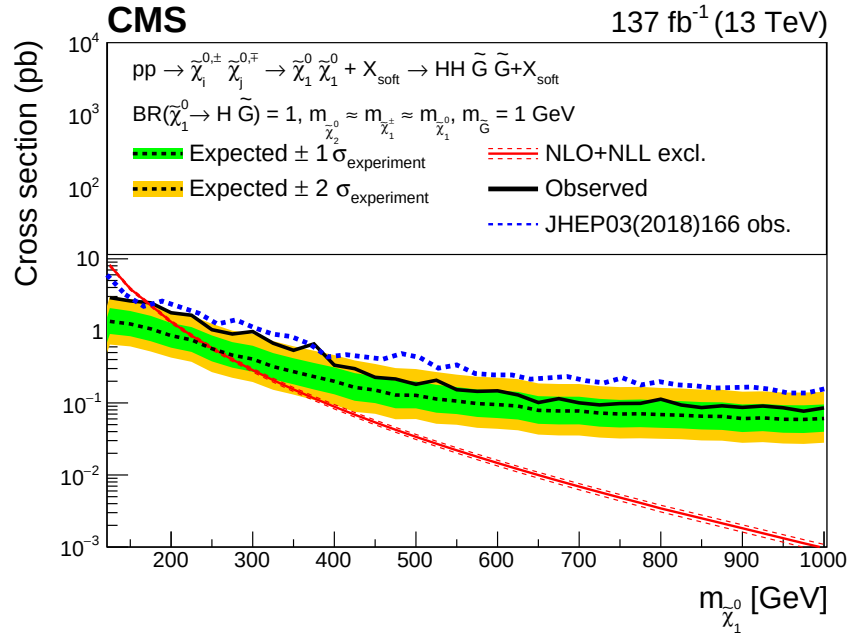


Figure 9.50: Interpretation of the search results for models of effective $\tilde{\chi}_1^0 \tilde{\chi}_1^0$ production with HH-mediated superpartner decays. The figure contents are explained in the caption of Fig. 9.48.

9.9 Summary and outlook

Supersymmetry is able to address many of the open problems in physics and was long deemed to be one of the most likely candidates for physics beyond the SM. The theory predicts the existence of superpartners: Fermionic partners to every SM boson and vice-versa for fermions. In the absence of evidence for the production of strongly interacting superpartners in a deluge of searches by the ATLAS and CMS collaborations, electroweak production of superpartners becomes the next search target.

An extensive search for the direct production of charginos and neutralinos, mass eigenstates composed of the superpartners of the electroweak gauge and H bosons, is presented. The number of possible manifestations of electroweak superpartner production is large, necessitating the analysis of a legion of final states. The search covers events with three or more leptons, including up to two hadronically decaying τ lepton (τ_h) candidates, as well as events with two light leptons of the same sign. The analyzed final states are relatively rare in the SM, while electroweak superpartner production might have large branching fractions to multilepton events.

Events are divided into twelve different categories by their lepton flavor and charge content. Each search category provides sensitivity to one or more distinct hypotheses of superpartner production. A set of orthogonal search region bins is optimized in each of the search categories. In events with three light leptons, at least two of which form an opposite sign and same flavor (OSSF) pair, parametric neural networks are used to search for signal events. This technique is adopted for the first time in such a search and results in massive improvements compared to using search region bins. For many signal hypotheses the expected analysis sensitivity is increased by more than 100%, and sometimes the improvements are even larger.

No significant excess of events over the SM expectation is observed in any of the search region bins or parametric neural network evaluations. The observed yields are used to set limits on a number of simplified benchmark models of chargino-neutralino pair production in which the lightest neutralino is the lightest supersymmetric particle, and models of effective neutralino pair production with a nearly massless gravitino as the lightest supersymmetric particle.

In supersymmetric scenarios where left-handed sleptons exist with masses lower than those of the charginos and heavy neutralinos, the chargino and neutralino decays could be mediated by intermediate sleptons. The branching fraction to multilepton events is high in such a scenario, and all lepton flavors would occur in the final state with equal probability. Under these circumstances the search sensitivity is dominated by events with three light leptons, at least two of which form an OSSF pair. With the parametric neural network analysis, lower limits in the chargino mass up to 1450 GeV are set. For compressed signal models, i.e. when some of the final state leptons are soft, events with two same-sign light leptons add significant sensitivity.

If only right-handed sleptons are lighter than the charginos, charginos would predominately decay to τ leptons. Under these conditions chargino masses up to 1150 GeV can be excluded. Another possibility is that the superpartners of the electrons and muons are out of reach, and only τ lepton superpartners are relatively light. In that case both charginos and neutralinos exclusively decay to τ leptons, and the lower limits on the chargino mass range up to 970 GeV. The sensitivity to all scenarios that favor τ leptons in the final state is significantly enhanced by analyzing events with τ_h candidates.

All slepton flavors could be heavy and decoupled. If that is the case, charginos and neutralinos have to decay via the emission of W, H, or Z bosons. For decays of chargino-neutralino pairs through W and Z bosons, events with three light leptons, two of which form an OSSF pair, once again make up the most important search category. The use of a parametric neural network in these events leads to lower limits in the chargino mass that extend up to 650 GeV. If the neutralino emits an H boson instead of a Z boson in its decay, chargino masses up to 300 GeV can be excluded. Given the multiplicity of H boson decay modes, many of the analyzed final states provide signal sensitivity under this decay hypothesis. The search region binning in events with three light leptons without an OSSF pair is tailored for identifying signals with H boson decays by exploiting the spin correlations in $H \rightarrow WW \rightarrow \ell^+ \nu \ell^- \nu$ decays.

In the studied models of effective neutralino pair production, both neutralinos decay via the emission of an H or Z boson. The analysis of events with four leptons in the final state is important for achieving maximum sensitivity to these models, but three lepton events also remain relevant. Neutralino masses up to 600 GeV are excluded in these scenarios.

The results of this analysis provide the strongest constraints on the production of charginos and neutralinos under several conditions. The exclusion limits under all hypotheses of slepton-mediated superpartner decays are at the time of writing the most stringent in existence. Under the hypothesis of chargino and neutralino decays via left-handed sleptons, superpartner masses up to 1450 GeV are excluded. These are the highest excluded masses for electroweak superpartners in any analysis at any experiment. When the chargino and neutralino decays are mediated by a W and a Z boson and the difference between the masses of the chargino and the lightest neutralino approaches the Z boson mass, the sensitivity of searches is significantly reduced. This is because signal events under these hypotheses look highly SM-like. This analysis provides by far the strongest limits in that challenging regime.

As in every analysis, further improvements can certainly be made. One area where there is a lot of potential for enhancing the sensitivity of the search is extending the usage of the parametric neural networks. Currently, parametric neural networks are only used in one final state and for a subset of the probed signal hypotheses. Using parametric neural networks brings about large improvements to the analysis sensitivity, in some cases more than 100% over the reach of search region bins, which are themselves partially inspired by the features learned by the parametric neural networks. It is not unlikely that similar improvements could be made for signal hypotheses for which the search currently relies only on orthogonal search region bins. Introducing more parametric neural networks would not come without cost: A great deal of additional complexity would be added to this already immense analysis.

The identification of τ_h candidates could be significantly improved, which would extend the search results for superpartner decay scenarios favoring τ leptons. Currently, we use BDT based τ_h identification algorithms, but CMS has recently introduced the DEEPTAU algorithm [355]. This is a deep neural network for identifying τ_h candidates, which has a vastly superior performance compared to BDT based methods. At a fixed efficiency for identifying prompt τ_h candidates, the background rate from misidentified jets could be reduced by about a factor two compared to the identification strategy used in the current search. Given the fact that the reducible background from nonprompt τ_h candidates is by far the largest background in the final states that include τ_h candidates this should be a promising and straightforward avenue for improving the presented results.

Significantly improving the analysis techniques for events with three light leptons, two of which form an OSSF pair, should be expected to be comparatively difficult. The parametric neural networks can certainly be improved, but they went through an extensive optimization procedure and seem to use all event information that could hint at the presence of charginos and neutralinos¹³. Further optimizations are likely to give only incremental improvements. Improving the lepton identification or event selection strategy will also give limited results since the irreducible background from SM WZ production already makes up the vast majority of background events.

Redoing the analysis once more data has been collected is a sure way of increasing the sensitivity of the search. The returns of a larger data set collected at the same center of mass energy are unfortunately quickly diminishing for supersymmetry searches such as this one. In the absence of changes in the analysis techniques, the expected sensitivity increases roughly with the square root of the data volume, as long as systematic uncertainties do not put a stop to this. Meanwhile, the cross sections for superpartner production decrease rapidly with the superpartner mass as shown in Fig. 9.1. Vastly better analysis techniques are therefore needed every time one wants to considerably increase the probed superpartner mass range. These odds are not favorable, and searches for supersymmetry in their current incarnation seem likely to fall out of favor until higher energy collisions than those currently provided by the LHC become available. Meanwhile, alternative final states, or kinematic

¹³At least to the extent that our simulation of chargino-neutralino production captures these features.

regions where the current sensitivity is limited seem to be more promising avenues for future work.

9.10 Acknowledgements

The completion of this analysis required a significantly greater effort than the other analyses presented in the dissertation, and more of the work was shared.

I was responsible for the design of the analysis in $3\ell A$ events, both the search region and the parametric neural network approaches. In particular I made the decision to use parametric neural networks, implemented, and optimized them. Since this is a novel technique that was never applied to such a search before, the verification and review of the results proved to be a long and arduous task. Producing the final analysis results and figures in $3\ell A$ events was also my responsibility. In addition, I measured the light lepton fake rate, fine tuned the loose lepton selection, measured the electron sign misidentification rates, and evaluated the $t\bar{t}Z$ and nonprompt control regions. I also designed and produced the test statistic plots which allow this massive analysis to be succinctly summarized. I served as the contact person for this analysis, wrote the first drafts of the paper, and was the main paper editor throughout the lengthy publication process. The majority of the internal documentation is also written by me.

Carlos Erice Cid's contributions to this analysis are equal to mine. He optimized the search region bins in all event categories other than $3\ell A$, and produced the analysis results in these categories. He evaluated the results of the WZ , ZZ , and $X\gamma$ control regions, measured the lepton selection efficiencies in collision data, and optimized the BDT thresholds used in the selection of leptons. The evaluation of the $W\gamma$ enriched data region used to derive uncertainties in the WZ background prediction was also done by Carlos, though by repeating a procedure I developed earlier. Carlos also made the exclusion limit plots. He helped editing the paper during the review process on several occasions. The preapproval and approval presentations of the analysis were given by Carlos in return for me being the contact person. Fig. 9.24 to 9.31 and 9.37 to 9.50 were made by Carlos.

Marco Peruzzi, Santiago Folgueras, and Pieter Everaerts were involved in many discussions relating to the analysis and gave important inputs on several occasions.

Liam Wezenbeek made some small contributions to this analysis. He determined which of the existing τ lepton identification thresholds would be the best to use in the search, and he measured the fake rate for τ_h candidates.

Chapter 10

Summary and outlook

*“You have not reached the end yet!
While you were on your way there
they were great marvels
three grievous wonders”*

Elias Lönnrot, *Kalevala*

Several pressing questions remain open in the SM. This dissertation highlighted a small part of the CMS Collaboration effort in addressing these questions. Three searches for distinct physical processes in the 13 TeV proton-proton collision data set collected with the CMS detector from 2016 to 2018 were presented. While the targets of the searches differ greatly, they share a large degree of experimental similarity. All of them exploit events with multiple charged leptons in the final state. Leptons do not interact via the strong interaction and therefore do not hadronize, providing an exceptionally clean signal in the detector. They are reconstructed and identified with high efficiency and low background rates. The added advantage of events with multiple leptons from the decay of heavy particles such as electroweak bosons, is that they are rare in the SM. Small irreducible and reducible background rates allow the presented searches to achieve a high sensitivity to their target signals, sometimes despite relatively low branching fractions.

Two of the discussed searches target physics beyond the SM. The first search aims to find sterile Majorana neutrinos in $W^{(*)} \rightarrow N\ell$ decays in which the sterile neutrino itself decays to two additional leptons (Chapter 7). Searches for sterile neutrinos in the multilepton final state had never been performed at the LHC up to the publication of this search. It was the first time sterile neutrinos with masses exceeding 500 GeV were directly probed, and the first search at a hadron collider sensitive to sterile neutrino masses in the range below 40 GeV. A more detailed summary of this analysis and a corresponding outlook is given in Section 7.9.

The last search presented in the dissertation targets the production of the superpartners of the electroweak bosons (Chapter 9). This search is the first of its kind to use parametric neural networks. The neural networks address difficulties in the optimization of orthogonal search region bins in searches with a broad and complex range of signal event topologies. Compared to classical analysis techniques based on orthogonal bins, the parametric neural networks result in large improvements to the reach of the search. Under many conditions the improvements brought by the neural networks are greater than what a quadrupling of the analyzed data volume would bring. The rest of the analysis strategy is also reoptimized compared to earlier work, and the search is the most powerful probe of electroweak superpartner production to date for a multitude of signal hypotheses. A more thorough summary can be found in Section 9.9.

In neither the search for sterile neutrinos nor the search for electroweak superpartner production, a significant excess of events over the SM expectation is observed. New physics remains elusive, and the observed event yields are used to place constraints on possible beyond the SM physics scenarios.

Aside from searches for physics beyond the SM, this dissertation covers a search for an extremely

rare SM process: Single top quark production in association with a Z boson (tZq). The process remained unobserved before the analysis presented in Chapter 8 was published. The use of machine learning in a crucial part of the analysis, the identification of signal leptons, proved to be the missing ingredient in earlier attempts to discover the process. We observe tZq production with a significance well above five standard deviations over the background-only hypothesis. The kinematics of the tZq process are extraordinarily sensitive to modifications to the SM by new physics, even if said physics would be beyond the direct reach of the LHC. We therefore study the Z boson p_T spectrum in very tZq-like events, but see no deviations from the SM expectation. Section 8.7 gives a comprehensive summary.

The searches for tZq production and electroweak production of superpartners both showcase the power of machine learning techniques. The main difference between the applied learning algorithms and classical particle identification or search strategies is that modern machine learning uses gradient based optimization schemes. The direction awareness in such optimization schemes results in performance that tends to be unilaterally better than what can be achieved by tuning sets of orthogonal thresholds. The search for sterile neutrinos that was presented could be improved greatly by taking inspiration from the two newer publications. The use of a machine learning based lepton identification, as well as a parametric neural network to identify sterile neutrino events, could probably make the analysis sensitive to sterile neutrino production cross sections several factors lower than those currently probed.

The analyses presented in this dissertation use the full 13 TeV proton-proton collision data set collected with the CMS detector from 2016 to 2018, or subsets thereof. This data set has a total integrated luminosity of 137 fb^{-1} , all of which is used in the search for electroweak superpartner production. The searches for sterile neutrinos and tZq production only use 35.9 and 77.4 fb^{-1} of pp collision data respectively because they were carried out while data was still being collected. Over its lifetime the LHC is planned to deliver an integrated luminosity of approximately 3000 fb^{-1} [372, 373]. This is an immense increase in the size of the data set, but not all of the covered research subjects will benefit from this equally.

The horizon might look the brightest for future searches for sterile neutrino production. As mentioned above, the use of machine learning techniques in the right parts of the analysis could extend the reach of the presented search by a similar amount as a data volume increase of several factors. The total data set expected to be collected with the LHC is almost 100 times larger than what was analyzed, so large improvements would obviously follow. The presented results also do not target sterile neutrinos that have a significant lifetime before decaying. This necessarily limits the reach at low sterile neutrino masses. This gap could be filled with dedicated searches for low-mass sterile neutrinos that target decay products originating from a vertex displaced with respect to the primary pp interaction vertex. Both ATLAS [276] and CMS [277] carried out such searches in the multilepton final state. At sterile neutrino masses of a few GeV these searches set upper limits on the square of the mixing parameter between sterile and SM neutrinos that are orders of magnitude lower than those set with the presented analysis. Twice as many sterile neutrino decays result in hadrons as in leptons, so further analyses of hadronic final states can further improve the coverage of sterile neutrino searches. These searches might target both high-mass sterile neutrinos resulting in jets originating from the primary pp interaction vertex and low-mass sterile neutrinos giving rise to displaced jets.

Sterile neutrinos with masses below several GeV can be probed at a number of different experiments than the LHC. Many of these provide very stringent exclusion limits, and a brief overview was given in Section 7.1. As soon as the neutrino masses one wants to probe exceed a few GeV however, the LHC is the only option. When it comes to the other two topics of the dissertation: Searches for tZq production and direct superpartner production, the LHC is the only machine that is currently able to provide results.

For tZq production we have now transitioned from searches to an era of precision measurements. The work presented here measured the total tZq production cross section with a relative uncertainty of 15%, twice as precise as any previous measurement. While the presented results constitute a

tremendous improvement over its predecessors, further improvements to the analysis techniques in the measurement of the integral tZq cross section will probably be small and incremental. This can be seen from the results obtained with the full Run 2 data set by the ATLAS [308] and CMS [245] Collaborations. The ATLAS result has the same relative uncertainty as the one presented in this dissertation, although 80% more data is analyzed. The newer CMS result is roughly equivalent to extrapolating the presented analysis to the greater integrated luminosity of the full Run 2 data set. Simply adding even more data will not improve the precision much further because the total statistical uncertainty in the latest tZq cross section measurements is almost as small as the systematic uncertainty. A reduction in the systematic uncertainties is necessary, though it must be noted that a larger data set might help with this indirectly. Differential tZq cross section measurements are a more promising avenue for future work. The kinematics of tZq production are highly sensitive to deviations from the SM [297], motivating such measurements. In addition, differential measurements of the tZq cross section are statistically limited and should be expected to become increasingly precise with the collection of more data for a long time to come. The latest CMS tZq measurement [245] sets the first step in this direction.

Searches for superpartner production seem to go towards a relatively uncertain future. Every search for both strong and electroweak superpartner production at the LHC has returned null results. In most analyzed final states results are now available that exploit the full LHC Run 2 data set. The search sensitivity one can gain from collecting more data is only proportional to the square root of the data volume. Meanwhile, superpartner production cross sections fall off sharply with the superpartner masses at a given collision center of mass energy. Any returns from the large data increase that is expected by the end of the LHC's operation should be expected to be small. Analysis improvements can undoubtedly be made, but it will be a hard battle against the rapidly decreasing cross sections. The situation could be somewhat improved if later runs of the LHC increase the center of mass energy of the collisions to 14 TeV. The increase in the window we can probe by going from 13 TeV to 14 TeV collisions is still rather limited. It might require a next generation collider becoming operational to take another leap in our knowledge about supersymmetry. In the meantime, alternative superpartner production and decay scenarios, and more exotic final states that are not covered by any of the current searches might be explored.

Bibliography

- [1] R. Brun and F. Rademakers. “ROOT: An object oriented data analysis framework”. *Nucl. Instrum. Meth. A* 389 (1997), p. 81. DOI: 10.1016/S0168-9002(97)00048-X.
- [2] I. Antcheva et al. “ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization”. *Comput. Phys. Commun.* 180 (2009), pp. 2499–2512. DOI: 10.1016/j.cpc.2009.08.005. arXiv: 1508.07749 [physics.data-an].
- [3] J. D. Hunter. “Matplotlib: A 2D graphics environment”. *Computing in Science & Engineering* 9 (2007), p. 90. DOI: 10.1109/MCSE.2007.55.
- [4] T. Tantau. “The TikZ and PGF packages”. 2013. URL: <https://github.com/pgf-tikz/pgf>.
- [5] CMS Collaboration. “Search for electroweak production of charginos and neutralinos in multi-lepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JHEP* 03 (2018), p. 166. DOI: 10.1007/JHEP03(2018)166. arXiv: 1709.05406 [hep-ex].
- [6] CMS Collaboration. “Search for heavy neutral leptons in events with three charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *Phys. Rev. Lett.* 120 (2018), p. 221801. DOI: 10.1103/PhysRevLett.120.221801. arXiv: 1802.02965 [hep-ex].
- [7] CMS Collaboration. “CMS hunts for heavy neutral leptons”. *CERN Courier* (2018). URL: <https://cerncourier.com/a/cms-hunts-for-heavy-neutral-leptons/>.
- [8] W. Verbeke. “Search for heavy neutral leptons with the CMS detector”. *53rd Rencontres de Moriond on Electroweak Interactions and Unified Theories* (2018), p. 431. arXiv: 1805.05084 [hep-ex].
- [9] CMS Collaboration. “The CMS collaboration has observed a new production mechanism of top quarks”. 2018. URL: <https://cms.cern/news/cms-collaboration-has-observed-new-production-mechanism-top-quarks>.
- [10] W. Verbeke. “Single top quark and rare top quark production at ATLAS and CMS”. *54th Rencontres de Moriond on QCD and High Energy Interactions* (2019), p. 43. arXiv: 1908.07271 [hep-ex].
- [11] W. Verbeke. “Observation of single top quark production in association with a Z boson”. *PoS EPS-HEP2019* (2020), p. 683. DOI: 10.22323/1.364.0683.
- [12] CMS Collaboration. “Search for electroweak production of charginos and neutralinos in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JHEP* 04 (2022), p. 147. DOI: 10.1007/JHEP04(2022)147. arXiv: 2106.14246 [hep-ex].
- [13] CMS Collaboration. “Deep learning tailors supersymmetry searches”. *CERN Courier* (2021). URL: <https://cerncourier.com/a/deep-learning-tailors-supersymmetry-searches/>.
- [14] 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* 716 (2012), p. 1. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].
- [15] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. *Phys. Lett. B* 716 (2012), p. 30. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

- [16] M. Gell-Mann. “The Eightfold Way: A Theory of strong interaction symmetry” (1961). DOI: 10.2172/4008239.
- [17] M. Gell-Mann. “A schematic model of baryons and mesons”. *Phys. Lett.* 8 (1964), p. 214. DOI: 10.1016/S0031-9163(64)92001-3.
- [18] R. Davis, D. S. Harmer, and K. C. Hoffman. “Search for neutrinos from the sun”. *Phys. Rev. Lett.* 20 (1968), p. 1205. DOI: 10.1103/PhysRevLett.20.1205.
- [19] SNO Collaboration. “Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory”. *Phys. Rev. Lett.* 87 (2001), p. 071301. DOI: 10.1103/PhysRevLett.87.071301. arXiv: nucl-ex/0106015.
- [20] C. S. Wu et al. “Experimental test of parity conservation in beta decay”. *Phys. Rev.* 105 (1957), p. 1413. DOI: 10.1103/PhysRev.105.1413.
- [21] J. H. Christenson et al. “Evidence for the 2π decay of the K_2^0 meson”. *Phys. Rev. Lett.* 13 (1964), p. 138. DOI: 10.1103/PhysRevLett.13.138.
- [22] KTeV Collaboration. “Observation of direct CP violation in $K_{S,L} \rightarrow \pi\pi$ decays”. *Phys. Rev. Lett.* 83 (1999), p. 22. DOI: 10.1103/PhysRevLett.83.22.
- [23] F. Englert and R. Brout. “Broken symmetry and the mass of gauge vector mesons”. *Phys. Rev. Lett.* 13 (1964), p. 321. DOI: 10.1103/PhysRevLett.13.321.
- [24] P. W. Higgs. “Broken symmetries and the masses of gauge bosons”. *Phys. Rev. Lett.* 13 (1964), p. 508. DOI: 10.1103/PhysRevLett.13.508.
- [25] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global conservation laws and massless particles”. *Phys. Rev. Lett.* 13 (1964), p. 585. DOI: 10.1103/PhysRevLett.13.585.
- [26] ATLAS Collaboration. “Evidence for the spin-0 nature of the Higgs boson using ATLAS data”. *Phys. Lett. B* 726 (2013), p. 120. DOI: 10.1016/j.physletb.2013.08.026. arXiv: 1307.1432 [hep-ex].
- [27] CMS Collaboration. “Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV”. *Phys. Rev. D* 92 (2015), p. 012004. DOI: 10.1103/PhysRevD.92.012004. arXiv: 1411.3441 [hep-ex].
- [28] D. G. Carsten Burgard. “The standard model infographic”. URL: <https://texample.net/tikz/examples/model-physics/>.
- [29] Particle Data Group. “Review of particle physics”. *PTEP* (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.
- [30] M. E. Peskin and D. V. Schroeder. “An introduction to quantum field theory”. Addison-Wesley, 1995. ISBN: 9780201503975.
- [31] F. Mandl and G. Shaw. “Quantum field theory”. 2nd ed. Wiley, 2010. ISBN: 9780471496830.
- [32] W. Pauli. “The connection between spin and statistics”. *Phys. Rev.* 58 (1940), p. 716. DOI: 10.1103/PhysRev.58.716.
- [33] J. Schwinger. “On quantum-electrodynamics and the magnetic moment of the electron”. *Phys. Rev.* 73 (1948), p. 416. DOI: 10.1103/PhysRev.73.416.
- [34] J. Schwinger. “Quantum electrodynamics. I. A covariant formulation”. *Phys. Rev.* 74 (1948), p. 1439. DOI: 10.1103/PhysRev.74.1439.
- [35] S. Tomonaga. “On a relativistically invariant formulation of the quantum theory of wave fields”. *Prog. Theor. Phys.* 1 (1946), p. 27. DOI: 10.1143/PTP.1.27.
- [36] R. P. Feynman. “Space-time approach to quantum electrodynamics”. *Phys. Rev.* 76 (1949), p. 769. DOI: 10.1103/PhysRev.76.769.

- [37] C. N. Yang and R. L. Mills. “Conservation of isotopic spin and isotopic gauge invariance”. *Phys. Rev.* 96 (1954), p. 191. DOI: 10.1103/PhysRev.96.191.
- [38] O. W. Greenberg. “Spin and unitary-spin independence in a paraquark model of baryons and mesons”. *Phys. Rev. Lett.* 13 (1964), p. 598. DOI: 10.1103/PhysRevLett.13.598.
- [39] M. Y. Han and Y. Nambu. “Three-triplet model with double SU(3) symmetry”. *Phys. Rev.* 139 (1965), B1006. DOI: 10.1103/PhysRev.139.B1006.
- [40] T. Sjöstrand, S. Mrenna, and P. Z. Skands. “PYTHIA 6.4 physics and manual”. *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175.
- [41] CMS Collaboration. “Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV”. *JHEP* 03 (2017), p. 156. DOI: 10.1007/JHEP03(2017)156. arXiv: 1609.05331 [hep-ex].
- [42] Gargamelle Collaboration. “Observation of neutrino like interactions without muon or electron in the Gargamelle neutrino experiment”. *Phys. Lett. B* 46 (1973), p. 138. DOI: 10.1016/0370-2693(73)90499-1.
- [43] R. P. Feynman and M. Gell-Mann. “Theory of the Fermi interaction”. *Phys. Rev.* 109 (1958), p. 193. DOI: 10.1103/PhysRev.109.193.
- [44] E. C. G. Sudarshan and R. E. Marshak. “Chirality invariance and the universal Fermi interaction”. *Phys. Rev.* 109 (1958), p. 1860. DOI: 10.1103/PhysRev.109.1860.2.
- [45] S. L. Glashow. “The renormalizability of vector meson interactions”. *Nucl. Phys.* 10 (1959), p. 107. DOI: 10.1016/0029-5582(59)90196-8.
- [46] A. Salam and J. C. Ward. “Weak and electromagnetic interactions”. *Nuovo Cim.* 11 (1959), p. 568. DOI: 10.1007/BF02726525.
- [47] S. Weinberg. “A model of leptons”. *Phys. Rev. Lett.* 19 (1967), p. 1264. DOI: 10.1103/PhysRevLett.19.1264.
- [48] C. Grupen. “Astroparticle physics”. Springer, 2005. ISBN: 9783540253129.
- [49] M. Thomson. “Modern particle physics”. Cambridge University Press, 2013. ISBN: 9781107034266.
- [50] UA1 Collaboration. “Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV”. *Phys. Lett. B* 122 (1983), p. 103. DOI: 10.1016/0370-2693(83)91177-2.
- [51] UA2 Collaboration. “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $\bar{p}p$ collider”. *Phys. Lett. B* 122 (1983), p. 476. DOI: 10.1016/0370-2693(83)91605-2.
- [52] UA1 Collaboration. “Experimental observation of lepton pairs of invariant mass around 95 GeV/c² at the CERN SPS collider”. *Phys. Lett. B* 126 (1983), p. 398. DOI: 10.1016/0370-2693(83)90188-0.
- [53] UA2 Collaboration. “Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN $\bar{p}p$ collider”. *Phys. Lett. B* 129 (1983), p. 130. DOI: 10.1016/0370-2693(83)90744-X.
- [54] ATLAS Collaboration. “Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector”. *Phys. Lett. B* 786 (2018), p. 59. DOI: 10.1016/j.physletb.2018.09.013. arXiv: 1808.08238 [hep-ex].
- [55] CMS Collaboration. “Observation of Higgs boson decay to bottom quarks”. *Phys. Rev. Lett.* 121 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex].
- [56] ATLAS Collaboration. “Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector”. *JHEP* 04 (2015), p. 117. DOI: 10.1007/JHEP04(2015)117. arXiv: 1501.04943 [hep-ex].

- [57] ATLAS Collaboration. “Cross-section measurements of the Higgs boson decaying into a pair of τ -leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *Phys. Rev. D* 99 (2019), p. 072001. DOI: 10.1103/PhysRevD.99.072001. arXiv: 1811.08856 [hep-ex].
- [58] CMS Collaboration. “Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector”. *Phys. Lett. B* 779 (2018), p. 283. DOI: 10.1016/j.physletb.2018.02.004. arXiv: 1708.00373 [hep-ex].
- [59] CMS Collaboration. “Measurement of the inclusive and differential Higgs boson production cross sections in the decay mode to a pair of τ leptons in pp collisions at $\sqrt{s} = 13$ TeV”. *Phys. Rev. Lett.* 128 (2022), p. 081805. DOI: 10.1103/PhysRevLett.128.081805. arXiv: 2107.11486 [hep-ex].
- [60] CMS Collaboration. “Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV”. *Eur. Phys. J. C* 81 (2021), p. 378. DOI: 10.1140/epjc/s10052-021-09014-x. arXiv: 2011.03652 [hep-ex].
- [61] CMS Collaboration. “Measurements of $t\bar{t}H$ production and the CP structure of the Yukawa interaction between the Higgs boson and top quark in the diphoton decay channel”. *Phys. Rev. Lett.* 125 (2020), p. 061801. DOI: 10.1103/PhysRevLett.125.061801. arXiv: 2003.10866 [hep-ex].
- [62] ATLAS Collaboration. “Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector” (2021). arXiv: 2111.06712 [hep-ex].
- [63] ATLAS Collaboration. “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector”. *Phys. Lett. B* 784 (2018), p. 173. DOI: 10.1016/j.physletb.2018.07.035. arXiv: 1806.00425 [hep-ex].
- [64] N. Cabibbo. “Unitary symmetry and leptonic decays”. *Phys. Rev. Lett.* 10 (1963), p. 531. DOI: 10.1103/PhysRevLett.10.531.
- [65] M. Kobayashi and T. Maskawa. “CP-Violation in the renormalizable theory of weak interaction”. *Prog. Theor. Phys.* 49 (1973), p. 652. DOI: 10.1143/PTP.49.652.
- [66] B. Pontecorvo. “Inverse beta processes and nonconservation of lepton charge”. *Zh. Eksp. Teor. Fiz.* 34 (1957), p. 247.
- [67] Z. Maki, M. Nakagawa, and S. Sakata. “Remarks on the unified model of elementary particles”. *Prog. Theor. Phys.* 28 (1962), p. 870. DOI: 10.1143/PTP.28.870.
- [68] D. J. Bird et al. “Detection of a cosmic ray with measured energy well beyond the expected spectral cutoff due to cosmic microwave radiation”. *Astrophys. J.* 441 (1995), p. 144. DOI: 10.1086/175344. arXiv: astro-ph/9410067.
- [69] J. C. Kapteyn. “First attempt at a theory of the arrangement and motion of the sidereal system”. *Astrophys. J.* 55 (1922), p. 302. DOI: 10.1086/142670.
- [70] K. Lundmark. “Über die bestimmung der entfernungen, dimensionen, massen und dichtigkeit für die nächstgelegenen anagalactischen sternsysteme.” *Meddelanden fran Lunds Astronomiska Observatorium Serie I* 125 (1930), p. 1.
- [71] J. H. Oort. “Some problems concerning the structure and dynamics of the galactic system and the elliptical nebulae NGC 3115 and 4494.” *Astrophys. J.* 91 (1940), p. 273. DOI: 10.1086/144167.
- [72] F. Zwicky. “On the masses of nebulae and of clusters of nebulae”. *Astrophys. J.* 86 (1937), p. 217. DOI: 10.1086/143864.
- [73] S. M. Faber and R. E. Jackson. “Velocity dispersions and mass-to-light ratios for elliptical galaxies.” *Astrophys. J.* 204 (1976), p. 668. DOI: 10.1086/154215.

- [74] H. W. Babcock. “The rotation of the Andromeda nebula”. *Lick Observatory Bulletin* 498 (1939), p. 41. DOI: 10.5479/ADS/bib/1939LicOB.19.41B.
- [75] V. C. Rubin and J. Ford W. Kent. “Rotation of the Andromeda nebula from a spectroscopic survey of emission regions”. *Astrophys. J.* 159 (1970), p. 379. DOI: 10.1086/150317.
- [76] K. C. Freeman. “On the disks of spiral and S0 galaxies”. *Astrophys. J.* 160 (1970), p. 811. DOI: 10.1086/150474.
- [77] V. C. Rubin, J. Ford W. K., and N. Thonnard. “Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc).” *Astrophys. J.* 238 (1980), p. 471. DOI: 10.1086/158003.
- [78] M. Milgrom. “MOND laws of galactic dynamics”. *Mon. Not. Roy. Astron. Soc.* 437 (2014), p. 2531. DOI: 10.1093/mnras/stt2066. arXiv: 1212.2568 [astro-ph.CO].
- [79] P. Natarajan et al. “Mapping substructure in the HST Frontier Fields cluster lenses and in cosmological simulations”. *Mon. Not. Roy. Astron. Soc.* 468 (2017), p. 1962. DOI: 10.1093/mnras/stw3385. arXiv: 1702.04348 [astro-ph.GA].
- [80] D. Clowe et al. “A direct empirical proof of the existence of dark matter”. *Astrophys. J. Lett.* 648 (2006), p. L109. DOI: 10.1086/508162. arXiv: astro-ph/0608407.
- [81] C. L. Bennett et al. “Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Final maps and results”. *Astrophys. J., Supplement* 208, 20 (2013), p. 20. DOI: 10.1088/0067-0049/208/2/20. arXiv: 1212.5225 [astro-ph.CO].
- [82] Planck Collaboration. “Planck 2018 results. VI. Cosmological parameters”. *Astron. Astrophys.* 641 (2020). [Erratum: *Astron. Astrophys.* 652, C4 (2021)], A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].
- [83] S. W. Allen, A. E. Evrard, and A. B. Mantz. “Cosmological parameters from observations of galaxy clusters”. *Annual Review of Astronomy and Astrophysics* 49 (2011), p. 409. DOI: 10.1146/annurev-astro-081710-102514.
- [84] A. Boveia and C. Doglioni. “Dark matter searches at colliders”. *Ann. Rev. Nucl. Part. Sci.* 68 (2018), p. 429. DOI: 10.1146/annurev-nucl-101917-021008. arXiv: 1810.12238 [hep-ex].
- [85] M. Schumann. “Direct detection of WIMP dark matter: Concepts and status”. *J. Phys. G* 46 (2019), p. 103003. DOI: 10.1088/1361-6471/ab2ea5. arXiv: 1903.03026 [astro-ph.CO].
- [86] R. D. Peccei and H. R. Quinn. “CP conservation in the presence of pseudoparticles”. *Phys. Rev. Lett.* 38 (1977), p. 1440. DOI: 10.1103/PhysRevLett.38.1440.
- [87] R. D. Peccei and H. R. Quinn. “Constraints imposed by CP conservation in the presence of pseudoparticles”. *Phys. Rev. D* 16 (1977), p. 1791. DOI: 10.1103/PhysRevD.16.1791.
- [88] P. W. Graham et al. “Experimental searches for the axion and axion-like particles”. *Ann. Rev. Nucl. Part. Sci.* 65 (2015), p. 485. DOI: 10.1146/annurev-nucl-102014-022120. arXiv: 1602.00039 [hep-ex].
- [89] T. Asaka, S. Blanchet, and M. Shaposhnikov. “The ν MSM, dark matter and neutrino masses”. *Phys. Lett. B* 631 (2005), p. 151. DOI: 10.1016/j.physletb.2005.09.070. arXiv: hep-ph/0503065 [hep-ph].
- [90] S. S. Team. “Observational evidence from supernovae for an accelerating universe and a cosmological constant”. *Astron. J.* 116 (1998), p. 1009. DOI: 10.1086/300499. arXiv: astro-ph/9805201.
- [91] S. C. Project. “Measurements of Ω and Λ from 42 high redshift supernovae”. *Astrophys. J.* 517 (1999), p. 565. DOI: 10.1086/307221. arXiv: astro-ph/9812133.
- [92] D. H. Weinberg et al. “Observational probes of cosmic acceleration”. *Phys. Rep.* 530 (2013), p. 87. DOI: 10.1016/j.physrep.2013.05.001.

- [93] A. Joyce et al. “Beyond the cosmological standard model”. *Phys. Rept.* 568 (2015), p. 1. DOI: 10.1016/j.physrep.2014.12.002. arXiv: 1407.0059 [astro-ph.CO].
- [94] R. J. Adler, B. Casey, and O. C. Jacob. “Vacuum catastrophe: An elementary exposition of the cosmological constant problem”. *Am. J. Phys.* 63 (1995), p. 620. DOI: 10.1119/1.17850.
- [95] A. O. Barvinsky, A. Y. Kamenshchik, and T. Vardanyan. “Comment about the vanishing of the vacuum energy in the Wess–Zumino model”. *Phys. Lett. B* 782 (2018), p. 55. DOI: 10.1016/j.physletb.2018.05.008. arXiv: 1804.03394 [hep-th].
- [96] AMS Collaboration. “Search for anti-helium in cosmic rays”. *Phys. Lett. B* 461 (1999), p. 387. DOI: 10.1016/S0370-2693(99)00874-6. arXiv: hep-ex/0002048.
- [97] L. Canetti, M. Drewes, and M. Shaposhnikov. “Matter and antimatter in the universe”. *New J. Phys.* 14 (2012), p. 095012. DOI: 10.1088/1367-2630/14/9/095012. arXiv: 1204.4186 [hep-ph].
- [98] A. D. Sakharov. “Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe”. *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), p. 32. DOI: 10.1070/PU1991v034n05ABEH002497.
- [99] M. Baldo-Ceolin et al. “A new experimental limit on neutron-antineutron oscillations”. *Z. Phys. C* 63 (1994), p. 409. DOI: 10.1007/BF01580321.
- [100] Super-Kamiokande Collaboration. “Search for proton decay via $p \rightarrow e^+ \pi^0$ and $p \rightarrow \mu^+ \pi^0$ in 0.31 megaton-years exposure of the Super-Kamiokande water Cherenkov detector”. *Phys. Rev. D* 95 (2017), p. 012004. DOI: 10.1103/PhysRevD.95.012004. arXiv: 1610.03597 [hep-ex].
- [101] S.-K. Collaboration. “The Search for $n - \bar{n}$ oscillation in Super-Kamiokande I”. *Phys. Rev. D* 91 (2015), p. 072006. DOI: 10.1103/PhysRevD.91.072006. arXiv: 1109.4227 [hep-ex].
- [102] T. Appelquist and R. Shrock. “Neutrino masses in theories with dynamical electroweak symmetry breaking”. *Phys. Lett. B* 548 (2002), p. 204. DOI: 10.1016/S0370-2693(02)02854-X. arXiv: hep-ph/0204141 [hep-ph].
- [103] T. Appelquist and R. Shrock. “Dynamical symmetry breaking of extended gauge symmetries”. *Phys. Rev. Lett.* 90 (2003), p. 201801. DOI: 10.1103/PhysRevLett.90.201801. arXiv: hep-ph/0301108 [hep-ph].
- [104] T. Asaka and M. Shaposhnikov. “The ν MSM, dark matter and baryon asymmetry of the universe”. *Phys. Lett. B* 620 (2005), p. 17. DOI: 10.1016/j.physletb.2005.06.020. arXiv: hep-ph/0505013 [hep-ph].
- [105] J. Schechter and J. W. F. Valle. “Neutrino masses in $SU(2) \otimes U(1)$ theories”. *Phys. Rev. D* 22 (1980), p. 2227. DOI: 10.1103/PhysRevD.22.2227.
- [106] ALEPH, DELPHI, L3, OPAL, and SLD Collaborations, LEP Electroweak Working Group, SLD Electroweak Group, and SLD Heavy Flavour Group. “Precision electroweak measurements on the Z resonance”. *Phys. Rept.* 427 (2006), p. 257. DOI: 10.1016/j.physrep.2005.12.006. arXiv: hep-ex/0509008.
- [107] KATRIN Collaboration. “Improved upper limit on the neutrino mass from a direct kinematic method by KATRIN”. *Phys. Rev. Lett.* 123 (2019), p. 221802. DOI: 10.1103/PhysRevLett.123.221802. arXiv: 1909.06048 [hep-ex].
- [108] ALEPH Collaboration. “An upper limit on the tau-neutrino mass from three-prong and five-prong tau decays”. *Eur. Phys. J. C* 2 (1998), p. 395. DOI: 10.1007/s100520050149.
- [109] P. Minkowski. “ $\mu \rightarrow e \gamma$ at a rate of one out of 1-billion muon decays?” *Phys. Lett. B* 67 (1977), p. 421. DOI: 10.1016/0370-2693(77)90435-X.
- [110] M. Gell-Mann, P. Ramond, and R. Slansky. “Supergravity: proceedings of the supergravity workshop at Stony Brook”. Ed. by P. V. Nieuwenhuizen and D. Z. Freedman. North-holland, 1979.

- [111] T. Yanagida. “Proceedings: Workshop on the unified theories and the baryon number in the universe”. Ed. by O. Sawada and A. Sugamoto. Natl. Lab. High Energy Phys., 1979.
- [112] R. N. Mohapatra and G. Senjanović. “Neutrino mass and spontaneous parity violation”. *Phys. Rev. Lett.* 44 (1980), p. 912. DOI: 10.1103/PhysRevLett.44.912.
- [113] M. Magg and C. Wetterich. “Neutrino mass problem and gauge hierarchy”. *Phys. Lett. B* 94 (1980), p. 61. DOI: 10.1016/0370-2693(80)90825-4.
- [114] T. P. Cheng and L.-F. Li. “Neutrino masses, mixings and oscillations in $SU(2) \otimes U(1)$ models of electroweak interactions”. *Phys. Rev. D* 22 (1980), p. 2860. DOI: 10.1103/PhysRevD.22.2860.
- [115] G. Lazarides, Q. Shafi, and C. Wetterich. “Proton lifetime and fermion masses in an $SO(10)$ model”. *Nucl. Phys. B* 181 (1981), p. 287. DOI: 10.1016/0550-3213(81)90354-0.
- [116] R. N. Mohapatra and G. Senjanović. “Neutrino masses and mixings in gauge models with spontaneous parity violation”. *Phys. Rev. D* 23 (1981), p. 165. DOI: 10.1103/PhysRevD.23.165.
- [117] J. Schechter and J. W. F. Valle. “Neutrino decay and spontaneous violation of lepton number”. *Phys. Rev. D* 25 (1982), p. 774. DOI: 10.1103/PhysRevD.25.774.
- [118] R. Foot et al. “See-saw neutrino masses induced by a triplet of leptons”. *Z. Phys. C* 44 (1989), p. 441. DOI: 10.1007/BF01415558.
- [119] S. P. Martin. “A supersymmetry primer”. *Advanced Series on Directions in High Energy Physics* (1998), p. 1. DOI: 10.1142/9789812839657_0001.
- [120] Muon g-2 Collaboration. “Measurement of the positive muon anomalous magnetic moment to 0.46 ppm”. *Phys. Rev. Lett.* 126 (2021), p. 141801. DOI: 10.1103/PhysRevLett.126.141801. arXiv: 2104.03281 [hep-ex].
- [121] H.-B. Zhang et al. “Muon anomalous magnetic dipole moment in the $\mu\nu$ SSM” (2021). arXiv: 2104.03489 [hep-ph].
- [122] S. Borsanyi et al. “Leading hadronic contribution to the muon magnetic moment from lattice QCD”. *Nature* 593 (2021), p. 51. DOI: 10.1038/s41586-021-03418-1. arXiv: 2002.12347 [hep-lat].
- [123] LHCb Collaboration. “Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\mu^-\bar{\nu}_\mu)$ ”. *Phys. Rev. Lett.* 115 (2015). [Erratum: *Phys.Rev.Lett.* 115, 159901 (2015)], p. 111803. DOI: 10.1103/PhysRevLett.115.111803. arXiv: 1506.08614 [hep-ex].
- [124] Belle Collaboration. “Measurement of the branching ratio of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle”. *Phys. Rev. D* 92 (2015), p. 072014. DOI: 10.1103/PhysRevD.92.072014. arXiv: 1507.03233 [hep-ex].
- [125] BaBar Collaboration. “Evidence for an excess of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ decays”. *Phys. Rev. Lett.* 109 (2012), p. 101802. DOI: 10.1103/PhysRevLett.109.101802. arXiv: 1205.5442 [hep-ex].
- [126] B. Capdevila et al. “Patterns of new physics in $b \rightarrow s\ell^+\ell^-$ transitions in the light of recent data”. *JHEP* 01 (2018), p. 093. DOI: 10.1007/JHEP01(2018)093. arXiv: 1704.05340 [hep-ph].
- [127] LHCb Collaboration. “Test of lepton universality in beauty-quark decays” (2021). arXiv: 2103.11769 [hep-ex].
- [128] CDF Collaboration. “High-precision measurement of the W boson mass with the CDF II detector”. *Science* 376 (2022), p. 170. DOI: 10.1126/science.abk1781.
- [129] D. López-Val and T. Robens. “ Δr and the W boson mass in the singlet extension of the standard model”. *Phys. Rev. D* 90 (2014), p. 114018. DOI: 10.1103/PhysRevD.90.114018. arXiv: 1406.1043 [hep-ph].
- [130] ATLAS Collaboration. “Measurement of the W boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”. *Eur. Phys. J. C* 78 (2018). [Erratum: *Eur.Phys.J.C* 78, 898 (2018)], p. 110. DOI: 10.1140/epjc/s10052-017-5475-4. arXiv: 1701.07240 [hep-ex].

- [131] ALEPH, DELPHI, L3, and OPAL Collaborations and LEP Electroweak Working Group. “Electroweak measurements in electron-positron collisions at W boson pair energies at LEP”. *Phys. Rept.* 532 (2013), p. 119. DOI: 10.1016/j.physrep.2013.07.004. arXiv: 1302.3415 [hep-ex].
- [132] D0 Collaboration. “Measurement of the W boson mass with the D0 detector”. *Phys. Rev. D* 89 (2014), p. 012005. DOI: 10.1103/PhysRevD.89.012005. arXiv: 1310.8628 [hep-ex].
- [133] S. Heinemeyer et al. “Precise prediction for m_W in the MSSM”. *JHEP* 08 (2006), p. 052. DOI: 10.1088/1126-6708/2006/08/052. arXiv: hep-ph/0604147.
- [134] IceCube Collaboration. “Cosmic ray spectrum and composition from PeV to EeV using 3 years of data from IceTop and IceCube”. *Phys. Rev. D* 100 (2019), p. 082002. DOI: 10.1103/PhysRevD.100.082002. arXiv: 1906.04317 [astro-ph.HE].
- [135] Pierre Auger Collaboration. “Measurement of the cosmic ray spectrum above 4×10^{18} eV using inclined events detected with the pierre auger observatory”. *JCAP* 08 (2015), p. 049. DOI: 10.1088/1475-7516/2015/08/049. arXiv: 1503.07786 [astro-ph.HE].
- [136] A. Bettini. “Introduction to elementary particle physics”. 2nd ed. Cambridge University Press, 2014. ISBN: 9781107779730.
- [137] T. Schoerner-Sadenius, ed. “The Large Hadron Collider: Harvest of Run 1”. Springer International Publishing, 2015. ISBN: 9783319150000.
- [138] L. Evans and P. Bryant. “LHC machine”. *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/s08001.
- [139] K. Long et al. “Muon colliders to expand frontiers of particle physics”. *Nature Phys.* 17 (2021), p. 289. DOI: 10.1038/s41567-020-01130-x. arXiv: 2007.15684 [physics.acc-ph].
- [140] SLAC-SP-017 Collaboration. “Discovery of a narrow resonance in e^+e^- annihilation”. *Phys. Rev. Lett.* 33 (1974), p. 1406. DOI: 10.1103/PhysRevLett.33.1406.
- [141] M. L. Perl et al. “Evidence for anomalous lepton production in $e^+ - e^-$ annihilation”. *Phys. Rev. Lett.* 35 (1975), p. 1489. DOI: 10.1103/PhysRevLett.35.1489.
- [142] F. Zimmermann. “Future colliders for particle physics —“big and small””. *Nucl. Instrum. Meth. A* 909 (2018), p. 33. DOI: 10.1016/j.nima.2018.01.034. arXiv: 1801.03170 [physics.acc-ph].
- [143] A. Nisati and V. Sharma, eds. “Discovery of the Higgs boson”. World Scientific, 2016.
- [144] D0 Collaboration. “Observation of the top quark”. *Phys. Rev. Lett.* 74 (1995), p. 2632. DOI: 10.1103/PhysRevLett.74.2632. arXiv: hep-ex/9503003.
- [145] CDF Collaboration. “Observation of top quark production in $\bar{p}p$ collisions”. *Phys. Rev. Lett.* 74 (1995), p. 2626. DOI: 10.1103/PhysRevLett.74.2626. arXiv: hep-ex/9503002.
- [146] E. Mobs. “The CERN accelerator complex”. 2016. URL: <http://cds.cern.ch/record/2197559>.
- [147] AC Team. “Diagram of an LHC dipole magnet.” 1999. URL: <https://cds.cern.ch/record/40524>.
- [148] CMS Collaboration. “Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV”. *JHEP* 07 (2018), p. 161. DOI: 10.1007/JHEP07(2018)161. arXiv: 1802.02613 [hep-ex].
- [149] CMS Collaboration. “CMS luminosity public results”. 2020. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [150] M. H. Seymour and M. Marx. “Monte Carlo event generators”. *69th Scottish Universities Summer School in Physics: LHC Physics*. 2013, p. 287. DOI: 10.1007/978-3-319-05362-2_8. arXiv: 1304.6677 [hep-ph].
- [151] T. Gleisberg et al. “Event generation with SHERPA 1.1”. *JHEP* 02 (2009), p. 007. DOI: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622 [hep-ph].

- [152] R. K. Ellis, W. J. Stirling, and B. R. Webber. “QCD and collider physics”. Cambridge University Press, 1996. ISBN: 9780511628788.
- [153] A. Buckley et al. “General-purpose event generators for LHC physics”. *Phys. Rept.* 504 (2011), p. 145. DOI: 10.1016/j.physrep.2011.03.005. arXiv: 1101.2599 [hep-ph].
- [154] P. Nason. “A New method for combining NLO QCD with shower Monte Carlo algorithms”. *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.
- [155] S. Frixione, P. Nason, and C. Oleari. “Matching NLO QCD computations with parton shower simulations: The POWHEG method”. *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].
- [156] T. Melia et al. “ W^+W^- , WZ and ZZ production in the POWHEG BOX”. *JHEP* 11 (2011), p. 078. DOI: 10.1007/JHEP11(2011)078. arXiv: 1107.5051 [hep-ph].
- [157] P. Nason and G. Zanderighi. “ W^+W^- , WZ and ZZ production in the POWHEG-BOX-V2”. *Eur. Phys. J. C* 74 (2014), p. 2702. DOI: 10.1140/epjc/s10052-013-2702-5. arXiv: 1311.1365 [hep-ph].
- [158] J. M. Campbell and R. K. Ellis. “MCFM for the Tevatron and the LHC”. *Nucl. Phys. B Proc. Suppl.* 205 (2010), p. 10. DOI: 10.1016/j.nuclphysbps.2010.08.011. arXiv: 1007.3492 [hep-ph].
- [159] S. Frixione and B. R. Webber. “Matching NLO QCD computations and parton shower simulations”. *JHEP* 06 (2002), p. 029. DOI: 10.1088/1126-6708/2002/06/029. arXiv: hep-ph/0204244.
- [160] 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”. *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].
- [161] T. Sjöstrand et al. “An introduction to PYTHIA 8.2”. *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].
- [162] P. Skands, S. Carrazza, and J. Rojo. “Tuning PYTHIA 8.1: the Monash 2013 tune”. *Eur. Phys. J. C* 74 (2014), p. 3024. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: 1404.5630 [hep-ph].
- [163] CMS Collaboration. “Event generator tunes obtained from underlying event and multiparton scattering measurements”. *Eur. Phys. J. C* 76 (2016), p. 155. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex].
- [164] CMS Collaboration. “Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements”. *Eur. Phys. J. C* 80 (2020), p. 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].
- [165] R. Frederix and S. Frixione. “Merging meets matching in MC@NLO”. *JHEP* 12 (2012), p. 061. DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215 [hep-ph].
- [166] J. Alwall et al. “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”. *Eur. Phys. J. C* 53 (2008), p. 473. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph].
- [167] NNPDF Collaboration. “Parton distributions for the LHC Run II”. *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].
- [168] NNPDF Collaboration. “Parton distributions from high-precision collider data”. *Eur. Phys. J. C* 77 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].
- [169] CMS Collaboration. “The CMS Experiment at the CERN LHC”. *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.
- [170] CMS Collaboration. “Detector drawings”. 2012. URL: <http://cds.cern.ch/record/1433717>.

- [171] K. Eskola et al. “Rapidity dependence of particle production in ultrarelativistic nuclear collisions”. *Phys. Lett. B* 543 (2002), p. 208. DOI: [https://doi.org/10.1016/S0370-2693\(02\)02457-7](https://doi.org/10.1016/S0370-2693(02)02457-7).
- [172] CMS Collaboration. “Precision measurement of the structure of the CMS inner tracking system using nuclear interactions”. *JINST* 13 (2018), P10034. DOI: [10.1088/1748-0221/13/10/P10034](https://doi.org/10.1088/1748-0221/13/10/P10034). arXiv: 1807.03289 [physics.ins-det].
- [173] CMS Collaboration. “CMS tracker detector performance results”. 2020. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK>.
- [174] CMS Collaboration. “Technical proposal for the upgrade of the CMS detector through 2020”. Tech. rep. CERN-LHCC-2011-006. LHCC-P-004. 2011. URL: <https://cds.cern.ch/record/1355706>.
- [175] T. A. Vami. “Calibration and performance of the CMS pixel detector in LHC Run 2”. *PoS LHCP2019* (2019). DOI: [10.22323/1.350.0010](https://doi.org/10.22323/1.350.0010). arXiv: 1909.12920 [physics.ins-det].
- [176] CMS Collaboration. “Commissioning and performance of the CMS silicon strip tracker with cosmic ray muons”. *JINST* 5 (2010), T03008. DOI: [10.1088/1748-0221/5/03/t03008](https://doi.org/10.1088/1748-0221/5/03/t03008).
- [177] CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. *JINST* 9 (2014), P10009. DOI: [10.1088/1748-0221/9/10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009). arXiv: 1405.6569 [physics.ins-det].
- [178] CMS Collaboration. “CMS physics: Technical design report volume 1: Detector performance and software”. Tech. rep. CERN-LHCC-2006-001. 2006. URL: <http://cds.cern.ch/record/922757>.
- [179] CMS Collaboration. “Technical proposal for the Phase-II upgrade of the CMS detector”. Tech. rep. CERN-LHCC-2015-010. 2015. URL: <https://cds.cern.ch/record/2020886>.
- [180] CMS Collaboration. “ECAL 2016 refined calibration and Run 2 summary plots”. Tech. rep. CMS-DP-2020-021. 2020. URL: <https://cds.cern.ch/record/2717925>.
- [181] CMS Collaboration. “A measurement of the Higgs boson mass in the diphoton decay channel”. *Phys. Lett. B* 805 (2020), p. 135425. DOI: [10.1016/j.physletb.2020.135425](https://doi.org/10.1016/j.physletb.2020.135425). arXiv: 2002.06398 [hep-ex].
- [182] CMS Collaboration. “Particle-flow reconstruction and global event description with the CMS detector”. *JINST* 12 (2017), P10003. DOI: [10.1088/1748-0221/12/10/P10003](https://doi.org/10.1088/1748-0221/12/10/P10003). arXiv: 1706.04965 [physics.ins-det].
- [183] CMS Collaboration. “Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV”. *JINST* 15 (2020), P05002. DOI: [10.1088/1748-0221/15/05/P05002](https://doi.org/10.1088/1748-0221/15/05/P05002). arXiv: 1910.00079 [physics.ins-det].
- [184] CMS Collaboration. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JINST* 13 (2018), P06015. DOI: [10.1088/1748-0221/13/06/P06015](https://doi.org/10.1088/1748-0221/13/06/P06015). arXiv: 1804.04528 [physics.ins-det].
- [185] CMS Collaboration. “The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC”. *JINST* 8 (2013), P11002. DOI: [10.1088/1748-0221/8/11/P11002](https://doi.org/10.1088/1748-0221/8/11/P11002). arXiv: 1306.6905 [physics.ins-det].
- [186] M. Abbrescia, V. Peskov, and P. Fonte. “Resistive gaseous detectors: Designs, performance, and perspectives”. John Wiley & Sons, Ltd, 2018. ISBN: 9783527340767.
- [187] CMS Collaboration. “The CMS trigger system”. *JINST* 12 (2017), P01020. DOI: [10.1088/1748-0221/12/01/P01020](https://doi.org/10.1088/1748-0221/12/01/P01020). arXiv: 1609.02366 [physics.ins-det].
- [188] CMS Collaboration. “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JINST* 15 (2020), P10017. DOI: [10.1088/1748-0221/15/10/P10017](https://doi.org/10.1088/1748-0221/15/10/P10017). arXiv: 2006.10165 [hep-ex].

- [189] GEANT4 Collaboration. “GEANT4—a simulation toolkit”. *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.
- [190] ALEPH Collaboration. “Performance of the ALEPH detector at LEP”. *Nucl. Instrum. Meth. A* 360 (1995), p. 481. DOI: 10.1016/0168-9002(95)00138-7.
- [191] D. Barney. “CMS detector slice”. 2016. URL: <https://cds.cern.ch/record/2120661>.
- [192] M. Swartz et al. “A new technique for the reconstruction, validation, and simulation of hits in the CMS pixel detector”. *PoS VERTEX2007* (2007), p. 035. DOI: 10.22323/1.057.0035.
- [193] R. E. Kalman. “A new approach to linear filtering and prediction problems”. *Journal of Basic Engineering* 82 (1960), p. 35. DOI: 10.1115/1.3662552.
- [194] R. Fruhwirth. “Application of Kalman filtering to track and vertex fitting”. *Nucl. Instrum. Meth. A* 262 (1987), p. 444. DOI: 10.1016/0168-9002(87)90887-4.
- [195] P. Billoir. “Progressive track recognition with a Kalman-like fitting procedure”. *Comput. Phys. Commun.* 57 (1989), p. 390. DOI: 10.1016/0010-4655(89)90249-X.
- [196] P. Billoir and S. Qian. “Simultaneous pattern recognition and track fitting by the Kalman filtering method”. *Nucl. Instrum. Meth. A* 294 (1990), p. 219. DOI: 10.1016/0168-9002(90)91835-Y.
- [197] C. A. Pruneau. “Data analysis techniques for physical scientists”. Cambridge University Press, 2017. ISBN: 9781108241922.
- [198] X. Ai et al. “A GPU-based Kalman filter for track fitting”. *Comput. Softw. Big Sci.* 5 (2021), p. 20. DOI: 10.1007/s41781-021-00065-z. arXiv: 2105.01796 [physics.ins-det].
- [199] CMS Collaboration. “CMS tracking performance results from early LHC operation”. *Eur. Phys. J. C* 70 (2010), p. 1165. DOI: 10.1140/epjc/s10052-010-1491-3. arXiv: 1007.1988 [physics.ins-det].
- [200] W. Adam et al. “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”. *J. Phys. G* 31 (2005), N9. DOI: 10.1088/0954-3899/31/9/n01.
- [201] H. Bethe and W. Heitler. “On the stopping of fast particles and on the creation of positive electrons”. *Proc. Roy. Soc. Lond. A* 146 (1934), p. 83. DOI: 10.1098/rspa.1934.0140.
- [202] CMS Collaboration. “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”. *JINST* 10 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det].
- [203] CMS Collaboration. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”. *JINST* 16 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].
- [204] K. Rose. “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”. *IEEE Proc.* 86 (1998), p. 2210. DOI: 10.1109/5.726788.
- [205] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. “Adaptive vertex fitting”. Tech. rep. CMS-NOTE-2007-008. 2007, N343. DOI: 10.1088/0954-3899/34/12/N01.
- [206] T. Miao et al. “Beam position determination using tracks”. Tech. rep. CMS-NOTE-2007-021, FERMILAB-FN-0816-E. 2007. URL: <http://cds.cern.ch/record/1061285>.
- [207] J. L. Bentley. “Multidimensional binary search trees used for associative searching”. *Commun. ACM* 18 (1975), p. 509. DOI: 10.1145/361002.361007.
- [208] CMS Collaboration. “Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”. *JINST* 10 (2015), P08010. DOI: 10.1088/1748-0221/10/08/P08010. arXiv: 1502.02702 [physics.ins-det].
- [209] M. Cacciari, G. P. Salam, and G. Soyez. “The anti- k_t jet clustering algorithm”. *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

- [210] M. Cacciari, G. P. Salam, and G. Soyez. “FastJet user manual”. *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].
- [211] T. H. Cormen et al. “Introduction to algorithms”. 3rd ed. MIT Press, 2009. ISBN: 9780262033848.
- [212] M. Cacciari and G. P. Salam. “Dispelling the N^3 myth for the k_t jet-finder”. *Phys. Lett. B* 641 (2006), p. 57. DOI: 10.1016/j.physletb.2006.08.037. arXiv: hep-ph/0512210.
- [213] CMS Collaboration. “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”. *JINST* 12 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].
- [214] CMS Collaboration. “Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018”. Tech. rep. CMS-DP-2020-019. 2020. URL: <https://cds.cern.ch/record/2715872>.
- [215] M. Cacciari and G. P. Salam. “Pileup subtraction using jet areas”. *Phys. Lett. B* 659 (2008), p. 119. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph].
- [216] CMS Collaboration. “Pileup mitigation at CMS in 13 TeV data”. *JINST* 15 (2020), P09018. DOI: 10.1088/1748-0221/15/09/P09018. arXiv: 2003.00503 [hep-ex].
- [217] CMS Collaboration. “Determination of jet energy calibration and transverse momentum resolution in CMS”. *JINST* 6 (2011), P11002. DOI: 10.1088/1748-0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det].
- [218] CMS Collaboration. “Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector”. *JINST* 14 (2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].
- [219] CMS Collaboration. “Performance of missing transverse momentum reconstruction in events containing a photon and jets collected by CMS during proton-proton collisions at $\sqrt{s} = 13$ TeV in 2018”. Tech. rep. CMS-DP-2020-031. 2020. URL: <https://cds.cern.ch/record/2723010>.
- [220] CMS Collaboration. “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”. *JINST* 13 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].
- [221] CMS Collaboration. “Measurement of $B\bar{B}$ angular correlations based on secondary vertex reconstruction at $\sqrt{s} = 7$ TeV”. *JHEP* 03 (2011), p. 136. DOI: 10.1007/JHEP03(2011)136. arXiv: 1102.3194 [hep-ex].
- [222] W. Waltenberger. “Adaptive vertex reconstruction”. Tech. rep. CMS-NOTE-2008-033. CERN, 2008. URL: <https://cds.cern.ch/record/1166320>.
- [223] CMS Collaboration. “Identification of b-quark jets with the CMS experiment”. *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].
- [224] D. Guest, K. Cranmer, and D. Whiteson. “Deep learning and its application to LHC physics”. *Ann. Rev. Nucl. Part. Sci.* 68 (2018), p. 161. DOI: 10.1146/annurev-nucl-101917-021019. arXiv: 1806.11484 [hep-ex].
- [225] A. Butter et al. “The machine learning landscape of top taggers”. *SciPost Phys.* 7 (2019). Ed. by G. Kasieczka and T. Plehn, p. 014. DOI: 10.21468/SciPostPhys.7.1.014. arXiv: 1902.09914 [hep-ph].
- [226] D. Guest et al. “Jet flavor classification in high-energy physics with deep neural networks”. *Phys. Rev. D* 94 (2016), p. 112002. DOI: 10.1103/PhysRevD.94.112002. arXiv: 1607.08633 [hep-ex].
- [227] Z. Li et al. “A survey of convolutional neural networks: Analysis, applications, and prospects”. *IEEE Transactions on Neural Networks and Learning Systems* (2020), p. 1. DOI: 10.1109/TNNLS.2021.3084827. arXiv: 2004.02806 [cs.CV].

- [228] K. Greff et al. “LSTM: A search space odyssey”. *IEEE Transactions on Neural Networks and Learning Systems* 28 (2017), p. 2222. DOI: 10.1109/tnnls.2016.2582924. arXiv: 1503.04069 [cs.NE].
- [229] E. Bols et al. “Jet flavour classification using DeepJet”. *JINST* 15 (2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519 [hep-ex].
- [230] CMS Collaboration. “Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector”. Tech. rep. CMS-DP-2018-058. 2018. URL: <http://cds.cern.ch/record/2646773>.
- [231] CMS Collaboration. “Measurement of associated W + charm production in pp collisions at $\sqrt{s} = 7$ TeV”. *JHEP* 02 (2014), p. 013. DOI: 10.1007/JHEP02(2014)013. arXiv: 1310.1138 [hep-ex].
- [232] CMS Collaboration. “Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV”. *JINST* 13 (2018), P10005. DOI: 10.1088/1748-0221/13/10/P10005. arXiv: 1809.02816 [hep-ex].
- [233] CMS Collaboration. “Performance of τ -lepton reconstruction and identification in CMS”. *JINST* 7 (2012), P01001. DOI: 10.1088/1748-0221/7/01/p01001.
- [234] CMS Collaboration. “Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS”. *JINST* 11 (2016), P01019. DOI: 10.1088/1748-0221/11/01/P01019. arXiv: 1510.07488 [physics.ins-det].
- [235] CMS Collaboration. “Performance of CMS muon reconstruction in cosmic-ray events”. *JINST* 5 (2010), T03022. DOI: 10.1088/1748-0221/5/03/t03022.
- [236] K. Rehermann and B. Tweedie. “Efficient identification of boosted semileptonic top quarks at the LHC”. *JHEP* 03 (2011), p. 059. DOI: 10.1007/JHEP03(2011)059. arXiv: 1007.2221 [hep-ph].
- [237] CMS Collaboration. “Search for new physics in same-sign dilepton events in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *Eur. Phys. J. C* 76 (2016), p. 439. DOI: 10.1140/epjc/s10052-016-4261-z. arXiv: 1605.03171 [hep-ex].
- [238] CMS Collaboration. “Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s} = 13$ TeV”. *JHEP* 08 (2018), p. 066. DOI: 10.1007/JHEP08(2018)066. arXiv: 1803.05485 [hep-ex].
- [239] CMS Collaboration. “Observation of $t\bar{t}H$ production”. *Phys. Rev. Lett.* 120 (2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801. arXiv: 1804.02610 [hep-ex].
- [240] CMS Collaboration. “Evidence for Higgs boson decay to a pair of muons”. *JHEP* 01 (2021), p. 148. DOI: 10.1007/JHEP01(2021)148. arXiv: 2009.04363 [hep-ex].
- [241] CMS Collaboration. “Summaries of CMS cross section measurements”. 2020. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined>.
- [242] CMS Collaboration. “Measurement of the WZ production cross section in pp collisions at $\sqrt{s} = 7$ and 8 TeV and search for anomalous triple gauge couplings at $\sqrt{s} = 8$ TeV”. *Eur. Phys. J. C* 77 (2017), p. 236. DOI: 10.1140/epjc/s10052-017-4730-z. arXiv: 1609.05721 [hep-ex].
- [243] ATLAS Collaboration. “Measurements of $Z\gamma$ and $Z\gamma\gamma$ production in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”. *Phys. Rev. D* 93 (2016), p. 112002. DOI: 10.1103/PhysRevD.93.112002. arXiv: 1604.05232 [hep-ex].
- [244] CMS Collaboration. “Measurement of the associated production of a single top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV”. *Phys. Lett. B* 779 (2018), p. 358. DOI: 10.1016/j.physletb.2018.02.025. arXiv: 1712.02825 [hep-ex].

- [245] CMS Collaboration. “Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JHEP* 02 (2022), p. 107. DOI: 10.1007/JHEP02(2022)107. arXiv: 2111.02860 [hep-ex].
- [246] CMS Collaboration. “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV”. *Eur. Phys. J. C* 75 (2015), p. 212. DOI: 10.1140/epjc/s10052-015-3351-7. arXiv: 1412.8662 [hep-ex].
- [247] K. Cranmer. “Practical statistics for the LHC”. *2011 European School of High-Energy Physics*. 2014, p. 267. DOI: 10.5170/CERN-2014-003.267. arXiv: 1503.07622 [physics.data-an].
- [248] CMS Collaboration. “CMS luminosity measurements for the 2016 data-taking period”. Tech. rep. CMS-PAS-LUM-17-001. 2017. URL: <https://cds.cern.ch/record/2257069>.
- [249] CMS Collaboration. “CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV”. Tech. rep. CMS-PAS-LUM-18-002. 2019. URL: <https://cds.cern.ch/record/2676164>.
- [250] CMS Collaboration. “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV”. Tech. rep. CMS-PAS-LUM-17-004. 2018. URL: <https://cds.cern.ch/record/2621960>.
- [251] J. Butterworth et al. “PDF4LHC recommendations for LHC Run II”. *J. Phys. G* 43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].
- [252] G. Bellini et al. “Neutrino oscillations”. *Adv. High Energy Phys.* (2014), p. 191960. DOI: 10.1155/2014/191960. arXiv: 1310.7858 [hep-ph].
- [253] E. K. Akhmedov, V. A. Rubakov, and A. Y. Smirnov. “Baryogenesis via neutrino oscillations”. *Phys. Rev. Lett.* 81 (1998), p. 1359. DOI: 10.1103/PhysRevLett.81.1359. arXiv: hep-ph/9803255 [hep-ph].
- [254] L. Canetti et al. “Dark matter, baryogenesis and neutrino oscillations from right-handed neutrinos”. *Phys. Rev. D* 87 (2013), p. 093006. DOI: 10.1103/PhysRevD.87.093006. arXiv: 1208.4607 [hep-ph].
- [255] M. Fukugita and T. Yanagida. “Baryogenesis without grand unification”. *Phys. Lett. B* 174 (1986), p. 45. DOI: 10.1016/0370-2693(86)91126-3.
- [256] L. Canetti, M. Drewes, and B. Garbrecht. “Probing leptogenesis with GeV-scale sterile neutrinos at LHCb and Belle II”. *Phys. Rev. D* 90 (2014), p. 125005. DOI: 10.1103/PhysRevD.90.125005. arXiv: 1404.7114 [hep-ph].
- [257] S. Antusch et al. “Probing Leptogenesis at Future Colliders”. *JHEP* 09 (2018), p. 124. DOI: 10.1007/JHEP09(2018)124. arXiv: 1710.03744 [hep-ph].
- [258] CHARM Collaboration. “A search for decays of heavy neutrinos in the mass range 0.5 GeV to 2.8 GeV”. *Phys. Lett. B* 166 (1986), p. 473. DOI: 10.1016/0370-2693(86)91601-1.
- [259] NuTeV and E815 Collaborations. “Search for neutral heavy leptons in a high-energy neutrino beam”. *Phys. Rev. Lett.* 83 (1999), p. 4943. DOI: 10.1103/PhysRevLett.83.4943. arXiv: hep-ex/9908011.
- [260] NA62 Collaboration. “Search for heavy neutrinos in $K^+ \rightarrow \mu^+ \nu_\mu$ decays”. *Phys. Lett. B* 772 (2017), p. 712. DOI: 10.1016/j.physletb.2017.07.055. arXiv: 1705.07510 [hep-ex].
- [261] S. Alekhin et al. “A facility to search for hidden particles at the CERN SPS: the SHiP physics case”. *Rept. Prog. Phys.* 79 (2016), p. 124201. DOI: 10.1088/0034-4885/79/12/124201. arXiv: 1504.04855 [hep-ph].
- [262] Belle Collaboration. “Search for heavy neutrinos at Belle”. *Phys. Rev. D* 87 (2013). [Erratum: *Phys.Rev.D* 95, 099903 (2017)], p. 071102. DOI: 10.1103/PhysRevD.87.071102. arXiv: 1301.1105 [hep-ex].

- [263] L3 Collaboration. “Search for isosinglet neutral heavy leptons in Z^0 decays”. *Phys. Lett. B* 295 (1992), p. 371. DOI: 10.1016/0370-2693(92)91579-X.
- [264] L3 Collaboration. “Search for heavy isosinglet neutrinos in e^+e^- annihilation at $130 < \sqrt{s} < 189$ GeV”. *Phys. Lett. B* 461 (1999), p. 397. DOI: 10.1016/S0370-2693(99)00852-7. arXiv: hep-ex/9909006.
- [265] L3 Collaboration. “Search for heavy isosinglet neutrino in e^+e^- annihilation at LEP”. *Phys. Lett. B* 517 (2001), p. 67. DOI: 10.1016/S0370-2693(01)00993-5. arXiv: hep-ex/0107014.
- [266] DELPHI Collaboration. “Search for neutral heavy leptons produced in Z decays”. *Z. Phys. C* 74 (1997). [Erratum: Z.Phys.C 75, 580 (1997)], p. 57. DOI: 10.1007/s002880050370.
- [267] ATLAS Collaboration. “Inclusive search for same-sign dilepton signatures in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”. *JHEP* 10 (2011), p. 107. DOI: 10.1007/JHEP10(2011)107. arXiv: 1108.0366 [hep-ex].
- [268] ATLAS Collaboration. “Search for heavy neutrinos and right-handed W bosons in events with two leptons and jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”. *Eur. Phys. J. C* 72 (2012), p. 2056. DOI: 10.1140/epjc/s10052-012-2056-4. arXiv: 1203.5420 [hep-ex].
- [269] ATLAS Collaboration. “Search for heavy Majorana neutrinos with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV”. *JHEP* 07 (2015), p. 162. DOI: 10.1007/JHEP07(2015)162. arXiv: 1506.06020 [hep-ex].
- [270] CMS Collaboration. “Search for heavy Majorana neutrinos in $\mu^\pm\mu^\pm$ + jets and $e^\pm e^\pm$ + jets events in pp collisions at $\sqrt{s} = 7$ TeV”. *Phys. Lett. B* 717 (2012), p. 109. DOI: 10.1016/j.physletb.2012.09.012. arXiv: 1207.6079 [hep-ex].
- [271] CMS Collaboration. “Search for heavy neutrinos and W_R bosons with right-handed couplings in a left-right symmetric model in pp collisions at $\sqrt{s} = 7$ TeV”. *Phys. Rev. Lett.* 109 (2012), p. 261802. DOI: 10.1103/PhysRevLett.109.261802. arXiv: 1210.2402 [hep-ex].
- [272] CMS Collaboration. “Search for heavy neutrinos and W bosons with right-handed couplings in proton-proton collisions at $\sqrt{s} = 8$ TeV”. *Eur. Phys. J. C* 74 (2014), p. 3149. DOI: 10.1140/epjc/s10052-014-3149-z. arXiv: 1407.3683 [hep-ex].
- [273] CMS Collaboration. “Search for heavy Majorana neutrinos in $\mu^\pm\mu^\pm$ + jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV”. *Phys. Lett. B* 748 (2015), p. 144. DOI: 10.1016/j.physletb.2015.06.070. arXiv: 1501.05566 [hep-ex].
- [274] CMS Collaboration. “Search for heavy Majorana neutrinos in $e^\pm e^\pm$ + jets and $e^\pm\mu^\pm$ + jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV”. *JHEP* 04 (2016), p. 169. DOI: 10.1007/JHEP04(2016)169. arXiv: 1603.02248 [hep-ex].
- [275] F. F. Deppisch, P. S. Bhupal Dev, and A. Pilaftsis. “Neutrinos and collider physics”. *New J. Phys.* 17 (2015), p. 075019. DOI: 10.1088/1367-2630/17/7/075019. arXiv: 1502.06541 [hep-ph].
- [276] ATLAS Collaboration. “Search for heavy neutral leptons in decays of W bosons produced in 13 TeV pp collisions using prompt and displaced signatures with the ATLAS detector”. *JHEP* 10 (2019), p. 265. DOI: 10.1007/JHEP10(2019)265. arXiv: 1905.09787 [hep-ex].
- [277] CMS Collaboration. “Search for long-lived heavy neutral leptons with displaced vertices in proton-proton collisions at $\sqrt{s} = 13$ TeV” (2022). arXiv: 2201.05578 [hep-ex].
- [278] CMS Collaboration. “Search for heavy Majorana neutrinos in same-sign dilepton channels in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JHEP* 01 (2019), p. 122. DOI: 10.1007/JHEP01(2019)122. arXiv: 1806.10905 [hep-ex].
- [279] LHCb Collaboration. “Search for heavy neutral leptons in $W^+ \rightarrow \mu^+\mu^\pm$ jet decays”. *Eur. Phys. J. C* 81 (2021), p. 248. DOI: 10.1140/epjc/s10052-021-08973-5. arXiv: 2011.05263 [hep-ex].

- [280] A. Atre et al. “The search for heavy Majorana neutrinos”. *JHEP* 05 (2009), p. 030. DOI: 10.1088/1126-6708/2009/05/030. arXiv: 0901.3589 [hep-ph].
- [281] C. Degrande et al. “UFO - The Universal FeynRules Output”. *Comput. Phys. Commun.* 183 (2012), p. 1201. DOI: 10.1016/j.cpc.2012.01.022. arXiv: 1108.2040 [hep-ph].
- [282] D. Alva, T. Han, and R. Ruiz. “Heavy Majorana neutrinos from $W\gamma$ fusion at hadron colliders”. *JHEP* 02 (2015), p. 072. DOI: 10.1007/JHEP02(2015)072. arXiv: 1411.7305 [hep-ph].
- [283] C. Degrande et al. “Fully-automated precision predictions for heavy neutrino production mechanisms at hadron colliders”. *Phys. Rev. D* 94 (2016), p. 053002. DOI: 10.1103/PhysRevD.94.053002. arXiv: 1602.06957 [hep-ph].
- [284] S. Pascoli, R. Ruiz, and C. Weiland. “Heavy neutrinos with dynamic jet vetoes: multilepton searches at $\sqrt{s} = 14, 27$, and 100 TeV”. *JHEP* 06 (2019), p. 049. DOI: 10.1007/JHEP06(2019)049. arXiv: 1812.08750 [hep-ph].
- [285] T. Sjöstrand, S. Mrenna, and P. Z. Skands. “A brief introduction to PYTHIA 8.1”. *Comput. Phys. Commun.* 178 (2008), p. 852. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].
- [286] A. Manohar et al. “How bright is the proton? A precise determination of the photon parton distribution function”. *Phys. Rev. Lett.* 117 (2016), p. 242002. DOI: 10.1103/PhysRevLett.117.242002. arXiv: 1607.04266 [hep-ph].
- [287] T. Junk. “Confidence level computation for combining searches with small statistics”. *Nucl. Instrum. Meth. A* 434 (1999), p. 435. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006.
- [288] A. L. Read. “Presentation of search results: the CL_s technique”. *J. Phys. G* 28 (2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313.
- [289] J. S. Conway. “Incorporating nuisance parameters in likelihoods for multisource spectra”. *Proceedings, workshop on statistical issues related to discovery claims in search experiments and unfolding (PHYSTAT 2011)*. 2011, p. 115. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354 [physics.data-an].
- [290] R. J. Barlow and C. Beeston. “Fitting using finite Monte Carlo samples”. *Comput. Phys. Commun.* 77 (1993), p. 219. DOI: 10.1016/0010-4655(93)90005-W.
- [291] G. Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. *Eur. Phys. J. C* 71 (2011). [Erratum: doi:10.1140/epjc/s10052-013-2501-z], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].
- [292] ATLAS and CMS Collaborations, and the LHC Higgs Combination Group. “Procedure for the LHC Higgs boson search combination in summer 2011”. Tech. rep. CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11. 2011. URL: <https://cds.cern.ch/record/1379837>.
- [293] C. O. Dib and C. S. Kim. “Discovering sterile neutrinos lighter than m_W at the LHC”. *Phys. Rev. D* 92 (2015), p. 093009. DOI: 10.1103/PhysRevD.92.093009. arXiv: 1509.05981 [hep-ph].
- [294] S. Dube, D. Gadkari, and A. M. Thalapillil. “Lepton-jets and low-mass sterile neutrinos at hadron colliders”. *Phys. Rev. D* 96 (2017), p. 055031. DOI: 10.1103/PhysRevD.96.055031. arXiv: 1707.00008 [hep-ph].
- [295] J. Ellis et al. “Top, Higgs, Diboson and electroweak fit to the standard model effective field theory”. *JHEP* 04 (2021), p. 279. DOI: 10.1007/JHEP04(2021)279. arXiv: 2012.02779 [hep-ph].
- [296] CMS Collaboration. “Using associated top quark production to probe for new physics within the framework of effective field theory”. Tech. rep. CMS-PAS-TOP-19-001. CERN, 2020. URL: <http://cds.cern.ch/record/2725399>.

- [297] F. Maltoni, L. Mantani, and K. Mimasu. “Top-quark electroweak interactions at high energy”. *JHEP* 10 (2019), p. 004. DOI: 10.1007/JHEP10(2019)004. arXiv: 1904.05637 [hep-ph].
- [298] J. Campbell, R. K. Ellis, and R. Röntsch. “Single top production in association with a Z boson at the LHC”. *Phys. Rev. D* 87 (2013), p. 114006. DOI: 10.1103/PhysRevD.87.114006. arXiv: 1302.3856 [hep-ph].
- [299] C. Degrande et al. “Single-top associated production with a Z or H boson at the LHC: the SMEFT interpretation”. *JHEP* 10 (2018), p. 005. DOI: 10.1007/JHEP10(2018)005. arXiv: 1804.07773 [hep-ph].
- [300] J. A. Aguilar-Saavedra. “Top flavor-changing neutral interactions: Theoretical expectations and experimental detection”. *Acta Phys. Polon. B* 35 (2004). Ed. by F. del Aguila et al., p. 2695. arXiv: hep-ph/0409342.
- [301] J.-L. Agram et al. “Probing top anomalous couplings at the LHC with trilepton signatures in the single top mode”. *Phys. Lett. B* 725 (2013), p. 123. DOI: 10.1016/j.physletb.2013.06.052. arXiv: 1304.5551 [hep-ph].
- [302] Y. Afik et al. “New flavor physics in di- and tri-lepton events from single-top at the LHC and beyond” (2021). arXiv: 2101.05286 [hep-ph].
- [303] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak interactions with lepton-hadron symmetry”. *Phys. Rev. D* 2 (1970), p. 1285. DOI: 10.1103/PhysRevD.2.1285.
- [304] D. Pagani, I. Tsiniikos, and E. Vryonidou. “NLO QCD+EW predictions for tHj and tZj production at the LHC”. *JHEP* 08 (2020), p. 082. DOI: 10.1007/JHEP08(2020)082. arXiv: 2006.10086 [hep-ph].
- [305] LHC Higgs Cross Section Working Group. “Handbook of LHC Higgs cross sections: 1. Inclusive observables”. Tech. rep. CERN-2011-002. 2011. DOI: 10.5170/CERN-2011-002. arXiv: 1101.0593 [hep-ph].
- [306] CMS Collaboration. “Search for associated production of a Z boson with a single top quark and for tZ flavour-changing interactions in pp collisions at $\sqrt{s} = 8$ TeV”. *JHEP* 07 (2017), p. 003. DOI: 10.1007/JHEP07(2017)003. arXiv: 1702.01404 [hep-ex].
- [307] ATLAS Collaboration. “Measurement of the production cross-section of a single top quark in association with a Z boson in proton-proton collisions at 13 TeV with the ATLAS detector”. *Phys. Lett. B* 780 (2018), p. 557. DOI: 10.1016/j.physletb.2018.03.023. arXiv: 1710.03659 [hep-ex].
- [308] ATLAS Collaboration. “Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *JHEP* 07 (2020), p. 124. DOI: 10.1007/JHEP07(2020)124. arXiv: 2002.07546 [hep-ex].
- [309] J. H. Friedman. “Greedy function approximation: A gradient boosting machine.” *The Annals of Statistics* 29 (2001), p. 1189. DOI: 10.1214/aos/1013203451. URL: <https://doi.org/10.1214/aos/1013203451>.
- [310] D. Sieling. “Minimization of decision trees is hard to approximate”. *Journal of Computer and System Sciences* 74 (2008). Computational Complexity 2003, p. 394. DOI: <https://doi.org/10.1016/j.jcss.2007.06.014>.
- [311] C. Zhang et al. “Understanding deep learning requires rethinking generalization”. *Communications of the ACM* 64 (2016), p. 107. DOI: 10.1145/3446776.
- [312] H. Voss et al. “TMVA, the toolkit for multivariate data analysis with ROOT”. *PoS ACAT* (2007), p. 40. DOI: 10.22323/1.050.0040. arXiv: physics/0703039 [physics.data-an].
- [313] T. Chen and C. Guestrin. “XGBoost: A scalable tree boosting system”. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016). DOI: 10.1145/2939672.2939785. arXiv: 1603.02754 [cs.LG].

- [314] M. Abadi et al. “TensorFlow: Large-scale machine learning on heterogeneous systems”. Software available from <https://www.tensorflow.org/>. 2016.
- [315] F. Chollet et al. “Keras”. Software available from <https://keras.io>. 2015.
- [316] CMS Collaboration. “Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JHEP* 03 (2020), p. 056. DOI: 10.1007/JHEP03(2020)056. arXiv: 1907.11270 [hep-ex].
- [317] P. Ramond. “Dual theory for free fermions”. *Phys. Rev. D* 3 (1971), p. 2415. DOI: 10.1103/PhysRevD.3.2415.
- [318] J. Wess and B. Zumino. “A lagrangian model invariant under supergauge transformations”. *Phys. Lett. B* 49 (1974), p. 52. DOI: [https://doi.org/10.1016/0370-2693\(74\)90578-4](https://doi.org/10.1016/0370-2693(74)90578-4).
- [319] J. Wess and B. Zumino. “Supergauge transformations in four dimensions”. *Nucl. Phys. B* 70 (1974), p. 39. DOI: [https://doi.org/10.1016/0550-3213\(74\)90355-1](https://doi.org/10.1016/0550-3213(74)90355-1).
- [320] P. Fayet. “Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino”. *Nucl. Phys. B* 90 (1975), p. 104. DOI: [https://doi.org/10.1016/0550-3213\(75\)90636-7](https://doi.org/10.1016/0550-3213(75)90636-7).
- [321] H. Nilles. “Supersymmetry, supergravity and particle physics”. *Phys. Rep.* 110 (1984), p. 1. DOI: [https://doi.org/10.1016/0370-1573\(84\)90008-5](https://doi.org/10.1016/0370-1573(84)90008-5).
- [322] ATLAS Collaboration. “Search for top squarks in events with a Higgs or Z boson using 139 fb⁻¹ of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *Eur. Phys. J. C* 80 (2020), p. 1080. DOI: 10.1140/epjc/s10052-020-08469-8. arXiv: 2006.05880 [hep-ex].
- [323] ATLAS Collaboration. “Search for a scalar partner of the top quark in the all-hadronic $t\bar{t}$ plus missing transverse momentum final state at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *Eur. Phys. J. C* 80 (2020), p. 737. DOI: 10.1140/epjc/s10052-020-8102-8. arXiv: 2004.14060 [hep-ex].
- [324] ATLAS Collaboration. “Search for squarks and gluinos in final states with same-sign leptons and jets using 139 fb⁻¹ of data collected with the ATLAS detector”. *JHEP* 06 (2020), p. 046. DOI: 10.1007/JHEP06(2020)046. arXiv: 1909.08457 [hep-ex].
- [325] ATLAS Collaboration. “Search for bottom-squark pair production with the ATLAS detector in final states containing Higgs bosons, b-jets and missing transverse momentum”. *JHEP* 12 (2019), p. 060. DOI: 10.1007/JHEP12(2019)060. arXiv: 1908.03122 [hep-ex].
- [326] CMS Collaboration. “Search for supersymmetry in proton-proton collisions at 13 TeV in final states with jets and missing transverse momentum”. *JHEP* 10 (2019), p. 244. DOI: 10.1007/JHEP10(2019)244. arXiv: 1908.04722 [hep-ex].
- [327] CMS Collaboration. “Searches for physics beyond the standard model with the M_{T2} variable in hadronic final states with and without disappearing tracks in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *Eur. Phys. J. C* 80 (2020), p. 3. DOI: 10.1140/epjc/s10052-019-7493-x. arXiv: 1909.03460 [hep-ex].
- [328] CMS Collaboration. “Search for supersymmetry in pp collisions at $\sqrt{s} = 13$ TeV with 137 fb⁻¹ in final states with a single lepton using the sum of masses of large-radius jets”. *Phys. Rev. D* 101 (2020), p. 052010. DOI: 10.1103/PhysRevD.101.052010. arXiv: 1911.07558 [hep-ex].
- [329] 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”. *JHEP* 05 (2020), p. 032. DOI: 10.1007/JHEP05(2020)032. arXiv: 1912.08887 [hep-ex].
- [330] CMS Collaboration. “Search for physics beyond the standard model in events with jets and two same-sign or at least three charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *Eur. Phys. J. C* 80 (2020), p. 752. DOI: 10.1140/epjc/s10052-020-8168-3. arXiv: 2001.10086 [hep-ex].

- [331] LHC SUSY Cross Section Working Group. “SUSY cross sections for 13, 14, 33 and 100 TeV pp collisions”. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections>. 2020.
- [332] B. Fuks et al. “Precision predictions for electroweak superpartner production at hadron colliders with Resummino”. *Eur. Phys. J. C* 73 (2013), p. 2480. DOI: 10.1140/epjc/s10052-013-2480-0. arXiv: 1304.0790 [hep-ph].
- [333] C. Borschensky et al. “Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13, 14, 33$ and 100 TeV”. *Eur. Phys. J. C* 74 (2014), p. 3174. DOI: 10.1140/epjc/s10052-014-3174-y. arXiv: 1407.5066 [hep-ph].
- [334] CMS Collaboration. “CMS supersymmetry physics results”. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>. 2021.
- [335] ATLAS Collaboration. “Supersymmetry searches”. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults>. 2021.
- [336] ATLAS Collaboration. “Search for electroweak production of supersymmetric particles in final states with two or three leptons at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *Eur. Phys. J. C* 78 (2018), p. 995. DOI: 10.1140/epjc/s10052-018-6423-7. arXiv: 1803.02762 [hep-ex].
- [337] ATLAS Collaboration. “Search for chargino-neutralino pair production in final states with three leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector”. *Eur. Phys. J. C* 81 (2021), p. 1118. DOI: 10.1140/epjc/s10052-021-09749-7. arXiv: 2106.01676 [hep-ex].
- [338] ATLAS Collaboration. “Search for chargino-neutralino production with mass splittings near the electroweak scale in three-lepton final states in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector”. *Phys. Rev. D* 101 (2020), p. 072001. DOI: 10.1103/PhysRevD.101.072001. arXiv: 1912.08479 [hep-ex].
- [339] ATLAS Collaboration. “Search for supersymmetry in events with four or more charged leptons in 139 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector”. *JHEP* 07 (2021), p. 167. DOI: 10.1007/JHEP07(2021)167. arXiv: 2103.11684 [hep-ex].
- [340] ATLAS Collaboration. “Search for long-lived, massive particles in events with a displaced vertex and a muon with large impact parameter in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *Phys. Rev. D* 102 (2020), p. 032006. DOI: 10.1103/PhysRevD.102.032006. arXiv: 2003.11956 [hep-ex].
- [341] ATLAS Collaboration. “Search for charginos and neutralinos in final states with two boosted hadronically decaying bosons and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. *Phys. Rev. D* 104 (2021), p. 112010. DOI: 10.1103/PhysRevD.104.112010. arXiv: 2108.07586 [hep-ex].
- [342] CMS Collaboration. “Search for chargino-neutralino production in events with Higgs and W bosons using 137 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JHEP* 10 (2021), p. 045. DOI: 10.1007/JHEP10(2021)045. arXiv: 2107.12553 [hep-ex].
- [343] CMS Collaboration. “Search for supersymmetry in final states with two or three soft leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JHEP* 04 (2022), p. 091. DOI: 10.1007/JHEP04(2022)091. arXiv: 2111.06296 [hep-ex].
- [344] CMS Collaboration. “Search for supersymmetry in final states with two oppositely charged same-flavor leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JHEP* 04 (2021), p. 123. DOI: 10.1007/JHEP04(2021)123. arXiv: 2012.08600 [hep-ex].
- [345] CMS Collaboration. “Search for electroweak production of supersymmetric particles in final states containing hadronic decays of WW, WZ, or WH and missing transverse momentum”. Tech. rep. CMS-PAS-SUS-21-002. 2021. URL: <http://cds.cern.ch/record/2779116>.

- [346] P. Baldi et al. “Parameterized neural networks for high-energy physics”. *Eur. Phys. J. C* 76 (2016), p. 235. DOI: 10.1140/epjc/s10052-016-4099-4. arXiv: 1601.07913 [hep-ex].
- [347] LHC New Physics Working Group. “Simplified models for LHC new physics searches”. *J. Phys. G* 39 (2012). Ed. by N. Arkani-Hamed et al., p. 105005. DOI: 10.1088/0954-3899/39/10/105005. arXiv: 1105.2838 [hep-ph].
- [348] CMS Collaboration. “Interpretation of searches for supersymmetry with simplified models”. *Phys. Rev. D* 88 (2013), p. 052017. DOI: 10.1103/PhysRevD.88.052017. arXiv: 1301.2175 [hep-ex].
- [349] K. T. Matchev and S. D. Thomas. “Higgs and Z boson signatures of supersymmetry”. *Phys. Rev. D* 62 (2000), p. 077702. DOI: 10.1103/PhysRevD.62.077702. arXiv: hep-ph/9908482.
- [350] J. T. Ruderman and D. Shih. “General neutralino NLSPs at the early LHC”. *JHEP* 08 (2012), p. 159. DOI: 10.1007/JHEP08(2012)159. arXiv: 1103.6083 [hep-ph].
- [351] P. Meade, M. Reece, and D. Shih. “Prompt decays of general neutralino NLSPs at the Tevatron”. *JHEP* 05 (2010), p. 105. DOI: 10.1007/JHEP05(2010)105. arXiv: 0911.4130 [hep-ph].
- [352] W. Beenakker et al. “The production of charginos/neutralinos and sleptons at hadron colliders”. *Phys. Rev. Lett.* 83 (1999). [Erratum: *Phys.Rev.Lett.* 100, 029901 (2008)], p. 3780. DOI: 10.1103/PhysRevLett.100.029901. arXiv: hep-ph/9906298 [hep-ph].
- [353] B. Fuks et al. “Gaugino production in proton-proton collisions at a center-of-mass energy of 8 TeV”. *JHEP* 10 (2012), p. 081. DOI: 10.1007/JHEP10(2012)081. arXiv: 1207.2159 [hep-ph].
- [354] S. Abdullin et al. “The fast simulation of the CMS detector at LHC”. *J. Phys. Conf. Ser.* 331 (2011), p. 032049. DOI: 10.1088/1742-6596/331/3/032049.
- [355] CMS Collaboration. “Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons”. Tech. rep. CMS-DP-2019-033. 2019. URL: <http://cds.cern.ch/record/2694158>.
- [356] CMS Collaboration. “Combined search for electroweak production of charginos and neutralinos in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JHEP* 03 (2018), p. 160. DOI: 10.1007/JHEP03(2018)160. arXiv: 1801.03957 [hep-ex].
- [357] P. Baldi, P. Sadowski, and D. Whiteson. “Searching for exotic particles in high-energy physics with deep learning”. *Nature Commun.* 5 (2014), p. 4308. DOI: 10.1038/ncomms5308. arXiv: 1402.4735 [hep-ph].
- [358] P. Baldi, P. Sadowski, and D. Whiteson. “Enhanced Higgs boson to $\tau^+\tau^-$ search with deep learning”. *Phys. Rev. Lett.* 114 (2015), p. 111801. DOI: 10.1103/PhysRevLett.114.111801. arXiv: 1410.3469 [hep-ph].
- [359] I. Goodfellow, Y. Bengio, and A. Courville. “Deep learning”. MIT Press, 2016. ISBN: 9780262035613.
- [360] P. T. Komiske, E. M. Metodiev, and J. Thaler. “Energy flow networks: Deep sets for particle jets”. *JHEP* 01 (2019), p. 121. DOI: 10.1007/JHEP01(2019)121. arXiv: 1810.05165 [hep-ph].
- [361] B. K. Iwana, R. Kuroki, and S. Uchida. “Explaining convolutional neural networks using softmax gradient layer-wise relevance propagation”. *Proceedings of the 2019 International Conference on Computer Vision Workshop, ICCVW 2019* (2019), p. 4176. DOI: 10.1109/ICCVW.2019.00513. arXiv: 1908.04351 [cs.CV].
- [362] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. 2014. arXiv: 1412.6980 [cs.LG].
- [363] Y. Nesterov. “A method for unconstrained convex minimization problem with the rate of convergence $\mathcal{O}(1/k^2)$ ”. *Soviet Math. Dokl.* 269 (1983), p. 543.
- [364] S. Ruder. “An overview of gradient descent optimization algorithms”. 2017. arXiv: 1609.04747 [cs.LG].

- [365] S. Ioffe and C. Szegedy. “Batch Normalization: Accelerating deep network training by reducing internal covariate shift”. *Proceedings of the 32nd International Conference on International Conference on Machine Learning* 37 (2015), p. 448. arXiv: 1502.03167 [cs.LG].
- [366] N. Srivastava et al. “Dropout: A simple way to prevent neural networks from overfitting”. *J. Mach. Learn. Res.* 15 (2014), p. 1929.
- [367] ATLAS Collaboration. “Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector”. *Phys. Rev. D* 92 (2015), p. 012006. DOI: 10.1103/PhysRevD.92.012006. arXiv: 1412.2641 [hep-ex].
- [368] C. Lester and D. Summers. “Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders”. *Phys. Lett. B* 463 (1999), p. 99. DOI: 10.1016/S0370-2693(99)00945-4. arXiv: hep-ph/9906349.
- [369] C. G. Lester. “The stransverse mass, M_{T2} , in special cases”. *JHEP* 05 (2011), p. 076. DOI: 10.1007/JHEP05(2011)076. arXiv: 1103.5682 [hep-ph].
- [370] W. Verbeke. “Data driven WZ background estimation for SUSY searches with multiple leptons at the energy frontier, LHC”. MA thesis. 2016. URL: <http://lib.ugent.be/catalog/rug01:002272601>.
- [371] R. D. Cousins. “Generalization of chisquare goodness-of-fit test for binned data using saturated models, with application to histograms”. 2010. URL: www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.
- [372] O. Brüning and L. Rossi. “The High-Luminosity Large Hadron Collider.” *Nat. Rev. Phys.* 1 (2019), p. 241. DOI: 10.1038/s42254-019-0050-6.
- [373] L. Rossi and O. Brüning. “The High Luminosity Large Hadron Collider: The new machine for illuminating the mysteries of Universe”. World Scientific, 2015. ISBN: 9789814675468.