

---

---

# The Dark Side of the Higgs

*Searches for dark matter with a focus on invisibly decaying Higgs bosons using the full Run-2 dataset of the CMS experiment at the LHC*

---

By

ESHWEN BHAL



School of Physics  
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Science.

DECEMBER 2020

Word count: thirty five thousand, five hundred



## ABSTRACT

Dark matter is a poorly-understood phenomenon in nature. Though substantial evidence corroborates its existence, only few characteristics have been determined. One goal of the Large Hadron Collider (LHC) at CERN is to produce dark matter in high energy proton-proton collisions, potentially allowing insight into its currently-mysterious origins. Many hypotheses have been postulated regarding its nature, two of which are investigated in this thesis: invisible decays of the Higgs boson, and the production of semi-visible jets. The data used is from the LHC Run-2 era and recorded by the CMS experiment, corresponding to an integrated luminosity of  $137 \text{ fb}^{-1}$  at a centre of mass energy of 13 TeV.

The branching ratio of the Higgs boson to invisible states is predicted to be 0.1 % in the standard model. Enhancements from a coupling to dark matter may be observable at the LHC. A search is performed in final states comprising jets and missing transverse momentum targeting the  $t\bar{t}H$  and  $VH$  Higgs boson production modes. With the full Run-2 dataset from CMS, no significant deviation from the standard model was observed. Results are presented as an upper limit on the measured cross section times branching ratio over the standard model Higgs boson cross section at the 95 % confidence level. For the  $t\bar{t}H$ -tagged and  $VH$ -tagged events, observed (expected) limits of 0.56 (0.50) and 0.32 (0.22), respectively were achieved. A combined Run-2 limit of 0.28 was observed and 0.20 expected. These are interpreted in simplified dark matter scenarios.

Dark matter may exist in a Hidden Valley dark sector connected to the visible universe via a leptophobic mediator. Analogous to QCD, dark quarks may be produced at the LHC, hadronising and decaying into a mixture of visible and invisible particles: a semi-visible jet. The behaviour of simulated signal from  $s$ - and  $t$ -channel production modes in the CMS detector—and variables discriminating it from background—is explored. The transverse mass of the dijet system was found to be the most effective.



## DEDICATION AND ACKNOWLEDGEMENTS

This work is dedicated to my grandfather, Dato' Mahindar Singh Bhal, who was able to begin this journey with me but sadly unable to finish it. There are far too many people and too little space to individually thank everyone who has accompanied me during this PhD, but I'll try my best.

Firstly, to my supervisor, Dr. Henning Flächer. Your advice and guidance over the course of this degree has been instrumental to achieving it, as well as encouraging my growth as a researcher. If you hadn't spent so many late nights working with me and providing feedback, this thesis would be in a much shoddier state.

Secondly, to all my colleagues at the University of Bristol. Very little would have been achieved without your help. From scientific discussions in the pub to pub discussions in the physics building, they have all been fruitful either directly or by helping me detach from work. Special thanks are in order to Simone and Sam Maddrell-Mander in my cohort, who have put up with my complaining during the stressful times, as well as Ben Krikler and Olivier Davignon—the postdocs who likely helped me solve the issues I was complaining about in the first place.

To the "Massive" from Monmouth that include my best friends in the world, Mike, James, Sneddon, (Mini) Sam, and Matt Bristow. You've been on this adventure with me since our school days and had my back the entire time. We've been through the highest, lowest, and most hilarious times together. I can definitively say I would not be the person I am today without you.

To my friends from LTA and those based at CERN, especially Dwayne, Vukasin, and Matt Heath. You became my second family while I was in Geneva. I'll miss the skiing, trips into the city, games, and general banter.

To my undergraduate cohort from Exeter. Though our meet ups are rare, I always look forward to our group holidays and hope they continue far into the future.

Lorenza Iacobuzio, you get a special mention. While we weren't especially close until recently, you've been my sage, partner in food, and exceptional friend when I've needed it the most.

Finally, and most importantly, my family deserve my thanks. You have been there for me since the very beginning supporting me through school, undergrad, PhD, and every other endeavour. To my grandparents, uncles, aunties, and cousins, my siblings Joe, Lydia, Sitara, and Arjen, my stepmum Nadia, and my cat Pixie, I am sincerely grateful to have you in my lives. But above all else, to my mum Meeta and dad Kiron, I could not have done any of this without you and as such, you have my deepest and most heartfelt love and gratitude.



## AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: .....



## TABLE OF CONTENTS

	Page
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 The standard model and beyond</b>	<b>5</b>
2.1 The standard model of particle physics . . . . .	5
2.1.1 Particles of the standard model . . . . .	6
2.1.2 Symmetries and gauge invariance . . . . .	7
2.1.3 Electroweak symmetry breaking and the Higgs mechanism . . . . .	8
2.1.4 Masses of the fermions . . . . .	11
2.1.5 Limitations of the standard model . . . . .	11
2.2 The dark matter landscape . . . . .	12
2.2.1 Evidence . . . . .	12
2.2.2 Theoretical descriptions . . . . .	14
2.2.3 Searches . . . . .	15
2.3 Measuring the branching ratio of invisibly decaying Higgs bosons . . . . .	18
2.3.1 Production modes of the Higgs boson . . . . .	19
2.3.2 Results of previous searches . . . . .	22
2.4 Searches for semi-visible jets . . . . .	23
2.4.1 Kinematics and free parameters of the model . . . . .	24
<b>3 The LHC and the CMS experiment</b>	<b>27</b>
3.1 The Large Hadron Collider . . . . .	27
3.1.1 A proton's journey . . . . .	28
3.1.2 Luminosity . . . . .	30

---

## TABLE OF CONTENTS

---

	3.1.3 Pileup . . . . .	31
	3.1.4 Evolution of the LHC . . . . .	31
3.2	The CMS experiment . . . . .	34
	3.2.1 The CMS detector . . . . .	34
	3.2.2 Data acquisition and triggering . . . . .	41
	3.2.3 Simulating CMS data . . . . .	44
	3.2.4 Jet energy corrections in the Level-1 Trigger . . . . .	44
<b>4</b>	<b>Definitions of physics objects and observables</b>	<b>49</b>
4.1	Ubiquitous observables and quantities in collider physics . . . . .	49
	4.1.1 Transverse momentum ( $\vec{p}_T$ ) . . . . .	49
	4.1.2 $H_T$ . . . . .	50
	4.1.3 Missing transverse momentum ( $\vec{p}_T^{\text{miss}}$ ) . . . . .	50
	4.1.4 Transverse mass ( $m_T$ ) . . . . .	50
4.2	Classification of analysis-level physics objects . . . . .	51
	4.2.1 Jets . . . . .	51
	4.2.2 Boosted decays of top quarks and electroweak bosons . . . . .	54
	4.2.3 Muons . . . . .	55
	4.2.4 Electrons . . . . .	55
	4.2.5 Photons . . . . .	56
	4.2.6 Tau leptons . . . . .	57
	4.2.7 Overlap removal . . . . .	58
	4.2.8 Revised energy sums . . . . .	58
<b>5</b>	<b>Simulation studies of semi-visible jets</b>	<b>59</b>
5.1	Analysis summary . . . . .	59
5.2	Signal simulation overview . . . . .	60
5.3	Generation in PYTHIA . . . . .	61
5.4	Showering in PYTHIA . . . . .	61
	5.4.1 Hadronisation of dark quarks . . . . .	61
	5.4.2 Jet clustering . . . . .	63
	5.4.3 Filtering events . . . . .	64
5.5	Generation in MADGRAPH . . . . .	64
	5.5.1 MADGRAPH run settings . . . . .	65
	5.5.2 $s$ -channel . . . . .	66
	5.5.3 $t$ -channel . . . . .	68

5.6	PYTHIA-MADGRAPH comparisons for $s$ -channel signal . . . . .	70
<b>6</b>	<b>Search for invisibly decaying Higgs bosons</b>	<b>75</b>
6.1	Overview of the analysis . . . . .	75
6.1.1	Analysis strategy . . . . .	76
6.2	Only tools and forces: software and toolkits . . . . .	77
6.3	Data and simulation . . . . .	78
6.3.1	Data . . . . .	78
6.3.2	Simulated signal processes . . . . .	79
6.3.3	Simulated background processes . . . . .	79
6.3.4	Cross section reweighting . . . . .	81
6.4	Event selection . . . . .	81
6.4.1	Preselection . . . . .	81
6.4.2	Additional filters . . . . .	82
6.4.3	Mitigating the HEM issue . . . . .	83
6.5	Categorisation of the production modes . . . . .	84
6.5.1	$t\bar{t}H$ categories . . . . .	86
6.5.2	$VH$ categories . . . . .	86
6.5.3	Optimisation of the categories . . . . .	86
6.5.4	Binning . . . . .	90
6.5.5	The signal region . . . . .	91
6.6	Statistical model for fit to data . . . . .	92
6.7	Background estimation . . . . .	95
6.7.1	Control regions . . . . .	96
6.7.2	Lost lepton background estimation . . . . .	99
6.7.3	$Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ background estimation . . . . .	103
6.7.4	QCD multijet background estimation . . . . .	108
6.8	Corrections and systematic uncertainties for simulation . . . . .	111
6.8.1	Efficiency of the triggers . . . . .	112
6.8.2	Pileup reweighting . . . . .	113
6.8.3	Higher order corrections to $V + \text{jets}$ samples . . . . .	114
6.8.4	Top quark processes . . . . .	116
6.8.5	Object-level scale factors and uncertainties . . . . .	119
6.8.6	Minor contributions . . . . .	120
6.8.7	Summary of systematic uncertainties . . . . .	122

---

**TABLE OF CONTENTS**

6.9	Results . . . . .	124
6.9.1	Results from the control region-only fit . . . . .	124
6.9.2	Results from the background-only fit . . . . .	128
6.9.3	Results from the $t\bar{t}H$ channel . . . . .	131
6.9.4	Results from the $VH$ channel . . . . .	134
6.9.5	Combined results . . . . .	137
6.10	Interpretations in simplified dark matter scenarios . . . . .	138
6.10.1	Higgs portal model with the standard model Higgs boson . . . . .	138
7	<b>Conclusions</b>	141
7.1	$H \rightarrow$ invisible . . . . .	141
7.2	Semi-visible jets . . . . .	143
A	<b>Additional material for the Higgs to invisible analysis</b>	145
A.1	Control region-only fits to the $t\bar{t}H$ categories . . . . .	146
A.2	Control region-only fits to the $VH$ categories . . . . .	150
A.3	Rate parameters from the control region-only fit . . . . .	154
A.4	Inputs to the QCD multijet prediction . . . . .	160
A.5	Background-only fits to the $t\bar{t}H$ categories . . . . .	163
A.6	Background-only fits to the $VH$ categories . . . . .	164
A.7	Limits and likelihood scans for each category . . . . .	165
	<b>Bibliography</b>	169
	<b>Glossary</b>	183
	<b>Acronyms</b>	185

## LIST OF FIGURES

FIGURE	Page
2.1 A depiction of the Higgs potential $V$ as a function of the component fields . . . . .	9
2.2 A measure of the comoving number density of WIMP dark matter as a function of time with projections for several particle masses . . . . .	15
2.3 A visual representation of the three main types of dark matter detection: direct, indirect, and production . . . . .	16
2.4 The expected masses and interaction cross sections of a set of dark matter candidates	17
2.5 A subset of the Feynman diagrams for the four predominant production mechanisms of the Higgs boson at the LHC . . . . .	20
2.6 Example Feynman diagrams for the two main production modes of semi-visible jets. A $Z'$ boson mediates the $s$ -channel process while a bi-fundamental $\Phi$ mediates the $t$ -channel process . . . . .	24
2.7 The constituents of a semi-visible jet as a function of its invisible fraction . . . . .	25
2.8 The typical direction of the missing transverse energy relative to the semi-visible jets as a function of the invisible fraction $r_{\text{inv}}$ . . . . .	25
3.1 A schematic of the CERN accelerator complex . . . . .	29
3.2 The average number of pileup interactions per bunch crossing at CMS during Run-2 of the LHC . . . . .	32
3.3 The integrated luminosity of $pp$ collision data delivered to CMS during Run-2 of the LHC . . . . .	33
3.4 A cutaway diagram of the CMS detector for the 2017–18 data taking years, with all of the principal components labelled . . . . .	35
3.5 A quadrant of the CMS detector illustrating the main subsystems with their radius, longitudinal distance, and pseudorapidity from the interaction point . .	36
3.6 A transverse slice through the barrel section of the CMS detector with the main subsystems and components visible . . . . .	40

## LIST OF FIGURES

---

3.7	A summary of the CMS Level-1 Trigger data flow from the hits recorded by the subsystems to the Global Trigger . . . . .	42
3.8	Examples of correction curves used to calibrate the jet energies in two $ \eta $ bins . . . . .	46
3.9	The energies of matched pairs of jets from simulation in the entire barrel and end cap, in the pileup 40–50 range, before and after jet energy corrections have been applied . . . . .	47
3.10	The response curves in each $ \eta $ bin as a function of $p_T^{L1}$ , in the pileup 40–50 range, before and after jet energy corrections are applied . . . . .	47
5.1	A diagram of the mass insertion decay of $\pi_{\text{dark}}$ mesons in the $s$ -channel semi-visible jet model . . . . .	62
5.2	The dependence of the dark force scale $\Lambda_{\text{dark}}$ on the dark hadron mass $m_{\text{dark}}$ for each value of $\alpha_{\text{dark}}$ used in the thesis . . . . .	63
5.3	Distributions of the transverse mass of the dijet system $m_T$ for $s$ -channel semi-visible jet samples emulating the 2016 data taking period. In each panel, one of the free parameters of the model is varied with respect to the benchmark point SVJ_3000_20_0.3_peak . . . . .	67
5.4	Distributions of the transverse mass of the dijet system $m_T$ for $s$ -channel semi-visible jet samples emulating the 2016 data taking period. In each panel, one of the free parameters of the model is varied with respect to the benchmark point SVJ_2000_20_0.5_peak . . . . .	69
5.5	Distributions of several observables for the models SVJ_1000_20_0.3_peak, SVJ_-3000_20_0.1_peak, and SVJ_3000_20_0.9_peak . . . . .	71
5.6	Distributions of several observables for the models SVJ_3000_20_0.5_peak, SVJ_-3000_50_0.3_peak, and SVJ_4000_20_0.3_peak . . . . .	72
6.1	The azimuthal angle of the $\vec{p}_T^{\text{miss}}$ inclusive in the $t\bar{t}H$ and $VH$ categories before applying the selections designed to mitigate the HEM issue in 2018 . . . . .	84
6.2	Distributions of $\tilde{\omega}_{\min}$ in the signal region inclusive in the $t\bar{t}H$ and $VH$ categories, along with the significance—from several figures of merit—if a cut is placed to the right of a given value . . . . .	89
6.3	Composition of each category in the signal region for simulated signal and background processes after the analysis-level selection . . . . .	92
6.4	An infographic showcasing the role of each analysis region in the fit to data . . . . .	96
6.5	Composition of each category in each of the control regions for simulated background processes after the analysis-level selection . . . . .	100

---

6.6	Post-fit yields for each category and $p_T^{\text{miss}}$ bin in the single lepton control regions of the $t\bar{t}H$ categories for the 2017 dataset . . . . .	101
6.7	Post-fit yields for each category and $p_T^{\text{miss}}$ bin in the single lepton control regions of the $VH$ categories for the 2017 dataset . . . . .	102
6.8	Fits of real and fake photon templates to data in the $\sigma_{inj}$ spectrum . . . . .	104
6.9	The fraction of impure photons in data as a function of $p_T$ for each data taking year in Run-2 . . . . .	105
6.10	Post-fit yields for each category and $p_T^{\text{miss}}$ bin in the dilepton control regions of the $t\bar{t}H$ categories for the 2017 dataset . . . . .	106
6.11	Post-fit yields for each category and $p_T^{\text{miss}}$ bin in the dilepton and photon control regions of the $VH$ categories for the 2017 dataset . . . . .	107
6.12	Data–simulation comparisons of the $p_T^{\text{miss}}$ distribution in the $t\bar{t}H$ and $VH$ sidebands, aggregated over the full Run-2 dataset . . . . .	110
6.13	Scale factors accounting for the efficiencies of the HLT $p_T^{\text{miss}} - H_T^{\text{miss}}$ cross triggers in each data taking year of Run-2 . . . . .	113
6.14	NLO $k$ -factors—as a function of the generator-level boson $p_T$ and leading jet $p_T$ —for reweighting events in LO QCD $W$ and $Z + \text{jets}$ simulation to NLO accuracy in QCD . . . . .	115
6.15	NLO $k$ -factors as a function of generator-level boson $p_T$ for reweighting QCD $V + \text{jets}$ events generated at LO electroweak accuracy . . . . .	116
6.16	The deviations relative to the nominal weight of the combinations of QCD renormalisation and factorisation scale variations. They are presented as a function of $p_T^{\text{miss}}$ inclusive over the $t\bar{t}H$ categories for the $t\bar{t}$ samples . . . . .	117
6.17	The ratio of the NNLO to NLO $t\bar{t}$ cross section as a function of top quark $p_T$ . The fit to the distribution—along with the uncertainties in the parameters—are shown . . . . .	119
6.18	Pre-fit and post-fit yields in the signal region for every $t\bar{t}H$ category and $p_T^{\text{miss}}$ bin in each year of Run-2 . . . . .	132
6.19	Observed and expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction $\mathcal{B}(H \rightarrow \text{invisible})$ (left) and the corresponding profile likelihood scan (right) in the $t\bar{t}H$ channel . . . . .	133
6.20	Pre-fit and post-fit yields in the signal region for every $VH$ category and $p_T^{\text{miss}}$ bin in each year of Run-2 . . . . .	135

---

## LIST OF FIGURES

---

6.21 Observed and expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction $\mathcal{B}(H \rightarrow \text{invisible})$ (left) and the corresponding profile likelihood scan (right) in the $VH$ channel . . . . .	136
6.22 Observed and expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction $\mathcal{B}(H \rightarrow \text{invisible})$ and the corresponding profile likelihood scan, for both the individual categories, as well as the combination of them, for the full Run-2 dataset . . . . .	137
6.23 90 % confidence level upper limits on the spin-independent dark matter-nucleon scattering cross section in Higgs portal models, where the standard model Higgs boson decays into a pair of scalar (solid orange) or fermion (dashed red) dark matter particles . . . . .	140
A.1 Post-fit yields for each $t\bar{t}H$ category and $p_T^{\text{miss}}$ bin in the single lepton control regions for the 2016 dataset . . . . .	146
A.2 Post-fit yields for each $t\bar{t}H$ category and $p_T^{\text{miss}}$ bin in the dilepton control regions for the 2016 dataset . . . . .	147
A.3 Post-fit yields for each $t\bar{t}H$ category and $p_T^{\text{miss}}$ bin in the single lepton control regions for the 2018 dataset . . . . .	148
A.4 Post-fit yields for each $t\bar{t}H$ category and $p_T^{\text{miss}}$ bin in the dilepton control regions for the 2018 dataset . . . . .	149
A.5 Post-fit yields for each $VH$ category and $p_T^{\text{miss}}$ bin in the single lepton control regions for the 2016 dataset . . . . .	150
A.6 Post-fit yields for each $VH$ category and $p_T^{\text{miss}}$ bin in the dilepton and photon control regions for the 2016 dataset . . . . .	151
A.7 Post-fit yields for each $VH$ category and $p_T^{\text{miss}}$ bin in the single lepton control regions for the 2018 dataset . . . . .	152
A.8 Post-fit yields for each $VH$ category and $p_T^{\text{miss}}$ bin in the dilepton and photon control regions for the 2018 dataset . . . . .	153
A.9 Results of the background-only fit to the $t\bar{t}H$ categories in each year of Run-2 .	163
A.10 Results of the background-only fit to the $VH$ categories in each year of Run-2 .	164
A.11 Observed and expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the $t\bar{t}H$ channel, for both the individual categories, and the combination of them, for each data-taking year in Run-2 . . . . .	165
A.12 Observed and expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the $VH$ channel, for both the individual categories, and the combination of them, for each data-taking year in Run-2 . . . . .	166

---

LIST OF FIGURES

A.13 Observed and expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction $\mathcal{B}(H \rightarrow \text{invisible})$ and the corresponding profile likelihood scan, in the $t\bar{t}H$ and $VH$ boosted and resolved category groupings and their combination, for the full Run-2 dataset . . . . .	167
--	-----



## LIST OF TABLES

TABLE	Page
2.1 A summary of the fermionic particles of the standard model . . . . .	7
2.2 A summary of the bosonic particles of the standard model . . . . .	7
2.3 Cross sections of the $H \rightarrow$ invisible signal processes . . . . .	20
2.4 The most recent searches for invisibly decaying Higgs bosons in hadronic channels with 2016 data from CMS, and the achieved upper limits on the $H \rightarrow$ invisible branching ratio at 95 % confidence level . . . . .	22
3.1 The integrated luminosity delivered by the LHC during Run-2 which were recorded and certified by CMS . . . . .	32
3.2 Some of the important parameters defining the performance of the LHC at the end of Run-2, for the High-Luminosity LHC, and the design specification . . .	33
4.1 The requirements for a jet to pass tight identification for data taken in 2016, and to Monte Carlo events emulating that year . . . . .	52
4.2 The requirements for a jet to pass tight identification for data taken in 2017, and to Monte Carlo events emulating that year . . . . .	53
4.3 The requirements for a jet to pass tight identification for data taken in 2018, and to Monte Carlo events emulating that year . . . . .	53
4.4 Criteria for defining the veto and tight electron IDs in the barrel region (supercluster $ \eta  \leq 1.479$ ) . . . . .	55
4.5 Criteria for defining the veto and tight electron IDs in the end cap region (supercluster $ \eta  > 1.479$ ) . . . . .	56
4.6 Criteria for defining the loose and medium photon IDs in the barrel region of the detector ( $ \eta  \leq 1.479$ ) . . . . .	57
4.7 Criteria for defining the loose and medium photon IDs in the end cap region of the detector ( $ \eta  > 1.479$ ) . . . . .	57

## LIST OF TABLES

---

6.1	Categorisation of the $t\bar{t}H$ and $VH$ production modes in the analysis . . . . .	85
6.2	The binning scheme implemented to categorise events in terms of $p_T^{\text{miss}}$ . . . . .	91
6.3	The trigger thresholds required for events to enter the signal region in each data taking year . . . . .	91
6.4	The trigger requirements for events to enter the $e + \text{jets}$ or $ee + \text{jets}$ control regions, if they originate from the dataset of $e$ -based triggers . . . . .	98
6.5	The trigger requirements for events to enter the $e + \text{jets}$ $ee + \text{jets}$ , or $\gamma + \text{jets}$ control regions, if they originate from the dataset of $\gamma$ -based triggers . . . . .	98
6.6	Sources of uncertainty in the luminosity measurements for each year of Run-2 .	122
6.7	Systematic uncertainties on the cross section of the simulated $H \rightarrow \text{invisible}$ signal samples . . . . .	122
6.8	The experimental uncertainties present in the analysis with the applicable samples and impact on the corresponding event counts . . . . .	123
6.9	Background event yields in the signal region as predicted by the control regions, and observed data event counts from the 2016 dataset . . . . .	125
6.10	Background event yields in the signal region as predicted by the control regions, and observed data event counts from the 2017 dataset . . . . .	126
6.11	Background event yields in the signal region as predicted by the control regions, and observed data event counts from the 2018 dataset . . . . .	127
6.12	Results of the combined fit to signal and control regions with the 2016 dataset .	128
6.13	Results of the combined fit to signal and control regions with the 2017 dataset .	129
6.14	Results of the combined fit to signal and control regions with the 2018 dataset .	130
6.15	Observed and median expected upper limits on $\mathcal{B}(H \rightarrow \text{invisible})$ at 95 % confidence level for each combination of channel and dataset analysed . . . . .	138
A.1	Monte Carlo and data yields in the $\mu + \text{jets}$ and $e + \text{jets}$ control regions of the 2016 dataset, and corresponding rate parameter $a_{\ell_{\text{lost}}}$ connecting them to the signal region, for each category and $p_T^{\text{miss}}$ bin . . . . .	154
A.2	Monte Carlo and data yields in the $\mu\mu + \text{jets}$ , $ee + \text{jets}$ , and $\gamma + \text{jets}$ control regions of the 2016 dataset, and corresponding rate parameter $a_{Z \rightarrow \nu\bar{\nu}}$ connecting them to the signal region, for each category and $p_T^{\text{miss}}$ bin . . . . .	155
A.3	Monte Carlo and data yields in the $\mu + \text{jets}$ and $e + \text{jets}$ control regions of the 2017 dataset, and corresponding rate parameter $a_{\ell_{\text{lost}}}$ connecting them to the signal region, for each category and $p_T^{\text{miss}}$ bin . . . . .	156

---

LIST OF TABLES

A.4 Monte Carlo and data yields in the $\mu\mu + \text{jets}$ , $ee + \text{jets}$ , and $\gamma + \text{jets}$ control regions of the 2017 dataset, and corresponding rate parameter $a_{Z \rightarrow \nu\bar{\nu}}$ connecting them to the signal region, for each category and $p_T^{\text{miss}}$ bin . . . . .	157
A.5 Monte Carlo and data yields in the $\mu + \text{jets}$ and $e + \text{jets}$ control regions of the 2018 dataset, and corresponding rate parameter $a_{\ell_{\text{lost}}}$ connecting them to the signal region, for each category and $p_T^{\text{miss}}$ bin . . . . .	158
A.6 Monte Carlo and data yields in the $\mu\mu + \text{jets}$ , $ee + \text{jets}$ , and $\gamma + \text{jets}$ control regions of the 2018 dataset, and corresponding rate parameter $a_{Z \rightarrow \nu\bar{\nu}}$ connecting them to the signal region, for each category and $p_T^{\text{miss}}$ bin . . . . .	159
A.7 Components of the QCD multijet prediction in the signal region for the 2016 dataset . . . . .	160
A.8 Components of the QCD multijet prediction in the signal region for the 2017 dataset . . . . .	161
A.9 Components of the QCD multijet prediction in the signal region for the 2018 dataset . . . . .	162



## INTRODUCTION

Wonder is the beginning of wisdom.

— Socrates

The universe, in all its vastness, structure, natural laws and chaos, is comprised of only three principal components: visible matter, the ingredients of stars, planets and life, is the only one we interact with on a regular basis; dark energy, a force or manifestation of something even more mysterious responsible for the accelerating expansion of the universe, is almost entirely unknown; and dark matter, a substance invisible in all sense of the word, that binds galaxies together and influences large scale structure in the cosmos, is the focus of this thesis.

The earliest evidence for a large, non-luminous component of the galaxy stretches back to the 1920s when Jacobus Kapteyn attempted to explain the motion of stars in the Milky Way [101]. Swiss astronomer Fritz Zwicky was the first to dub it as *dunkle Materie*, or *dark matter*. Since then, a wealth of independent astrophysical observations have reinforced the existence of this aggregation not just in our own, but in countless other galaxies and cosmological bodies. Despite considerable examination, relatively little is truly known about dark matter; from its origins and its place among the particles of the standard model (SM), to the reason its total mass dwarfs visible matter both in our galaxy and the universe as a whole. Given its apathy toward interacting with light—the primary medium by which we observe the dynamics of the cosmos—and the other forces that particle physics leverages to probe its foundations, the puzzling nature of dark matter is a tantalizing lure for any physicist.

While observational evidence has so far lain with astrophysics, a theoretical description and discovery of dark matter may fall into the realm of particle physics with the numerous, novel experimental searches underway. Since dark matter may be produced in measurable quantities by a particle accelerator, it is natural to utilise data from high energy collisions to conduct searches.

This thesis focuses predominantly on a search for invisibly decaying Higgs bosons in the  $t\bar{t}H$  and  $VH$  channels at the LHC. In the standard model of particle physics, the only avenue by which the Higgs boson can decay invisibly is  $H \rightarrow ZZ \rightarrow 4\nu$  with a branching ratio of  $\mathcal{O}(0.1\%)$  [92]. The leading experimental upper limits on this measurement are 19 % from CMS [139] and 11 % from ATLAS [24], far higher than the predicted value. The VBF channel gives the greatest contribution to the sensitivity in both results. If undiscovered invisible particles—such as dark matter—couple to the Higgs field the branching ratio will be enhanced. As the Higgs field bestows mass to the other known elementary particles, it is possible that dark matter obtains its mass from the same mechanism. A considerably large enhancement to the invisible state branching ratio may allow for this process to be observed at the LHC. At the very least, a more accurate constraint is able to exclude some models of dark matter in which the Higgs boson acts as a portal between the visible and dark sectors, such as those described in Refs. 76 and 100.

Each of the above production modes of the Higgs boson yields distinct signatures in a particle detector, with hadronic final states investigated in this thesis given their high rates. If the Higgs decays invisibly, the hadronic activity would be accompanied by large “missing” transverse momentum (explained further in Chpt. 4.1.3), a quantity that represents the momenta of particles invisible to the detector. The unique topologies of the production mechanisms can be characterised by the multiplicities of light flavour,  $b$ -tagged, and large radius jets, all of whose mass and substructure is compatible with hadronic decays of vector bosons and top quarks, and produced in conjunction with missing transverse momentum.  $t\bar{t}H$  presents a particularly formidable search with the smallest cross section and final states with many jets. With the full Run-2 dataset of  $137\text{ fb}^{-1}$  collected by the CMS experiment at the LHC, there is great potential to constrain the upper limit on the Higgs boson to invisible state branching fraction, and as a consequence various Higgs portal models.

In addition to the Higgs sector, a generator study is presented on a new physics model resulting in *semi-visible jets*: a hadronic final state in which dark matter particles are interspersed within jets [69, 70]. It is a novel and experimentally challenging signature easily mistaken for mismeasured QCD multijet processes. Understanding how the expected signal would manifest in CMS is an important part of constructing an analysis to best identify it.

---

There is significant motivation to study dark matter from a wider, as well as a more personal viewpoint. It is important to understand how the universe operates, and dark matter opens up the potential for new physics that improves our understanding of nature. My personal interests include the blend of particle physics and astrophysics, the opportunity to discover, and add to humanity’s collective wisdom. An outline of this thesis is given as follows. A theoretical foundation is laid out in Chpt. 2 for the characteristics of the expected signal from invisibly decaying Higgs bosons and semi-visible jets; they are built on top of the standard model which is recapitulated. The landscape of evidence, possible descriptions, and searches for dark matter are also summarised to provide further context for the motivations of this work. Chpt. 3 consists of the predominant aspects of the design and capabilities of the Large Hadron Collider, and the detector composition and data acquisition at the CMS experiment. Physics objects and quantities are defined in Chpt. 4 that are subsequently utilised in the analysis chapters: signal simulation studies of semi-visible jets are presented in Chpt. 5, and the search for invisibly decaying Higgs bosons is described in Chpt. 6 (for which App. A is reserved, containing supplementary material). Finally, a summary of this thesis and conclusions drawn from the obtained results form Chpt. 7.



## THE STANDARD MODEL AND BEYOND

It doesn't stop being magic just because you know how it works.

— Terry Pratchett

This thesis is comprised of experimental searches for dark matter and new physics. Chpt. 6 delves deeply into the predominantly-featured search for an invisibly decaying Higgs boson. A small chapter is dedicated to the pursuit of semi-visible jets with Chpt. 5. Before visiting either of which, the theoretical and phenomenological motivations must be understood to corroborate the incentives for searches at the Large Hadron Collider. In this chapter, a brief recap of the standard model—with emphasis on the Higgs mechanism—will be presented along with its shortcomings, foremost the lack of a dark matter candidate. Evidence and the characteristics of dark matter are then discussed, underpinning theoretical descriptions that best fit the relic density and astrophysical observations. A short summary is also given of the state of dark matter searches at the LHC. Specific interpretations in the forms of semi-visible jets and invisibly decaying Higgs bosons are lastly examined that provide the background for the respective analysis chapters.

### 2.1 The standard model of particle physics

The standard model of particle physics is the best description of nature the human race has to offer. Three of the four fundamental forces are encapsulated by it: the strong (nuclear)

force, the weak (nuclear) force, and electromagnetism. The latter two may instead be considered components of a single *electroweak* force [86, 123, 145], unified above an energy of  $\mathcal{O}(100\text{ GeV})$ ; the LHC, with collision energies far above this threshold, observes much electroweak physics. All of the elementary particles—the quarks, leptons, gauge bosons, and the Higgs boson—and their interactions with each other are contained within the standard model. These are described initially in Chpt. 2.1.1, followed by a recap of the role symmetries and gauge invariance play in their interactions with Chpt. 2.1.2. Then, in Chpt. 2.1.3, the Higgs mechanism and its eponymous boson, and how they factor in to the decomposition of the electroweak force into the electromagnetic and weak interactions are explained. In the aforementioned passage and in Chpt. 2.1.4, the acquisition of mass by the electroweak gauge bosons and fermions are respectively described.

### 2.1.1 Particles of the standard model

The standard model contains a relatively small, but diverse “zoo” of particles. They can be divided into two distinct categories based on their internal property *spin*: fermions with half-integer spin, comprise the quarks and leptons that constitute matter; and bosons with integer spin, mediate the interactions (i.e., the forces) between the fermions.

Six types, or *flavours*, of quark and six flavours of lepton exist, arranged in three generations. Particles between generations share many similarities, the primary differentiator being mass. In the quark sector, the first generation contains the up  $u$  and down  $d$ , the second the charm  $c$  and strange  $s$ , and the third the top  $t$  and bottom  $b$ . The former particle in each generation carries an electric charge of  $+2/3$  of the elementary charge  $e$  while the latter a value of  $-1/3$ . All quarks also carry colour charge, allowing them to combine into colour-neutral composite *hadron* particles such as protons and neutrons. Quarks can interact via all of the forces in the standard model.

In the lepton sector, each generation consists of a massive particle with electric charge  $-1e$ , and an associated electrically neutral, and extremely light, neutrino. The three generations consist of the electron  $e$  and electron neutrino  $\nu_e$ , the muon  $\mu$  and muon neutrino  $\nu_\mu$ , and the tau  $\tau$  with the tau neutrino  $\nu_\tau$ . Charged leptons can interact via electromagnetism and the weak force. Though the neutrinos, lacking any electric charge, only interact via the weak force. The main properties of the fermions are summarised in Tab. 2.1.

Each force is carried by one or more spin-1 gauge boson, acting to mediate the fermions’ interactions. Eight flavours of the massless, electrically neutral gluon  $g$  carries the strong force, while the equally massless and chargeless photon  $\gamma$  mediates the electromagnetic interaction. Three massive bosons mediate the weak interaction, the charged  $W^\pm$  and neutral  $Z$ . A scalar

Type	Generation	Particle	Spin	Electric charge	Mass
Quark	1	Up ( $u$ )	1/2	+2/3 e	$2.2^{+0.5}_{-0.4} \text{ MeV}/c^2$
		Down ( $d$ )	1/2	-1/3 e	$4.7^{+0.5}_{-0.3} \text{ MeV}/c^2$
	2	Charm ( $c$ )	1/2	+2/3 e	$1.28^{+0.03}_{-0.04} \text{ GeV}/c^2$
		Strange ( $s$ )	1/2	-1/3 e	$95^{+9}_{-3} \text{ MeV}/c^2$
	3	Top ( $t$ )	1/2	+2/3 e	$173 \pm 0.4 \text{ GeV}/c^2$
		Bottom ( $b$ )	1/2	-1/3 e	$4.18^{+0.04}_{-0.03} \text{ GeV}/c^2$
Lepton	1	Electron ( $e$ )	1/2	-1 e	$0.511 \text{ MeV}/c^2$
		Electron neutrino ( $\nu_e$ )	1/2	0	$< 0.2 \text{ eV}/c^2$
	2	Muon ( $\mu$ )	1/2	-1 e	$106 \text{ MeV}/c^2$
		Muon neutrino ( $\nu_\mu$ )	1/2	0	$< 0.2 \text{ eV}/c^2$
	3	Tau ( $\tau$ )	1/2	-1 e	$1.777 \text{ GeV}/c^2$
		Tau neutrino ( $\nu_\tau$ )	1/2	0	$< 0.2 \text{ eV}/c^2$

Table 2.1: A summary of the fermionic particles of the standard model. Masses obtained from Ref. 143.

(spin-0) Higgs boson  $H$  is the excitation of the Higgs field that acts to bestow mass to the elementary particles. All of the bosons are summarised in Tab. 2.2.

Force	Particle	Spin	Electric charge	Mass
Strong	Gluon ( $g$ )	1	0	0
	Photon ( $\gamma$ )	1	0	0
Weak	$W$ bosons ( $W^\pm$ )	1	$\pm 1 e$	$80.38 \pm 0.01 \text{ GeV}/c^2$
Weak	$Z$ boson ( $Z$ )	1	0	$91.19 \text{ GeV}/c^2$
—	Higgs boson ( $H$ )	0	0	$125.18 \pm 0.16 \text{ GeV}/c^2$

Table 2.2: A summary of the bosonic particles of the standard model. Masses obtained from Ref. 143.

### 2.1.2 Symmetries and gauge invariance

The standard model is a gauge quantum field theory: particles are characterised by excitations of quantum fields and their interactions described by continuous gauge symmetry groups. A symmetry is a feature in a theory where a quantity is preserved under specific transformations. This is significant for understanding what kind of interactions are allowed by each force. An  $SU(3)$  gauge group represents the interactions in the strong force, and an  $SU(2) \times U(1)$  gauge group for the electroweak force. Decomposition into the weak force and electromagnetism

results in separate  $SU(2)$  and  $U(1)$  groups, respectively. The interactions of the standard model as a whole can therefore be expressed as the product  $SU(3) \times SU(2) \times U(1)$ .

Noether's theorem associates a continuous symmetry of a physical system, that does not affect its Lagrangian, to a conserved charge or current [117]. A consequence of which is that an interaction represented by a particular group must conserve the charges associated with the symmetries of said group. The generators of the group<sup>1</sup> correspond to the gauge invariant fields that mediate the associated force, i.e., the gauge bosons in the SM, enforcing the conservation of the charges. With the  $SU(3)$  strong force, the mediating gluons carry colour charge. The electroweak force conserves weak isospin and weak hypercharge from the  $SU(2)$  and  $U(1)$  components, respectively. Mediated by the progenitors of the  $W$ ,  $Z$ , and photon, the boson fields in the  $SU(2)$  group are referred to as  $W_1$ ,  $W_2$ , and  $W_3$ , while the  $B$  field is the generator of the  $U(1)$  group. As the SM is described by the product of these gauge groups, it includes only gauge invariant fields that preserve the Lagrangian and equations of motions under the allowed transformations or interactions.

Symmetries in the standard model require the gauge bosons to be massless. For electromagnetism and the strong force, this is no issue. However, the bosons mediating the weak force have been determined experimentally to be massive [21, 26], posing a problem. One solution is to introduce a new field that can break the symmetry of the  $SU(2)$  group without affecting the gauge invariance elsewhere in the SM. This led to the introduction of the Higgs field.<sup>2</sup>

### 2.1.3 Electroweak symmetry breaking and the Higgs mechanism

As mentioned previously, the electromagnetic and weak forces are unified into the electroweak force above a certain energy threshold. In this regime, the electroweak bosons and fermions must be massless to maintain gauge invariance. The former obtain their mass by interacting with the Higgs field, the means by which is labelled the *Higgs mechanism* [81, 88, 95].

The simplest model by which the Higgs mechanism can be accommodated in the SM is by introducing an  $SU(2)$  doublet of complex scalar Higgs fields:

$$(2.1) \quad \Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}$$

---

<sup>1</sup>There are  $N^2 - 1$  generators for  $SU(N)$  and  $N^2$  for  $U(N)$ . This is what gives rise to the eight gluons of the strong force, three ( $W^+$ ,  $W^-$ , and  $Z$ ) bosons of the weak force, and the single photon of electromagnetism.

<sup>2</sup>There are many who deserve credit for formulating the theory. However, for concision, the corresponding paradigms will be referred to as the *Higgs field* or *Higgs mechanism* henceforth.

It possesses four degrees of freedom, i.e., one per gauge boson in the electroweak sector. The fields introduce additional terms in the standard model Lagrangian, most importantly a potential of the form

$$(2.2) \quad V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$$

which is the most general scalar potential that is also  $SU(2) \times U(1)$  invariant. With both quadratic and quartic components, positive values of both free parameters  $\mu^2$  (related to mass) and  $\lambda$  (related to self-interaction) would yield a single, stable minimum for the potential. In a physical context, this corresponds to the universe pre-electroweak symmetry breaking. But by setting  $\lambda > 0$  and  $\mu^2 < 0$ , what was once the minimum becomes an unstable local maximum and a new, degenerate circle of minima in the  $\phi_1 - \phi_2$  plane can be found with values

$$(2.3) \quad \Phi^\dagger \Phi|_{\min} = -\frac{\mu^2}{2\lambda}$$

forming the familiar “wine bottle” potential. An illustration with a toy Higgs boson is given in Fig. 2.1.

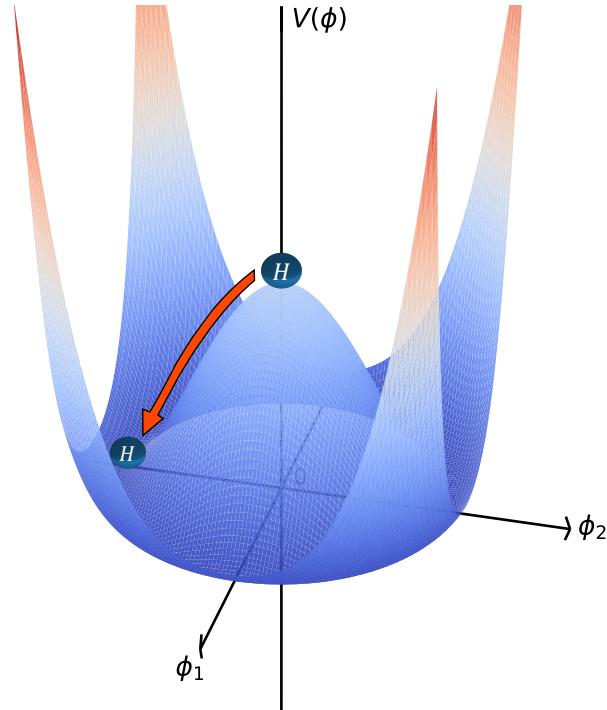


Figure 2.1: A depiction of the Higgs potential  $V$  as a function of the component fields. The Higgs boson is initially at  $(\phi_1, \phi_2) = (0, 0)$ , an unstable local maximum. It then tumbles into the minimum value of  $V$  and triggers electroweak symmetry breaking.

To spontaneously break the  $SU(2) \times U(1)$  symmetry of the electroweak force, a non-zero vacuum expectation value  $v$  must be assigned from the set of minima belonging to the potential  $V$ . Since the ability to select any potential is a symmetry, assigning a value breaks said symmetry. For the fields  $\phi_1$ ,  $\phi_2$ , and  $\phi_4$ ,  $v$  can simply be set to zero. However, as a neutral field,  $\phi_3$  must be assigned a non-zero value in order for the photon to remain massless after the symmetry is broken. It is allocated a value

$$(2.4) \quad \langle 0 | \phi_3 | 0 \rangle = v^2 = -\frac{\mu^2}{\lambda}$$

By expanding  $\Phi$  around the chosen vacuum and using the unitary gauge, it can be seen that all of the fields other than  $\phi_3$ —known as *Goldstone fields* in this paradigm—are eliminated. The originally massless  $W_{1,2,3}$  and  $B$  fields then mix, with some states gain mass by “eating” the Goldstone bosons. The  $W_1$  and  $W_2$  fields transform into the  $W^\pm$  bosons:

$$(2.5) \quad W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2)$$

While the  $W_3$  and  $B$  fields mix to produce the massless photon and massive  $Z$  boson:

$$(2.6) \quad \begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos \theta_w & \sin \theta_w \\ -\sin \theta_w & \cos \theta_w \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix}$$

In the new  $U(1)$  symmetry group of electromagnetism (mediated by the photon), electric charge is the conserved current, manifesting as a combination of weak hypercharge and the third component of weak isospin. For the  $SU(2)$  weak interaction, weak isospin becomes the conserved quantity. Due to only left-handed chiral states of fermions (and right-handed antifermions) possessing a non-zero weak isospin, certain interactions are not allowed by the weak force. Only left-handed fermions and right-handed antifermions couple to the  $W$  bosons that participate in *charged current* interactions. This maximally violates the parity symmetry conserved by the other forces. Left- and right-handed fermions also couple differently to the  $Z$  boson that mediates *neutral current* interactions.

The masses of the electroweak bosons can be calculated using the vacuum expectation value, and the coupling strengths of the weak force  $g$  and electromagnetism  $g'$ :

$$(2.7) \quad \begin{aligned} m_W &= \frac{1}{2}gv, \\ m_Z &= \frac{1}{2}(g \cos \theta_w + g' \sin \theta_w)^2 v = \frac{1}{2} \frac{gv}{\cos \theta_w} \end{aligned}$$

where  $\theta_w$  is the *weak mixing angle* or *Weinberg angle*. It is a quantity that represents the degree to which the weak and electromagnetic forces mix, defined as

$$(2.8) \quad \cos \theta_w = \frac{m_W}{m_Z} \text{ or } \tan \theta_w = \frac{g'}{g}$$

It is a free parameter in the standard model with its value constrained by measuring the  $W$  and  $Z$  boson masses. With the latest results in Ref. 143,  $\theta_w = 28.74^\circ$  and  $v = 246 \text{ GeV}/c^2$ . The mass of the Higgs boson is given by

$$(2.9) \quad m_H = \sqrt{2\lambda v^2} = \sqrt{-2\mu^2}$$

### 2.1.4 Masses of the fermions

The quarks and leptons in the SM acquire their masses from the Higgs field by means of a Yukawa coupling  $y_f$ , which describes the interaction between a scalar (Higgs) field and Dirac (fermion) field. As a function of  $y_f$  and the vacuum expectation value of the Higgs field, the fermion masses are calculated as

$$(2.10) \quad m_f = \frac{1}{\sqrt{2}} y_f v$$

They are directly proportional to their coupling strength to the Higgs, and are all free parameters in the SM so must be constrained by experiment. The decay width  $\Gamma$  of the Higgs boson into fermions is computed as

$$(2.11) \quad \Gamma(H \rightarrow f\bar{f}) = \frac{m_H}{8\pi} \left( \frac{m_f}{v} \right)^2 N_c \left( 1 - \frac{4m_f^2}{m_H^2} \right)^{\frac{3}{2}}$$

which is proportional to the squared mass of the decay products. The *branching ratio*  $\mathcal{B}$ —the probability of a particle decaying into a given final state—is simply the ratio of the partial width to the total width, the latter being the sum of partial widths for all the particle’s decay modes.

### 2.1.5 Limitations of the standard model

Despite the standard model providing precise predictions of three of the four fundamental forces and the particles interacting through them, there are many experimental observations that it cannot currently explain. Neutrino masses, dark matter, dark energy, and gravity all escape its description. Many of the important parameters in the SM do not have predicted values and so must be experimentally measured.

The Hierarchy Problem is one of the more serious issues facing the standard model. It may be explained in different manners that emphasize certain aspects. But inherently, it is a question of the disparity between energy scales of the fundamental forces—particularly relating to the weak force and gravity. The masses of the electroweak and Higgs bosons of  $\mathcal{O}(100 \text{ GeV}/c^2)$  are much smaller than the Planck mass of  $\mathcal{O}(10^{19} \text{ GeV}/c^2)$ . The mass term for the Higgs boson is  $\mu^2 \Phi^\dagger \Phi$  in the SM. Invariance under a gauge or global symmetry in the Higgs field leads to the mass being open to radiative corrections up to the Planck scale. It appears that, in nature, these very large corrective terms to the Higgs boson mass cancel to give the familiar  $m_H = 125 \text{ GeV}/c^2$  [4, 46]. It is deemed unnatural to expect cancellations to such a degree, i.e., one part in  $10^{17}$ . This *fine-tuning* of parameters in the standard model is something that unified or natural theories desperately try to avoid.

Some theories beyond the standard model (BSM) like supersymmetry (SUSY) provide well-motivated cancellations by introducing supersymmetric particles. In certain scenarios, some of these particles should exist at the TeV scale. In the SM, the largest correction to the Higgs mass derives from the top quark, since its Yukawa coupling to the Higgs is the strongest. At one-loop order, new physics at the  $\mathcal{O}(\text{TeV})$  scale is required, with new particles coupling to the Higgs field to prevent these corrections from being unreasonably large [82]. Arguments such as this give credence to new physics being discoverable at particle accelerators such as the Large Hadron Collider.

## 2.2 The dark matter landscape

### 2.2.1 Evidence

There is a wealth of evidence affirming the existence of dark matter in the universe. The Coma Cluster is a famous example: 90 % of its mass is thought to arise from dark matter, confirmed by its large mass-to-light ratio of  $400 M_\odot/L_\odot$  [147]. Further evidence is that the rotation curves of most galaxies are roughly flat [118], contrary to the expected Keplerian relationship ( $v_r \propto r^{-1/2}$ ) expected from solely visible matter. On a galactic scale, dark matter is sprinkled in a mostly spherical halo that spans beyond the observable disc. The inclusive dark matter mass increases linearly [79] to compensate for the decline expressed by visible matter [38, 84]. Gravitational lensing is another observational tool subject to influence from dark matter. Images of galaxies and other objects captured by this method appear distorted from a large gravitational field between the source and observer warping its local spacetime [97]. Arcs, ellipses and Einstein rings of smeared galaxies are often seen when dark matter is present.

While there are no widely-accepted estimations, it is believed that 85–95 % of the Milky Way is comprised of dark matter [29, 30, 99]. Though these approximations include non-visible identifiable matter such as dim stars, black holes and neutron stars, the term *dark matter* is typically reserved for the non-luminous, *non-baryonic* segment that pervades the cosmos. From the latest results of the Planck mission, the energy density of the observable universe is composed of 26.5 % dark matter [8]. This result follows the Lambda cold dark matter ( $\Lambda$ CDM) model to describe the constituents and evolution of the universe, which is often referred to as the cosmological analog of the standard model. From the calculations, postulations, and observations presented above, the following properties of dark matter can be deduced:

- It is electrically neutral as it does not interact with electromagnetic radiation. Hence, the adjective “dark” in dark matter.
- It is non-relativistic, or *cold*. Its velocity within galaxies is similar to the inhabiting stars [35, 94], since the combination of visible and dark matter drives the measured rotation curves. From classical mechanics, galactic dark matter *must* be cold since a velocity above the gravitational escape velocity of the galaxy would eject high speed particles.
- It is stable, at least on the timescale of the age of the universe. Dark matter production is postulated to have occurred only in the early universe via a thermal freeze-out mechanism. Hence, the remaining fraction has been present for a considerable time. Since most galaxies are dominated by dark matter and the gravitational influence from only the visible matter is too small to maintain itself, they could not have developed without it. This supports the idea of bottom-up structure formation in the universe: smaller galaxies form around gravitational potential wells induced by coalescing dark matter, then merge to form larger structures [146].
- Its interaction with matter and itself is very weak, or even non-existent. The Bullet Cluster—an astronomical object consisting of two colliding galaxy clusters—is the best example of this inference. From measurements of, predominantly, x-ray emission and gravitational lensing, it was found that while there is a substantial amount of dark matter present, interaction with itself and the visible matter surrounding it was minimal at most [49]. A kinematic explanation for the spherical distribution and low velocity of dark matter in galaxies can be explained by its collisionless nature. During the formation of a galaxy or planetary system, visible matter frequently collides, dissipating angular momentum and collapsing into a disc.

### 2.2.2 Theoretical descriptions

Dark matter may have been forged in the universe via one of many possible mechanisms. The most popular is described as a *thermal freeze-out* process. In the hot, early universe when the thermal background allowed spontaneous pair production of particle dark matter, it was generated in abundance. During this period, the particles may also have frequently annihilated seeing that the cosmos was still small. Inevitably, the universe expanded and cooled; the temperature became too low to allow significant production [27]. Matter was further separated and the dark matter annihilation rate decreased, leaving a behind the *thermal relic* that is observed today. These remaining particles were attracted via gravity, forming filaments throughout the universe. The potential wells they induced allowed the progenitors of galaxies to form within.

Full derivations of the thermal freeze-out of dark matter can be found in literature [32, 87], with the WIMP Miracle as a consequence: with relatively few assumptions, the correct dark matter relic abundance can be recovered by requiring a Weakly Interacting Massive Particle (WIMP) with a mass of  $\mathcal{O}(\text{GeV-TeV})$  dependent on the annihilation cross section, and an interaction strength similar to that of the weak force. This mass range is accessible at contemporary colliders such as the LHC, and perhaps coincidentally, around the electroweak energy scale. It is common for figures that depict the WIMP dark matter density over time to plot the yield  $n_\chi/s$  as a function of the dimensionless parameter  $x = m_\chi/T$ . In the former variable,  $n_\chi$  is the number density and  $s$  is the entropy density. In the latter,  $m_\chi$  is the dark matter mass and  $T$  is the average temperature of the universe, which serves as a measure of its age due to the temperature decreasing over time. An example is given in Fig. 2.2.

The time of the dark matter freeze-out epoch is somewhat insensitive to the mass and annihilation cross section. Approximate solutions to the Boltzmann equation for a time-dependent  $n_\chi$ —where dark matter is modelled as a weakly-interacting, diffuse gas of particles—suggest  $x_f \sim 20$  [32, 113]. Stronger dark matter interaction leads to decoupling at a later time and a lower number density. The approximate value of  $x_f$  is significant in that it supports the electroweak-scale mass of WIMPs. A higher mass must be balanced by a larger annihilation cross section to achieve the correct relic density, to which it tends asymptotically from the point of decoupling.

Another popular mechanism, targeting low-mass dark matter, is the *freeze-in* process [89, 109]. In this postulate, dark matter is not produced thermally in the early universe. Instead, it emerges through interactions between SM particles such as collisions, or decays of those heavier than dark matter. The comoving density increases with time until it plateaus from the cooling of the universe, where SM particles are generally stable enough and too low

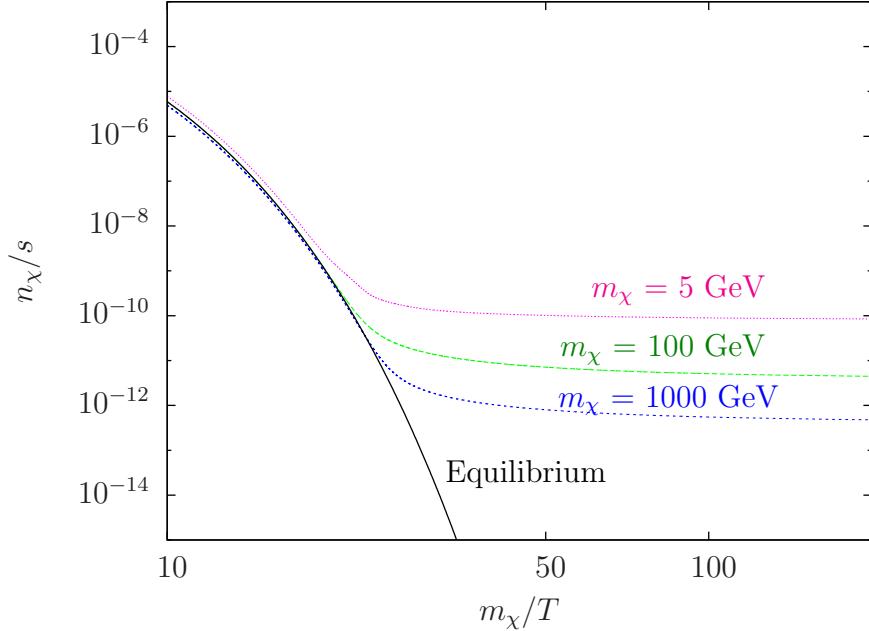


Figure 2.2: A measure of the comoving number density of WIMP dark matter as a function of time with projections for several particle masses. The black curve represents the scenario in which dark matter remains in equilibrium with the standard model. Figure taken from Ref. 90.

energy to produce dark matter in any meaningful quantity. The relic abundance can therefore be reclaimed from a combination of the initial thermal distributions, the dark matter mass, and the interaction strength, similar to the freeze-out process. In order to obey cosmological observations, particularly the fact that it is cold, the masses expected for freeze-in dark matter particles are of  $\mathcal{O}(\text{keV})$  or heavier.

### 2.2.3 Searches

A coalescence of observation from astrophysics and application from particle physics has paved the way for a nimiety of dark matter models that can be tested by either discipline. Searches dark matter can be classified into three distinct methods with unique signatures (paired with a visual summary in Fig. 2.3):

- **Direct:** dark matter may interact with visible matter on small scales, scattering from standard model particles [125]. The recoil the SM particles experience could be detected by highly-sensitive, low background experiments such as LUX-ZEPLIN (LZ) [11] which specialises in the search for WIMP dark matter within a wide range of masses.
- **Indirect:** if dark matter interacts with itself, it may annihilate to produce showers of

high energy photons or pions. Background estimation is difficult since the signatures can be highly model-dependent. The particles may be of a continuum—from hadronisation and radiation of the decay products—or contain features, such as internal radiation from the propagator in the interaction or from loop-level processes [71]. Large ranges of the annihilation cross section and dark matter mass can be probed with telescopes already searching for these characteristic events.

- **Production:** dark matter may have been abundantly produced in the hot, early universe. High energy particle accelerators such as the LHC can reproduce these conditions, with the WIMP Miracle reinforcing the idea that dark matter may exist in these accessible mass ranges. Many BSM theories accommodate dark matter candidates with a diverse spectrum of final states that can be investigated by analysing LHC data.

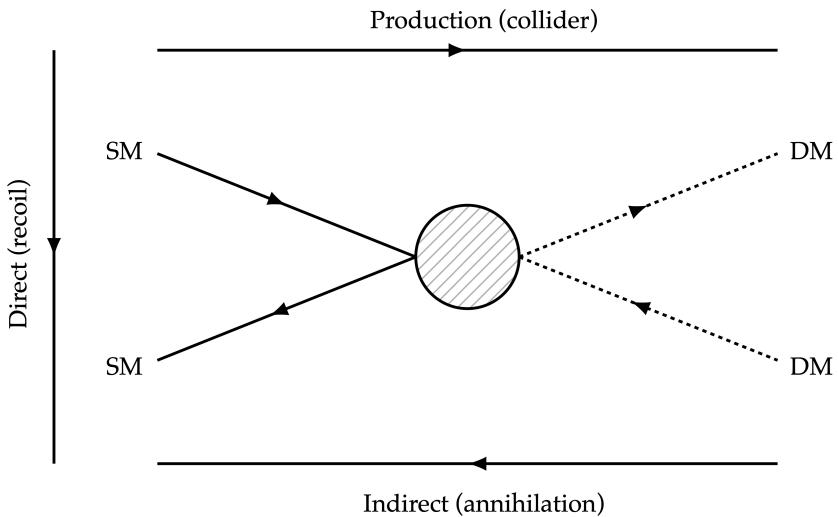


Figure 2.3: A visual representation of the three main types of dark matter detection: direct (dark matter recoiling from standard model particles); indirect (annihilation of dark matter); and production (dark matter created in high energy particle collisions).

Since detection at the CMS experiment from the production mechanism is the subject of this thesis, it is important to establish the current state of dark matter searches at the LHC, the world’s most powerful particle accelerator that provides the infrastructure for CMS to collect data. The LHC principally collides protons at centre of mass energies up to  $\sqrt{s} = 13$  TeV. These exceptionally high energies allow the conditions in the very early universe to be simulated in which heavy, unstable particles were produced plentifully. As a result, many theories can be investigated that predict heavy particles that do not exist in the universe today. Some of these, such as SUSY [114], sterile neutrinos [78], and Kaluza-Klein states [91] contain dark matter candidates that can be specifically searched for, or indirectly

inferred if a theory is experimentally proven. Despite the success of the standard model in explaining much of the natural world, it does not substantiate the existence of dark matter. BSM theories can therefore gain traction. Fig. 2.4 illustrates the masses and interaction cross sections of many dark matter candidates. WIMPs (highlighted by the purple rectangle) are the subject of several searches at the LHC since the expected mass ranges and cross sections are accessible there.

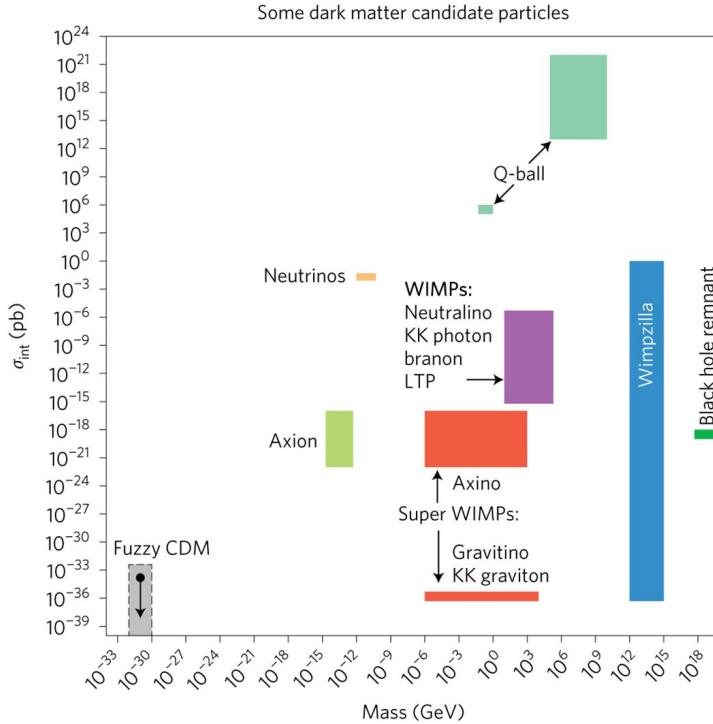


Figure 2.4: The expected masses and interaction cross sections of a set of dark matter candidates. The LHC, with its centre of mass energy of 13 TeV, is best suited to targeting Weakly Interacting Massive Particles (WIMPs). Figure acquired from Ref. 71.

Two avenues are usually considered when attempting to discover dark matter: explicit searches for the signatures of dark matter production, and anomalies in precision measurements. The former is quite common, with many theories and models tested at the LHC's general purpose detectors, ATLAS and CMS. Searches at CMS have been performed for promptly-decaying and *long-lived* supersymmetry in hadronic final states [103, 129]. Searches for specific supersymmetric particles in a variety of decay modes have been conducted by both experiments [42]. In many of these cases, the lightest supersymmetric particle (LSP) is considered to be a dark matter candidate.  $R$ -parity conservation is predicted (or even enforced) in many SUSY models [114], which prevents the decay of LSP and any lighter, standard model particles that have been observed to be stable. While supersymmetry is

the most popular BSM theory, due in part to its numerous interpretations and approaches for discovery, many others have also been explored at the LHC. From microscopic black holes [106] to dark photons [131], there are extensive propositions that have the potential to uproot the standard model. The analyses above are usually characterised by large missing transverse momentum ( $\vec{p}_T^{\text{miss}}$ ). Both its magnitude and direction in relation to the visible event content are important, especially when searching for topologies like semi-visible jets.

Precise measurements of standard model parameters is the other method often consulted in the hopes of attributing discrepancies to new physics. For example, attempts to explain anomalies in the  $b \rightarrow s$  transition include dark matter candidates [44, 144]. An investigation into the measurement of the Higgs boson to invisible state branching ratio is extensively detailed in Chpt. 6 along with interpretations that can accommodate dark matter. Since the same laws of physics and accepted descriptions of the universe are shared across fields, results from new measurements or searches in one sector can influence others. Notably, stronger constraints on standard model measurements like the above can exclude BSM theories searched for in direct and indirect dark matter detection.

## 2.3 Measuring the branching ratio of invisibly decaying Higgs bosons

The Higgs boson has caught the attention of the high energy physics community, and even the public eye, like no other particle in recent memory. Its discovery in the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ \rightarrow 4\ell$  channels in 2012—Independently by both CMS [46] and ATLAS [4]—realised one of the paramount goals of the LHC’s construction. The particle itself is not necessarily exciting. Rather, it confirms the existence of the Higgs *field* that pervades the universe and gives mass to the elementary particles via the exchange of its eponymous boson [81, 88, 95]. Its discovery, one might think, was the end of the discussion of the Higgs boson. However, it was only the beginning.

Many observations of the Higgs, such as its predominant decay mode  $H \rightarrow b\bar{b}$ , were not seen until recently by CMS [127] or ATLAS [3]. Constraints on its other properties have also been placed, such as its resonance width and branching ratios to several final states [143]. Fully understanding the Higgs boson is important to understanding the Higgs field and the wider standard model. Precision measurements in tension with SM predictions can also be a window to new physics. Measuring the  $H \rightarrow$  invisible branching ratio aims to do just that.

The only SM process in which Higgs boson can decay invisibly is to four neutrinos via a pair of  $Z$  bosons, possessing a branching ratio of only  $\mathcal{O}(0.1\%)$  while leading experimental

upper limits are  $\mathcal{O}(10\%)$ . If the Higgs field couples to dark matter, the overall  $\mathcal{B}(H \rightarrow \text{invisible})$  could increase substantially with possible observations at the LHC. If not, the upper limit on the branching ratio can still be chipped away, further constraining or excluding a selection of dark matter models.

Since the other elementary particles also acquire mass from the Higgs field, the same may be true for dark matter. Higgs *portal* models have been theorised that connect the visible sector of the standard model to a dark sector where particle dark matter resides [20, 23]. Certain models also predict a detectable presence at the LHC from a sufficient production rate [37], perhaps even with data obtained during Run-2 [7]. A dark matter mass of  $m_\chi < m_H/2$  would allow for direct pair production from the Higgs boson—a much stronger signal than via an intermediate decay, as is the case for the  $4\nu$  final state.

An analysis in search of this decay is provided thoroughly in Chpt. 6. Constraints on the experimental side stem largely from the different channels in which a Higgs boson can be produced. These are outlined in Chpt. 2.3.1, and must all be considered when examining such a rare process that is also difficult to distinguish amongst a large background. Previous results from searches for individual modes, including subsequent combinations, are documented in Chpt. 2.3.2.

### 2.3.1 Production modes of the Higgs boson

At the LHC, the most common mechanisms for producing a Higgs boson are vector boson fusion (VBF), gluon-gluon fusion ( $ggF$  or  $ggH$ ), associated production from top quarks ( $t\bar{t}H$ ), and associated production from a vector boson ( $VH$ ). Feynman diagrams of these processes are shown in Fig. 2.5. Additional diagrams for  $ggH$  involve a square top quark loop and/or initial state radiation. The  $ZH$  process can be initiated by  $gg \rightarrow ZH$  as well as  $pp \rightarrow ZH$ . They all have very different characteristics, production rates, and event signatures, complementing each other and allowing analyses to cover all bases with orthogonal parameter spaces to target them individually. One common feature of these final states is the presence of at least one quark. The hadronic constituents in the decay products of a collision often shower due to colour confinement, producing collimated sprays of hadrons called *jets*. In a detector, these are represented by clusters of hadronic energy deposits. Algorithms at each stage of data acquisition (see Chpt. 3.2.2 for those in CMS) can reliably connect these back to the individual quark decays so one has some certainty of the process they are observing. The cross section of each mechanism at  $\sqrt{s} = 13$  TeV is detailed in Tab. 2.3.

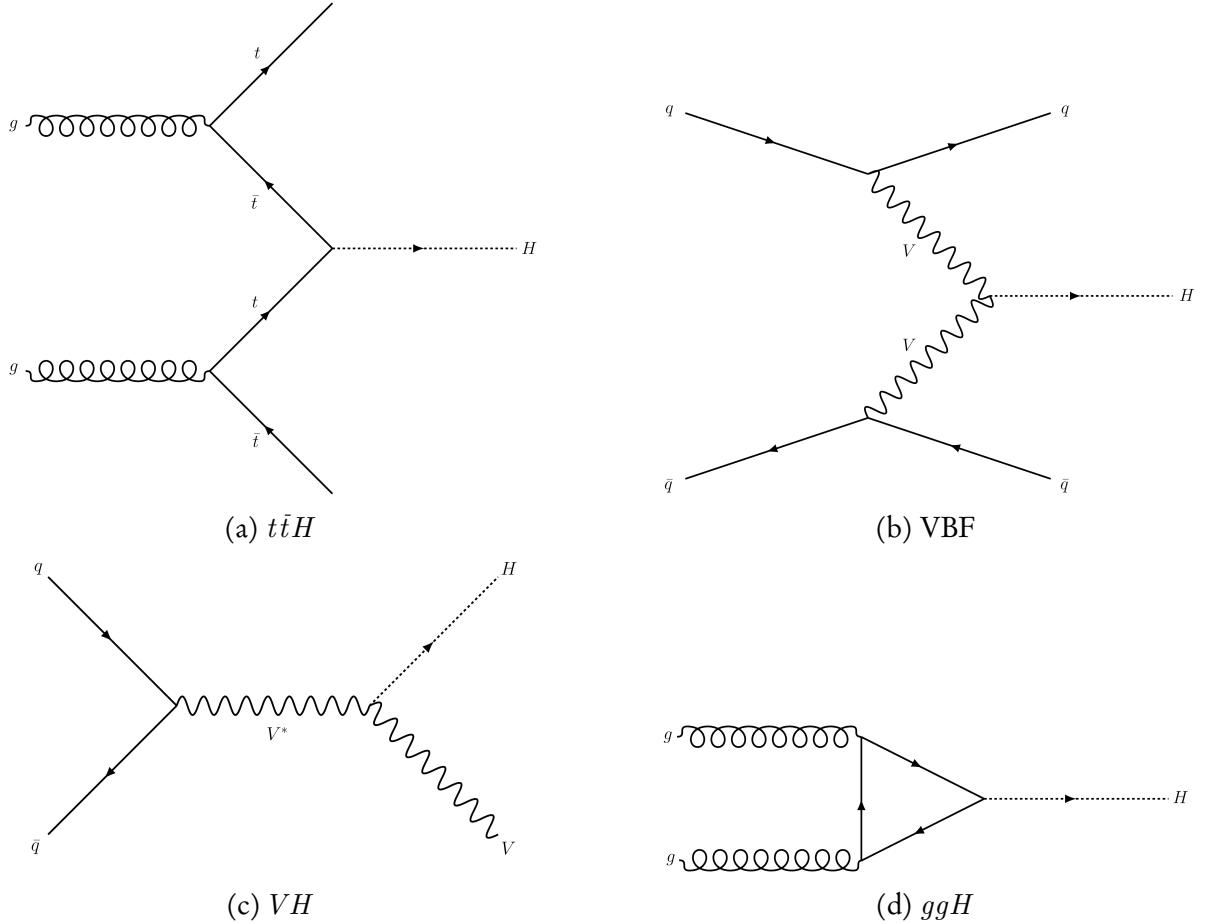


Figure 2.5: A subset of the Feynman diagrams for the four predominant production mechanisms of the Higgs boson at the LHC.

Mode ( $H \rightarrow$ invisible)	Cross section (pb)	Accuracy
VBF	3.77	NNLO QCD and NLO electroweak
$t\bar{t}H$	$5.07 \times 10^{-1}$	NLO QCD and NLO electroweak
$W^+H$	$8.31 \times 10^{-1}$	NNLO QCD and NLO electroweak
$W^-H$	$5.27 \times 10^{-1}$	NNLO QCD and NLO electroweak
$pp \rightarrow ZH$	$8.84 \times 10^{-1}$	NNLO QCD and NLO electroweak
$gg \rightarrow ZH$	$1.23 \times 10^{-1}$	LO QCD (NLO + NLL corrections)
$ggH$	$4.86 \times 10^1$	N3LO QCD and NLO electroweak

Table 2.3: Cross sections of the  $H \rightarrow$  invisible signal processes. They are calculated at  $\sqrt{s} = 13$  TeV at the highest orders available and obtained from Ref. 43.

### 2.3.1.1 Vector boson fusion (VBF)

A VBF topology is exhibited by a  $t$ -channel exchange of two vector bosons radiated by the incident quarks, which then combine to form a new particle such as a Higgs boson. Since the masses of the  $W$  and  $Z$  bosons are more than half the Higgs' mass, it can easily be produced on shell. The recoil of the quarks from the Higgs boson characterises the visible system: two jets with a large combined invariant mass, usually with a large separation in pseudorapidity  $\eta$  but small in azimuthal angle.<sup>3</sup> The jets move in opposite directions, one in  $+\eta$  and the other in  $-\eta$ , but are usually contained in the same horizontal half of the detector.

### 2.3.1.2 Associated production from top quarks ( $t\bar{t}H$ )

In  $t\bar{t}H$ , a  $t\bar{t}$  pair is produced from the collision. A virtual top quark  $t$  and antiquark  $\bar{t}$  produced in association with their real counterparts annihilate to produce the Higgs boson. As it decays invisibly, it is the remaining  $t$  and  $\bar{t}$  in the event that lead to three classes of final state. The  $t$  quark decays almost exclusively to  $b W^+$  (and  $\bar{t} \rightarrow \bar{b} W^-$ ) [143]. In a resolved system where top quarks possess low to moderate momentum, the multitude of available  $b$ -tagging algorithms can distinguish the decays of the  $b$  quark. The products of the  $W$  boson are the determining factor of the final state. Hadronically-decaying  $W$ s ( $W \rightarrow q\bar{q}$ ), of course, produce pairs of jets. But they can also decay into a lepton and neutrino. The final states then all have  $\vec{p}_T^{\text{miss}}$  and up to two  $b$ -jets in common. Several jets may accompany them (the hadronic channel), or fewer jets with a single lepton (the *semi-leptonic* channel), or simply two leptons (the *dileptonic* channel). The magnitude and direction of the  $\vec{p}_T^{\text{miss}}$  in the latter two channels may be affected by the neutrinos, depending on their direction and energy.

In a boosted system where the top quarks have significant  $p_T$ , it is often difficult to tag  $b$ -jets, especially if one is searching in the hadronic channel. The decay products are not well separated and can merge into large, “fat” jets. Recently-developed algorithms can assist in this case by inspecting the substructure of these fat jets to classify, for example, boosted topologies originating from  $t$  quarks as well as  $V$  bosons to identify  $t\bar{t}H$  events.

### 2.3.1.3 Associated production from a vector boson ( $VH$ )

A Higgs boson is radiated by the vector boson  $V$  in the  $VH$  mechanism. Parallels can be drawn with  $t\bar{t}H$  as the decay of the  $V$  determines the search channel. Resolved and boosted systems are also possible. In the resolved case, a dijet pair with an invariant mass close to that of the parent boson would distinguish the hadronic channel.  $b$ -taggers can be exploited if the

---

<sup>3</sup>These coordinates are described in Chpt. 3.2.1.1.

decay is to a  $b$  quark, i.e.,  $Z \rightarrow b\bar{b}$ .<sup>4</sup> Single lepton channels are possible for  $WH$  and dilepton for  $ZH$ . For a boosted  $V$ , one expects the products to be collimated into a fat jet, at least in the hadronic channel. As with  $t\bar{t}H$ , one can take advantage of novel tagging algorithms to capture these scenarios.

### 2.3.1.4 Gluon-gluon fusion ( $ggH$ )

Despite  $ggH$  having the largest cross section of the four modes, its upper limit on  $\mathcal{B}(H \rightarrow \text{invisible})$  is the weakest. The Higgs boson is created through the loop-level fusion of the initial state gluons, normally mediated by a top quark since it has the largest coupling to the Higgs. With no additional final state particles to first order, searches for this production mode usually involve initial state radiation from the gluons or the loop. As such, the signature is at least one jet and  $\vec{p}_T^{\text{miss}}$ .

### 2.3.2 Results of previous searches

Many previous analyses have investigated the  $H \rightarrow \text{invisible}$  decay, in some cases from dedicated searches, but often as an afterthought or interpretation of the main analysis. VBF is the most sensitive production mode. This is demonstrated in Tab. 2.4 by the upper limits attained compared to the other mechanisms. In Ref. 139, a combination was performed by CMS over all the production modes detailed in the table (with the exception of  $t\bar{t}H$ ). Using the recent 2016 measurements as well as data taken from Run-1 and 2015, this combined observed upper limit sits at 19 %, while the expected is 15 %. With only data from Run-1 and 2015 (the previous combination) the observed and expected upper limits of 24 % and 23 %, respectively, were found [107]. Similar sensitivities were achieved for VBF and  $t\bar{t}H$  with the corresponding datasets by ATLAS [24].

Targeted mode	Analysis	Final state	Observed (%)	Expected (%)
VBF	Ref. 139	VBF-jets + $\vec{p}_T^{\text{miss}}$	33	25
$VH(V \rightarrow q\bar{q})$	Ref. 130	$V(\rightarrow q\bar{q}) + \vec{p}_T^{\text{miss}}$	50	48
$ggH$	Ref. 130	jets + $\vec{p}_T^{\text{miss}}$	66	59
$t\bar{t}H$	Ref. 83	$t\bar{t}(\rightarrow \text{jets}) + \vec{p}_T^{\text{miss}}$	85	73

Table 2.4: The most recent searches for invisibly decaying Higgs bosons in hadronic channels with 2016 data from CMS, and the achieved upper limits on the  $H \rightarrow \text{invisible}$  branching ratio at 95 % confidence level.

<sup>4</sup>Other potential decay modes, such as  $W \rightarrow bu$  and  $W \rightarrow bc$  are suppressed in the CKM matrix, yielding small production rates at the LHC.

## 2.4 Searches for semi-visible jets

Many searches for dark matter presume it is a WIMP-like particle because of the considerations discussed in Chpt. 2.2. In the LHC, the signatures of WIMPs would be driven by large missing transverse momentum recoiling from visible matter in the event. Monojet [105] and dijet [137] searches are able to exploit this, for example. However, no sign of WIMPs have been observed yet. Thankfully, a boundless supply of alternative theories exist, with possible signatures equally as varied. Though the  $\vec{p}_T^{\text{miss}}$  could still be one of the characteristics by which the dark matter can be inferred, a plethora of topologies and discriminating observables are possible. The dynamics that govern dark matter may be confined to a *dark sector* or *hidden sector*, inhabited by new forces and particles.

A dark sector may be largely inaccessible, as in some Hidden Valley scenarios,<sup>5</sup> but communicate with the visible sector through a portal interaction [142]. An example from SM particles could be the Higgs boson bridging the visible and hidden sectors, as mentioned in Chpt. 2.3. Many interesting and novel signatures can be probed by LHC experiments from models like these. Dark forces with energy scales in the tens of GeV and mediator masses up to several TeV may be accessible. If they share parallels with the standard model, the mechanisms for the dark matter presence and relic density can be explained as arising from a baryon-like asymmetry.

Proposed in Refs. 69 and 70, a strongly-coupled dark sector in a Hidden Valley is imagined with interactions analogous to QCD. Its internal dynamics are described by an SU(2) gauge group with dark gluons mediating the interactions, conserving a dark colour charge. The portals allowing the dark and visible sectors to communicate can be decomposed into a leptophobic  $Z'$  (*s*-channel) and bi-fundamental  $\Phi$  (*t*-channel) mediating a dark weak force. In the *t*-channel case,  $\Phi$  is a representation of both the visible and dark QCD gauge groups. Depictions of the processes above are given in Fig. 2.6. In the LHC, protons could collide at energies high enough to access the dark sector. From either the resonant production of a  $Z'$  or exchange of a  $\Phi$ , dark quarks  $\chi$  are produced. Below a dark confinement scale  $\Lambda_{\text{dark}}$ , hadronisation takes place to coalesce them into dark hadrons. Depending on the species, some of these dark hadrons are stable (i.e., a source of dark matter), while others are unstable and decay back into visible sector particles, namely standard model quarks. The final state is then a shower of two jets each interspersed with dark matter: semi-visible jets.

---

<sup>5</sup>A Hidden Valley is a schema where the standard model is extended by a non-abelian group. SM particles are uncharged under this group. The new, light particles from this extension are the opposite: charged under the new group and neutral under the SM gauge group. A heavy mediator carries both charges, acting as a portal between the standard model and Hidden Valley particles.

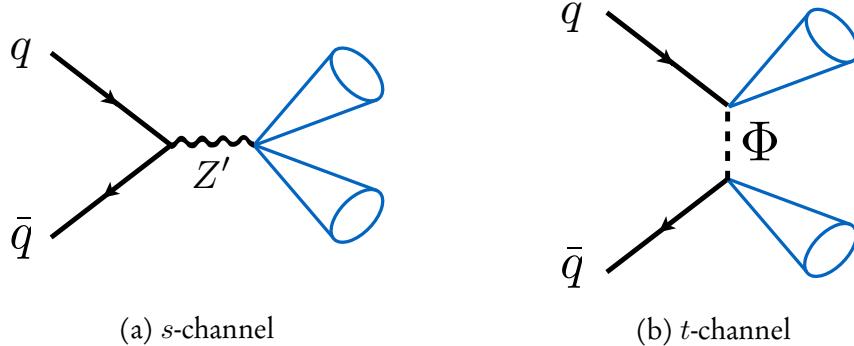


Figure 2.6: Example Feynman diagrams for the two main production modes of semi-visible jets. A  $Z'$  boson mediates the  $s$ -channel process while a bi-fundamental  $\Phi$  mediates the  $t$ -channel process. Figure from Ref. 70.

#### 2.4.1 Kinematics and free parameters of the model

The kinematics of semi-visible jets are heavily influenced by the following free parameters of the model: the mass of the mediator ( $m_{Z'}$  or  $m_\Phi$ ), the dark coupling strength ( $\alpha_{\text{dark}}$ ), the dark quark mass ( $m_\chi$ ), and the invisible fraction ( $r_{\text{inv}}$ ).

-  $m_{Z'}/m_\Phi$ : Since the energies of the colliding protons have an upper limit, the conservation of energy (or momentum) imposes one for the on-shell production/exchange of the mediator particle. In the  $s$ -channel process, production of the  $Z'$  is resonant. Consequently, its mass is possible to recover by calculating the dijet mass  $m_{jj}$  or transverse mass  $m_T$ .

- $\alpha_{\text{dark}}$ : In Ref. 70, this is defined as  $g_\chi^2/4\pi$  (where  $g_\chi$  is the coupling constant between the dark quarks and mediator). Analogous to QCD, the dark coupling runs as a function of the energy scale, influencing  $\Lambda_{\text{dark}}$ . At 1 TeV,

$$(2.12) \quad \Lambda_{\text{dark}} = 1000 \text{ [GeV]} \exp\left(\frac{-2\pi}{\alpha_{\text{dark}} b}\right)$$

where  $b = \frac{11}{3}N_c - \frac{2}{3}N_f$  is related to the number of dark colours and flavours, respectively.

- $m_\chi$ : This parameter does not directly affect much, but is related to the dark hadron mass ( $m_{\text{dark}} = 2m_\chi$ ) and  $\Lambda_{\text{dark}}$ . The combination of the two properties affects the shower dynamics. Note that while Ref. 70 describes some of these to be insensitive, a parameter scan over these two variables are necessary in the study described in Chpt. 5.

- $r_{\text{inv}}$ : This is defined as the fraction of produced invisible particles that remain stable, at least over timescales where they would interact with a detector. When generating simulated samples,  $r_{\text{inv}}$  can be interpreted as the *probability* of a dark hadron being stable.

While this variable is not inherent within the model, it is one that can parametrise many underlying components. As a result, visualisation of the shower and direction of  $\vec{p}_T^{\text{miss}}$  is much more intuitive, as demonstrated in Figs. 2.7 and 2.8, respectively. A large value of  $r_{\text{inv}}$  would yield a similar final state to a WIMP search.

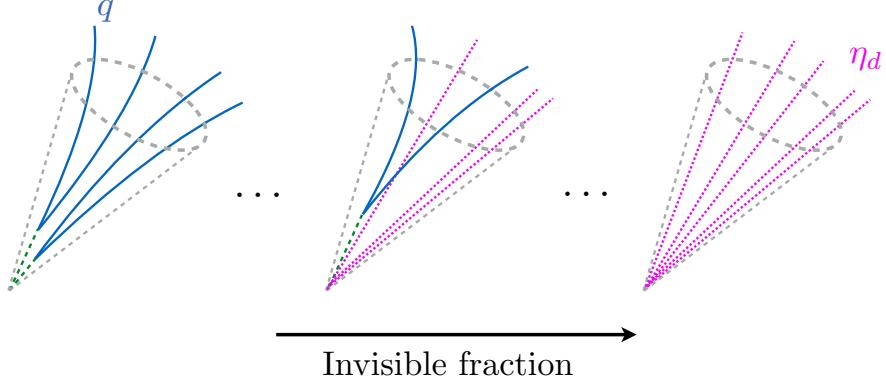


Figure 2.7: The constituents of a semi-visible jet as a function of its invisible fraction  $r_{\text{inv}}$ . The green dashed lines signify visibly decaying dark hadrons, blue the SM quarks, and pink the stable dark hadrons. Figure taken from Ref. 70.

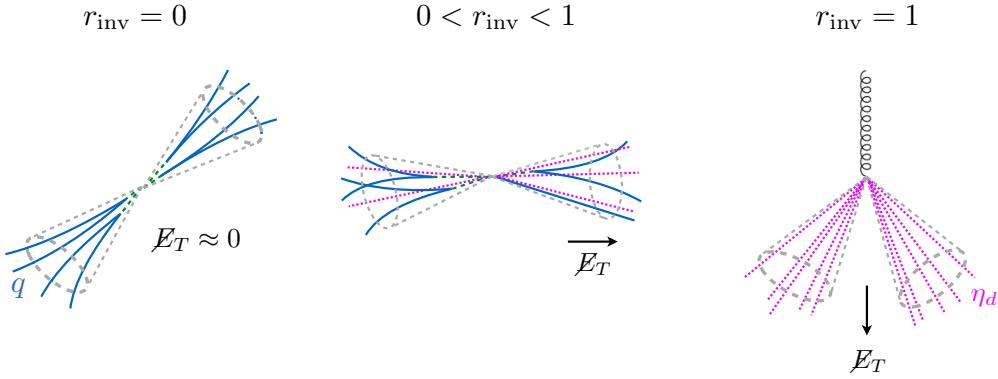


Figure 2.8: The typical direction of the missing transverse energy  $\cancel{E}_T$  (or  $\vec{p}_T^{\text{miss}}$ ) relative to the semi-visible jets as a function of their invisible fraction  $r_{\text{inv}}$ . The green dashed lines signify visibly decaying dark hadrons, blue the SM quarks, and pink the stable dark hadrons. Figure from Ref. 70.

In the studies of simulation for semi-visible jets in Chpt. 5, only the  $s$ -channel process has been analysed with LHC data with publication on the horizon. Generator studies have been additionally performed for the  $t$ -channel interaction and the analysis is underway. In the  $s$ -channel search, mediator masses of up to several TeV are accessible, and intermediate values of  $r_{\text{inv}}$  are most sensitive. Hence, the typical signature is a dijet pair with each jet

likely to contain a different invisible fraction, leading to the  $\vec{p}_T^{\text{miss}}$  aligned with one of the jets. WIMPs, on the other hand, completely recoil from the visible matter, and so jets may be more collimated with small separation. The  $\vec{p}_T^{\text{miss}}$  is also larger in magnitude and likely more isolated. The phase space exploited by this model is often rejected by dark matter searches since the final state can be easily mimicked by mismeasured QCD. A sizeable background from this process would therefore be present. However, jet substructure techniques and machine learning algorithms have developed rapidly in the recent years, and it is possible to disentangle signal and background with some certainty [119].

One interesting aspect of the model is the potential for signatures with displaced vertices, so called long-lived particles or *emerging jets* on account of the decay to visible states occurs a sufficient distance from the primary vertex. Some searches have already been performed for this final state from a different interpretation of a strongly-coupled dark force [140] to supersymmetry contexts [129]. These are not considered in Chpt. 5, so the dark hadrons are assumed to decay promptly. Long-lived interpretations have been noted as possible extensions to the search, however.

## THE LHC AND THE CMS EXPERIMENT

No one can whistle a symphony. It takes a whole orchestra to play it.

— H.E. Luccock

This chapter concerns the experimental setup. CERN, the European Organisation for Nuclear Research (*Organisation Européenne pour la Recherche Nucléaire*), is the body that manages the machines discussed and is a pioneer in the high energy physics community. Its most notable construction—the Large Hadron Collider—provides the CMS experiment with proton-proton collisions whose data is then stored, corrected, and finally used by physicists for analysis. Ranging from standard model precision measurements, searches for new physics, and the development of tools and algorithms to aid the previous two, these and much more are studied by the collaboration. CMS is described in detail, from its hardware and subdetectors to its data acquisition and trigger system. Special attention is given to the derivation of jet energy corrections (JEC) in the Level-1 Trigger (L1T) as a portion of this PhD has been devoted to them.

### 3.1 The Large Hadron Collider

Deep underground beneath the Franco-Swiss border lies the Large Hadron Collider (LHC), a synchrotron particle accelerator 27 km in circumference. As the largest machine in the world, the LHC stands as a testament to the importance of fundamental science and the dedication to which it is pursued. Predominantly a proton collider, lead and xenon ions

have also been injected for novel and unique studies. Four primary experiments are situated at their own interaction points where the two beams of particles are brought into contact: CMS (Compact Muon Solenoid), a general purpose detector with interests in precision measurements, searches for new physics, and many other avenues; ATLAS (A Toroidal LHC ApparatuS), a counterpart to CMS at its antipode on the LHC ring; LHCb, designed to study the decay of  $B$  hadrons; and ALICE (A Large Ion Collider Experiment), primarily studying heavy ion collisions and the quark-gluon plasma.

Four additional, smaller experiments are stationed in the LHC tunnel that are much more specialised than those aforementioned: TOTEM (TOTal Elastic and diffractive cross section Measurement) shares the CMS cavern with three subdetectors positioned near the beam line, performing proton structure and interaction cross section studies; LHCf shares the ATLAS cavity and is concerned with detecting neutral pions in the forward direction to explain the origins of high energy cosmic rays; FASER (ForwArd Search ExpeRiment) is another forward-based detector near ATLAS and searches for light, weakly interacting particles; finally, MoEDAL (Monopole and Exotics Detector At the LHC) is installed in proximity to LHCb, and aims to detect magnetic monopoles and other exotic particles.

The technical design report for the LHC can be found at Ref. 39, detailing the ring itself, infrastructure, general services, and the injector chain.

### 3.1.1 A proton’s journey

A proton destined for the LHC begins its journey as a hydrogen atom in a little, red bottle. Around  $3 \times 10^{14}$  protons are supplied to each beam in the LHC, with billions of refills available in this single container. Once the hydrogen atoms leave the source, they are stripped of their electrons, and propelled to the linear accelerator LINAC2.<sup>1</sup> This is the start of a long voyage through the accelerator complex (visualised in Fig. 3.1).

Given a modest boost to 50 MeV, the protons are sequentially fed from LINAC2 into the Proton Synchrotron Booster (PSB) that accelerates them further to 1.4 GeV. Another upsurge is provided once the protons travel to the Proton Synchrotron (PS), this time to 26 GeV. Then, the final energy increase received before entering the LHC comes from the Super Proton Synchrotron (SPS), leaving the protons at 450 GeV. Once injected into the LHC, they are finally accelerated to their peak: up to 6.5 TeV from a sequence of radio frequency cavities over the course of twenty minutes. The oscillation frequency of these cavities is precisely tuned and timed to give protons the appropriate kicks and accelerate them

---

<sup>1</sup>As of the end of Run-2, LINAC2 has been decommissioned. It has been replaced by LINAC4 in preparation for Run-3.

### 3.1. THE LARGE HADRON COLLIDER

to the desired energy. Since there is a distribution of proton energies in the beam, those that enter a cavity slightly out of time with a different energy than expected consequently become sorted into *bunches*. The remainder of the LHC ring is used to steer the beam with the aid of over 1,200 liquid helium-cooled superconducting dipole magnets. It is also focused by almost 400 equivalently-cooled quadrupole magnets to increase the rate of proton collisions.

One beam consists of a *train* of up to 2,808 bunches spaced 25 ns apart, each with 115 billion protons. Using bunches provides an advantage to the experiments at each of the interaction points. Discrete collisions take place between bunches in the opposing beams at 40 MHz as opposed to continuous streams of protons. This allows for estimates of pileup interactions that can be filtered out, and would otherwise introduce miscalculations of sums like  $p_T^{\text{miss}}$  and  $H_T$ . The data recorded by each experiment is naturally separated by the bunch crossing or *event*—an event being the products of the collisions in a bunch crossing, including any pileup interactions along with the nominal collision.

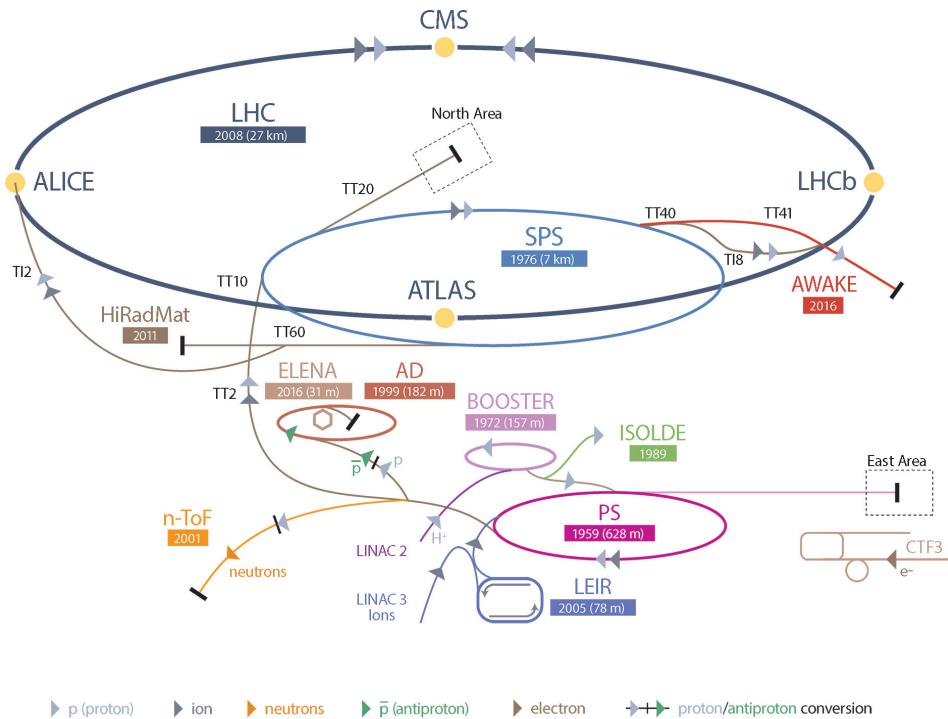


Figure 3.1: A schematic of the CERN accelerator complex. Various particles are shown from their sources to the interaction points enveloped by the detectors. Figure obtained from Ref. 126.

### 3.1.2 Luminosity

The *luminosity* of a particle accelerator is, along with centre of mass energy, a quantitative measure of its performance. It also often used to denote the amount of data delivered to, or collected by, the receiving detector. A full derivation of this quantity can be found in Ref. 93, and as such, only a summary is given here. The *instantaneous luminosity*  $\mathcal{L}$ , typically quoted in  $\text{cm}^{-2} \text{s}^{-1}$ , is defined as

$$(3.1) \quad \mathcal{L} = \frac{1}{\sigma_p} \frac{dN}{dt}$$

where  $N$  is the number of collisions and  $\sigma_p$  is the production cross section. In the LHC, because many final states are possible,  $\sigma_p$  is not generally known. It can be measured, however. A recent paper from LHCb measured the extrapolated inelastic cross section (which is the important component for LHC collisions) to be  $75.4 \pm 5.4 \text{ mb}$  at  $\sqrt{s} = 13 \text{ TeV}$  [5], allowing an estimate of the collision rate. For two colliding beams split into bunches with the constituents in each bunch distributed according to a gaussian profile,

$$(3.2) \quad \mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi \sigma_x \sigma_y} \cdot F$$

where  $N_1$  and  $N_2$  are the number of particles per bunch in beams 1 and 2, respectively,  $N_b$  is the number of bunches per beam,  $f$  is the revolution frequency (11.2 kHz for protons in the LHC), and  $\sigma_x$  and  $\sigma_y$  are the horizontal and vertical sizes of the beam, respectively.

The geometric reduction factor  $F$  is unity for beams colliding head on. However, in the LHC, the beams collide at an angle known as the *crossing angle* to reduce the effects of pileup. At the start of a fill when the number of protons in the LHC is at its maximum, the crossing angle at each interaction point is approximately  $300 \mu\text{rad}$ . As the runs progress, more protons have collided and the luminosity decreases due to the lower particle content. One of measures taken to recover luminosity and keep the LHC as efficient as possible is to reduce the crossing angle. By the end of the fill, it can be as small as  $240 \mu\text{rad}$ .

The instantaneous luminosity can be increased further by the inclusion of more bunches per beam, decreasing the size of a beam through improved quadrupole magnets, and by reducing the crossing angle. Tuning these parameters led the LHC to reach its design luminosity of  $10^{34} \text{ cm}^{-2} \text{s}^{-1}$  in 2016. It had more than doubled by the end of Run-2.

Integrating the instantaneous luminosity over a period of time yields the *integrated luminosity*  $\mathcal{L}_{\text{int.}}$ :

$$(3.3) \quad \mathcal{L}_{\text{int.}} = \int \mathcal{L} dt$$

These values are often quoted by experiments in units of *inverse femtobarn* ( $\text{fb}^{-1}$ ) as an indicator of the amount of collision data collected. A barn is a unit equal to  $10^{-28} \text{ m}^2$  and widely used to express cross sectional area in nuclear and particle physics. Thus its reciprocal, the same units as integrated luminosity, gives a good sense of scale that relates cross sections of individual processes to the total amount of data. Tab. 3.1 gives the integrated luminosities over Run-2 delivered by the LHC and collected by CMS. Taking  $139 \text{ fb}^{-1}$  from this table, and assuming the inelastic  $pp$  cross section above gives an estimate of  $10^{16}$  collisions recorded by CMS over Run-2 (excluding pileup interactions).

### 3.1.3 Pileup

At the high instantaneous luminosity of the LHC, multiple interactions per bunch crossing—known as *pileup*—are frequent. Due to all the factors designed to increase the luminosity, many “softer” collisions take place amidst the hard scatters. Pileup interactions can produce many low- $p_T$  objects, and near-collinear to the beam (detecting particles in this region is essential for some analyses). These additional objects can contaminate the reconstruction of the final state from the nominal collision, and as such affect the direction and magnitude of the  $\vec{p}_T^{\text{miss}}$ . When one aims to increase the luminosity to collect more data, the more severe pollution from pileup is an inevitable consequence. The distribution of pileup events in CMS during Run-2 can be seen in Fig. 3.2.

There are strategies in place in the LHC and the anchored experiments to mitigate pileup. Introducing, or increasing, a crossing angle between the beams reduces the effect. Decreasing the crossing angle increases the overlap between bunches at each intersection—since they have a non-negligible length—leading to more collisions. Track reconstruction algorithms are very efficient at connecting particles to their primary vertices, and therefore particles originating from pileup vertices can be identified and removed [1, 45]. At CMS, various methods are available in the calorimeter triggers to subtract the effects of pileup.

### 3.1.4 Evolution of the LHC

The LHC began operating in 2010 at a centre of mass energy of  $\sqrt{s} = 7 \text{ TeV}$ ;  $3.5 \text{ TeV}$  per beam. A modest increase to  $8 \text{ TeV}$  was achieved for 2012 until the end of Run-1 in 2013. Long Shutdown 1 then commenced where maintenance work was carried out, and upgrades to the accelerator and experiments were performed. Notably, the superconducting magnets were improved to safely handle much more energetic beams.

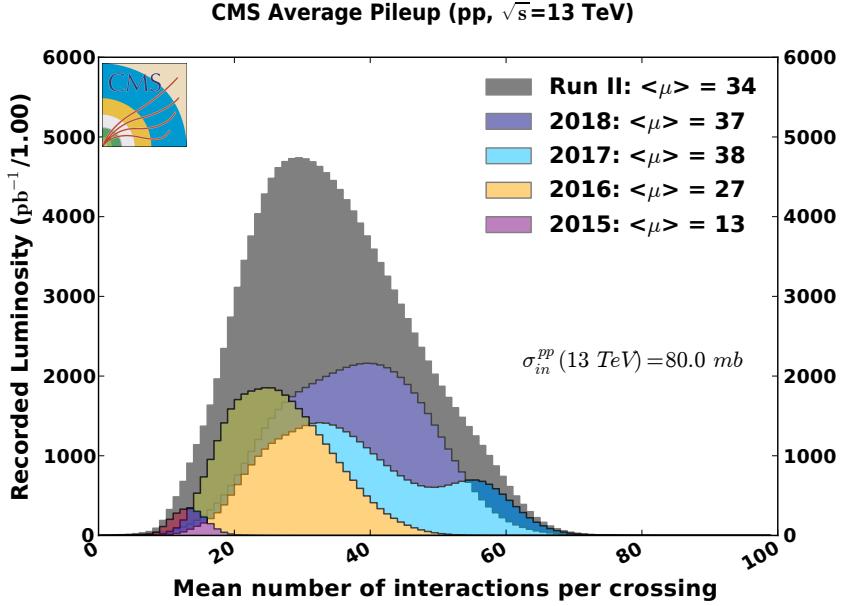


Figure 3.2: The average number of pileup interactions per bunch crossing at CMS during Run-2 of the LHC. Figure obtained from Ref. 53.

The LHC resumed operation in 2015, marking the start of Run-2 and further pushing the frontiers of high energy physics with a centre of mass energy of  $\sqrt{s} = 13$  TeV. While valuable data was taken that year, it was not until 2016 when substantial datasets were amassed. Therefore, when analysts (and this thesis) refer to the “full Run-2 dataset,” they typically do not include data from 2015. It requires as much care as the other years with respect to implementing corrections and addressing systematic uncertainties, and so on, for only a very small gain in integrated luminosity. Run-2 ended in 2018 with—omitting the 2015 dataset— $158.6 \text{ fb}^{-1}$  of  $pp$  collisions delivered,  $146.5 \text{ fb}^{-1}$  of which were recorded by CMS who certified  $137.2 \text{ fb}^{-1}$  suitable for analysis [53, 59]. A breakdown by year is presented in Tab. 3.1 with a visual chart in Fig. 3.3. By comparison, the LHC delivered only  $6.2 \text{ fb}^{-1}$  at 7 TeV and  $23.3 \text{ fb}^{-1}$  at 8 TeV over the course of Run-1.

Integrated luminosity	2015	2016	2017	2018	Full Run-2
Delivered by LHC ( $\text{fb}^{-1}$ )	4.2	41.0	49.8	67.9	162.9
Recorded by CMS ( $\text{fb}^{-1}$ )	3.8	37.8	45.0	63.7	150.3
Certified by CMS ( $\text{fb}^{-1}$ )	2.3	35.9	41.5	59.7	139.5

Table 3.1: The integrated luminosity delivered by the LHC during Run-2 which were recorded and certified by CMS. Typically, only the 2016–18 datasets are analysed since 2015 accrued little data compared to the other years. Numbers obtained from Refs. 53, 59.

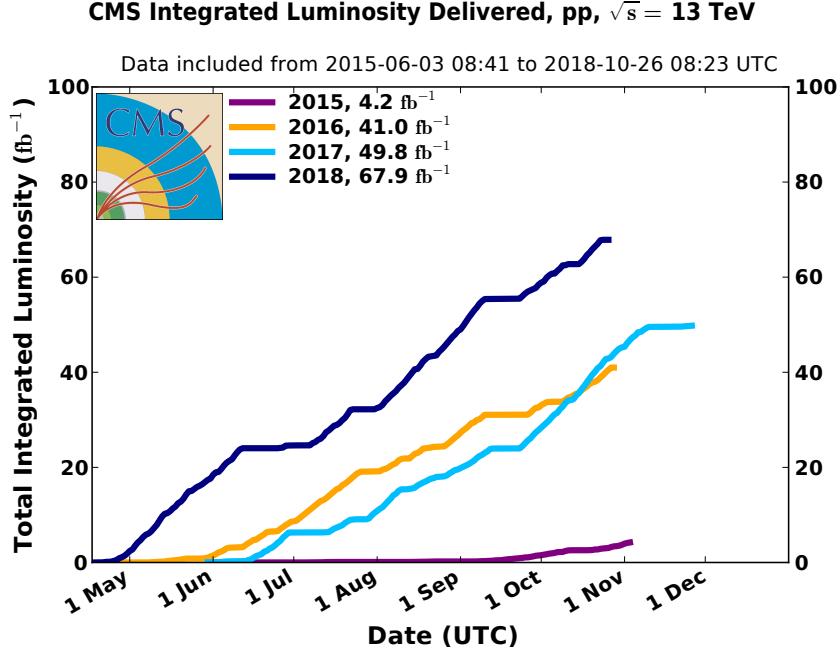


Figure 3.3: The integrated luminosity of  $pp$  collision data delivered to CMS during Run-2 of the LHC. Figure obtained from Ref. 53.

In addition to the upgrades that have so far been performed, a much grander change is on the horizon. In 2027, the High Luminosity Large Hadron Collider (HL-LHC) improvements are set to be introduced with the primary purpose of increasing the luminosity by up to a factor of ten: more powerful magnets for focusing the beam, crab cavities for increasing the bunch overlap area, shielding and more radiation-resistant components, an upgrade to many of the boosters in the accelerator complex, and much more. A summary of the important parameters at three stages of the LHC’s life is given in Tab. 3.2.

Parameter	End of Run-2	HL-LHC	Design
Beam energy (TeV)	6.5	7	7
Bunches per beam	2,556	2,748	2,808
Protons per bunch	$1.5 \times 10^{11}$	$2.2 \times 10^{11}$	$1.5 \times 10^{11}$
Crossing angle ( $\mu\text{rad}$ )	300	590	285
Instantaneous luminosity ( $\text{cm}^{-2}\text{s}^{-1}$ )	$2.1 \times 10^{34}$	$7.2 \times 10^{34}$	$1.0 \times 10^{34}$
Integrated luminosity per year ( $\text{fb}^{-1}$ )	68	250	40
Average pileup per bunch crossing	34	140	27

Table 3.2: Some of the important parameters defining the performance of the LHC at the end of Run-2, for the High Luminosity Large Hadron Collider, and the design specification. Several values in this table were obtained from Refs. 17, 33.

## 3.2 The CMS experiment

The Compact Muon Solenoid (CMS) Collaboration was formed in 1992 with a letter of intent circulated to the high energy physics community for a general purpose detector to be built at the LHC [74]. Though its emphasis was the detection of the Higgs boson, the myriad avenues for which physics could be studied was also acknowledged. Just under five hundred people from sixty two institutes were part of the original letter. At the time of writing over five thousand physicists, engineers, technical and administrative staff, and students, spanning over two hundred institutes in fifty countries, comprise the collaboration. Working tirelessly in harmony, with ingenuity, cooperation, and a drive for exploring the frontiers of particle physics, almost one thousand publications have been produced with data collected by the experiment. These range from exotic/BSM searches, standard model precision measurements,  $B$ -physics, top quark physics, heavy ions, and more. This wide scope of topics cements CMS as a versatile and world-leading experiment, and collaboration as a whole.

### 3.2.1 The CMS detector

CMS itself lives a hundred metres underground at Interaction Point 5, just outside the town of Cessy in France. Constructed over several years, the detector was separated into fifteen sections. Each was built on the surface and lowered into the experimental cavern to assemble the 14,000 tonne goliath. As a hermetic cylinder enclosing the LHC beam pipe, the machine is 21 m in length (plus a little extra from the bolt-on forward calorimeters) and 15 m in diameter. Its etymology stems from how *compact* (relatively speaking) the detector is since its weight is twice that of ATLAS in only one sixth of the volume, it has a subsystem dedicated to detecting *muons*, and a *solenoid* is employed to generate the magnetic field.

A kaleidoscope through we which see the subatomic, CMS is designed to detect all manner of particles as accurately as possible. To accomplish this, the detector is divided into four major subsystems (or subdetectors): the silicon tracker, the electromagnetic calorimeter (ECAL), the hadron calorimeter (HCAL), and the muon chambers. These are each explained in more detail below, with a graphic of the entire detector presented in Fig. 3.4 and a transverse slice through it in Fig. 3.6.

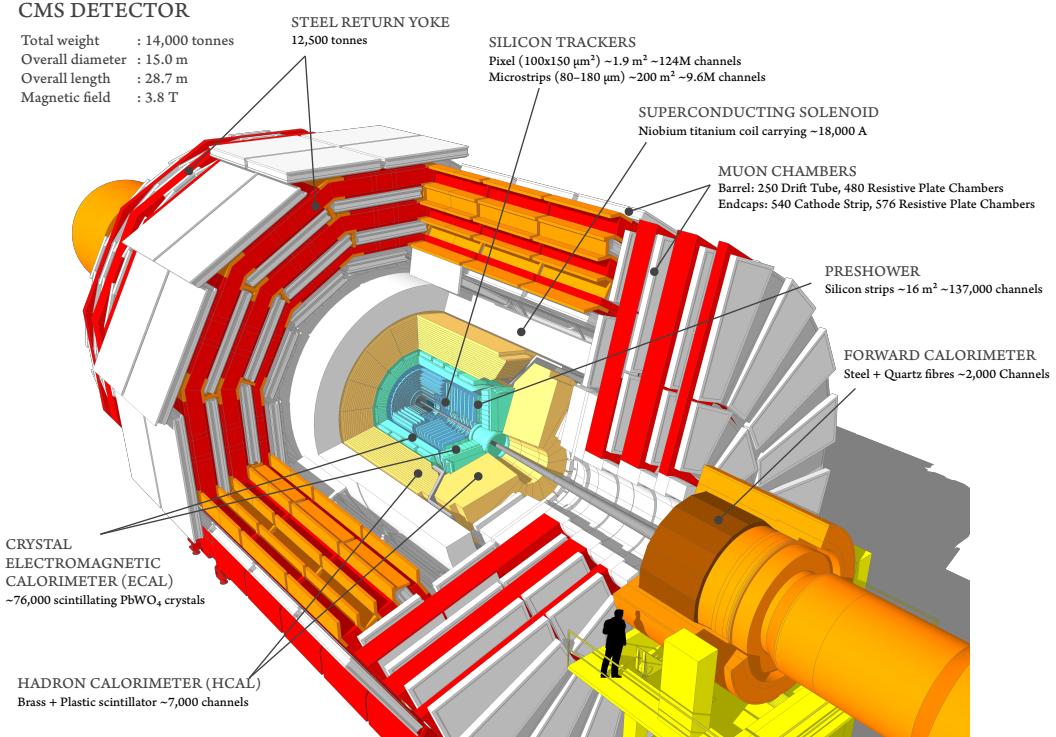


Figure 3.4: A cutaway diagram of the CMS detector with all of the principal components labelled. This detector configuration was used for the 2017–18 data taking years, where the coverage of the pixel detectors in the silicon tracker were upgraded. Image taken from Ref. 121.

### 3.2.1.1 Geometry

Collider physics tends to use certain conventions when describing the positions of particles in a detector. The azimuthal angle  $\phi$  is the same variable as in cylindrical coordinates with a range  $[-\pi, \pi]$ . Pseudorapidity, denoted by  $\eta$ , is a coordinate that describes the angle between a particle and the longitudinal axis of the detector (the beam line):

$$(3.4) \quad \eta = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right]$$

where  $\theta$  is the angle between the particle's three-momentum  $\vec{p}$  and the positive direction of the beam axis. So as  $\theta \rightarrow 0^\circ$ ,  $\eta \rightarrow \infty$ . Generally, particles with large  $\eta$  escape the detector which is why forward calorimeters are in place. The transverse momentum of a particle can be found with  $p_T = |\vec{p}| / \cosh \eta$ . The convention in hadron colliders is to use  $\eta$  over the rapidity  $y$ , where the latter is Lorentz invariant and tends to the former in the limit  $|\vec{p}| \gg m$ .

However, in cases such as jet clustering, soft particles are included and so  $y$  is required:

$$(3.5) \quad y = \frac{1}{2} \ln \left( \frac{E + p_L}{E - p_L} \right)$$

where  $p_L$  is the longitudinal component of the momentum. The subdetectors of CMS are nominally separated into several sections dependent on their geometry or layout, and it is useful to divide them into  $\eta$  ranges to demonstrate their coverage. *Barrel* sections are cylindrical around the beam line while *end caps* are usually discs or plates perpendicular to it. The following subsections describe them in detail. One quadrant of CMS showing the  $\eta$  divisions can be seen in Fig. 3.5.

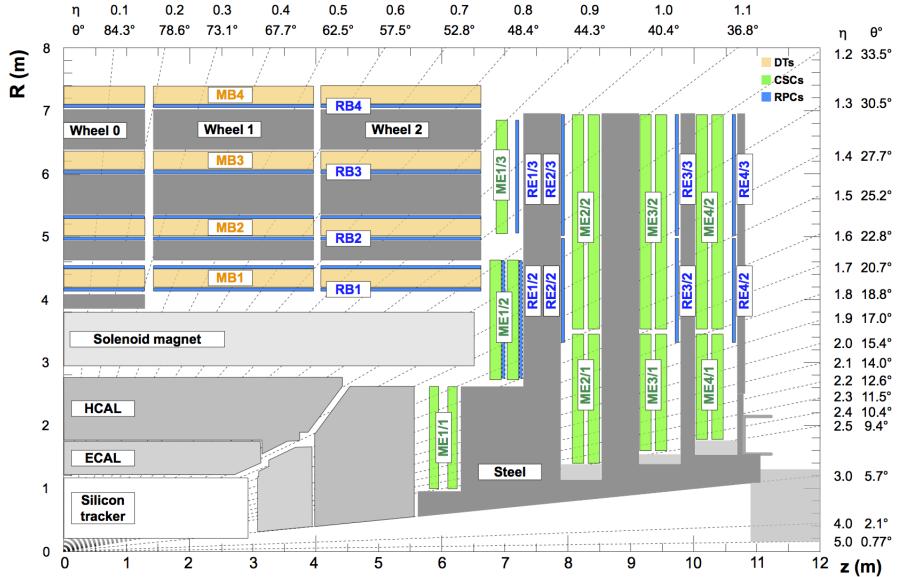


Figure 3.5: A quadrant of the CMS detector illustrating the main subsystems with their radius  $R$ , longitudinal distance  $z$ , and pseudorapidity  $\eta$  from the interaction point. The grey box at  $3 < |\eta| < 5$  and  $11 < z < 12$  m is the hadron forward calorimeter. Figure taken from Ref. 115.

The distance between two objects can be found with the variable  $\Delta R$ :

$$(3.6) \quad \Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

It is often used as a parameter in jet clustering to define the enclosing cone, or between reconstructed objects for assessing overlap.

### 3.2.1.2 The tracker

When a collision occurs, the ejected particles first encounter the silicon tracker that extends from 4 cm to just under 1 m in radius [68, 102]. Longitudinally, it is confined to  $|\eta| < 2.5$ ,

known as the *central region*. The barrel and end cap sections are bound by  $|\eta| \lesssim 1.6$  and  $1.6 \lesssim |\eta| < 2.5$ , respectively.

Consisting of tens of millions of pixels and microstrips, the positions of particles can be recorded to within  $10\text{ }\mu\text{m}$ . Their transverse momenta can be measured from the curvature of the track from the magnetic field induced by the solenoid (see Chpt. 3.2.1.5). During Run-1, the momentum resolution in the central region was 0.7 % for  $1\text{ GeV}/c$  particles and 5 % for those of  $1\text{ TeV}/c$ . Initially, the pixel detector was arranged in three layers, covering  $1\text{ m}^2$  and consisting of 66 million channels. It was upgraded for the 2017–18 run period to introduce a fourth layer, increasing its coverage to  $1.9\text{ m}^2$  with 124 million channels. The many layers of strip modules enclose the pixel detector. Since the tracker is the closest subsystem to the interaction point, the components must be extremely radiation-tolerant to withstand the bombardment of particles during every collision.

Track-finding algorithms use pattern recognition to determine whether the hits from multiple layers in the tracker can be correlated and reconstructed to originate from individual particles. Each is suited to different types of track,  $p_{\text{T}}$  threshold, and location. Vertex reconstruction algorithms are employed to determine the primary vertex. Any secondary vertices transverse (for example, in  $b$ -jets) or longitudinal (pileup) to the beam line may also be distinguished. One of the algorithms, the gaussian sum filter (GSF) method, is applied extensively in the HLT for triggering on electrons and photons [16].

### 3.2.1.3 The ECAL

After the tracker, the particles can interact with the ECAL [36, 65]. The 76,000 lead-tungstate crystals scintillate the incoming light, where the attached avalanche photodiodes record energy deposits in the barrel region ( $|\eta| < 1.48$ ) belonging to charged particles and photons. Vacuum phototriodes are instead glued to the crystals in the end caps ( $1.48 < |\eta| < 2.96$ ) as they are subject to higher radiation doses. A lead-silicon preshower detector precedes the end cap crystals to improve spatial resolution and discrimination of multiple particles. Electronics for groups of  $5 \times 5$  crystals (the ECAL component of a *trigger tower*) read out the data and deliver it to the trigger system (see Chpt. 3.2.2).

The ECAL is designed to identify charged particles (primarily electrons) and photons. The  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ \rightarrow 4\ell$  processes were kept in mind as an ardent focus was on discovering the Higgs boson. These particles are usually stopped entirely in the subdetector. Charged hadrons, muons, and tau leptons ( $\tau$ ), however, only deposit a sliver of energy in the ECAL, and so information from the other subdetectors is required in their cases. Measuring the energies of electrons and photons is complex and corrects for many factors to maintain a

high precision. Among them are a cluster correction (as several hits in a cluster can belong to one object, especially when they electromagnetically shower), a time-dependent response correction from the laser monitoring system that measures the crystal transparency, and an intercalibration correction that accounts for the non-uniform response of the crystals and photodetectors. The energy resolution in the ECAL  $\sigma_E^{\text{ECAL}}/E$  can be quantified as

$$(3.7) \quad \left( \frac{\sigma_E^{\text{ECAL}}}{E} \right)^2 = \left( \frac{S}{\sqrt{E}} \right)^2 + \left( \frac{N}{E} \right)^2 + C^2$$

where  $E$  is the energy in GeV,  $S$  is a stochastic term,  $N$  represents noise in the detector electronics, and  $C$  a constant. For electrons in the barrel during Run-1, the terms were measured to be  $S = 0.028$ ,  $N = 0.12$ , and  $C = 0.003$  [98], yielding a resolution of 0.5 % for a 100 GeV electron.

### 3.2.1.4 The HCAL

Hadronic particles propagate past the ECAL into the HCAL [66]. The barrel ( $|\eta| < 1.30$ ) and end cap ( $1.30 < |\eta| < 2.96$ ) regions both consist of brass absorber<sup>2</sup> and plastic scintillator with wavelength-shifting fibre. Arranged in wedges in the barrel and discs in the end cap, hybrid photodiodes record the signal in either case. This yields 7,000 detection channels. Due to colour confinement, hadrons are much more prone to showering than charged leptons. Additional scintillator tiles are placed outside the yoke of solenoid in the barrel region to capture the tails of these showers.

Unlike the ECAL which only covers the barrel and end cap, the HCAL also has a component in the forward region of the detector: the hadron forward calorimeter (HF). Composed of steel absorber and quartz fibres with 1,000 channels on each side, and occupying  $3.0 < |\eta| < 5.0$ , it is designed to detect particles with trajectories close to the beam line. The additional function of measuring the luminosity delivered to, and collected by, the experiment is also served. Cherenkov light kindled in the fibres is funnelled into photomultiplier tubes that collect the signal. Particles that enter the HF are usually not supplemented by tracker or ECAL information. As such, the energy and momentum of particles can be determined only by the subdetector.

Each cell of the HCAL maps onto a given number of ECAL crystals, forming trigger towers or calorimeter towers that are important when collecting data (see Chpt. 3.2.2). The towers project mostly-radially from the interaction point. Their granularity is greatest at low

---

<sup>2</sup>Famously, a large component of the brass used to build the end caps came from over 1 million decommissioned Russian naval artillery shells left over from World War II.

$\eta$ , and decreases with increasing  $\eta$ . Information from both subdetectors is utilised to identify jets, muons, and taus, as well as any electrons or photons that penetrate through the ECAL. Readout electronics are mounted to the various components of the HCAL to deliver to the trigger system. The response and transparency of modules is corrected for in a similar way to the ECAL, as is intercalibration.

The hadronic energy resolution for the combined HCAL and ECAL  $\sigma_E^h/E$  in the barrel section can be estimated as

$$(3.8) \quad \left( \frac{\sigma_E^h}{E} \right)^2 = \left( \frac{S}{\sqrt{E}} \right)^2 + C^2$$

where, as in Eq. 3.7,  $S$  and  $C$  are stochastic and constant terms, respectively. In Run-1, the terms were measured to be  $S = 0.847$  and  $C = 0.074$  in the barrel with similar performance in the end caps [63]. In the HF,  $S = 1.98$  and  $C = 0.09$ .

### 3.2.1.5 The magnet and muon chambers

The three subsystems described above are encased in a massive superconducting solenoid [67], making up 12,500 of the 14,000 tonne detector. It generates a magnetic field of 3.8 T and  $2.4 \times 10^9$  J of energy, making it the largest store of energy in the world for a single magnet. The field allows for the transverse momentum of a charged particle to be precisely measured, since the field propels it into a circular orbit. The magnetic force  $qvB$  exerted is equal to the centripetal force  $mv^2/r$ , and therefore the radius of the orbit is proportional to the momentum.

An iron return yoke interspersed with the muon chambers [111] constitute the rest of the detector, extending from a radius of around 3 m to 7.5 m. The iron yoke confines the magnetic field to the volume of the detector and provides a 2 T field in the opposite direction to that found farther inward. Since muons are much heavier than electrons and have high  $p_T$ , they penetrate farther than the bounds of the ECAL and deposit little energy in the process. The final states of several interesting decays include muons, such as with the Higgs boson. Constructing a dedicated subdetector is therefore essential.

There are four muon stations, as can be seen by the orange strips in Fig. 3.6, separated by layers of the return yoke. 480 resistive plate chambers and 250 drift tubes reside in the central barrel region of  $|\eta| < 1.2$ . The end caps are populated with 576 resistive plate chambers, complemented by 540 cathode strip chambers. Its limits are  $0.9 < |\eta| < 2.4$ , though the resistive plate chambers terminate at  $|\eta| = 1.9$ . Like many of CMS' subdetector components, they are designed parallel and perpendicular to the beam line, not according to  $\eta$ . With such

large segments at high radius in the muon chambers, a small overlap is therefore present. This  $0.9 < |\eta| < 1.2$  sector is sometimes referred to as the *overlap region*. The pseudorapidity range of the muon chambers' end cap is restricted compared to the ECAL and HCAL to (mostly) align with the tracker, and the steel that surrounds the structure takes up the remainder of the  $\eta$  that the former two occupy.

All of the detectors in the muon chambers take advantage of gaseous ionisation, as it is much less expensive than the silicon used in the tracker while still providing excellent resolution. If a muon enters a cell in a drift tube or cathode strip chamber, its position can be measured by the length of time taken for the gas it ionises to drift to the anode wire of said cell in the presence of a strong electric field. The latter can additionally utilise the induced charge on the cathode strips for improved spatial resolution. The resistive plate chambers work on a similar principle, but use parallel plate capacitors and are predominantly used for timing purposes when the data is sent to the trigger system. To measure the momentum of the muons, information from both the tracker and muon detector are used where possible. In Run-2, for muons with  $p_T < 100 \text{ GeV}$  the momentum resolution was 1 % in the barrel and 3 % in the end caps. Using cosmic ray muons for high  $p_T$  measurements, a resolution of 6 % was found for 1 TeV muons in the barrel [134].

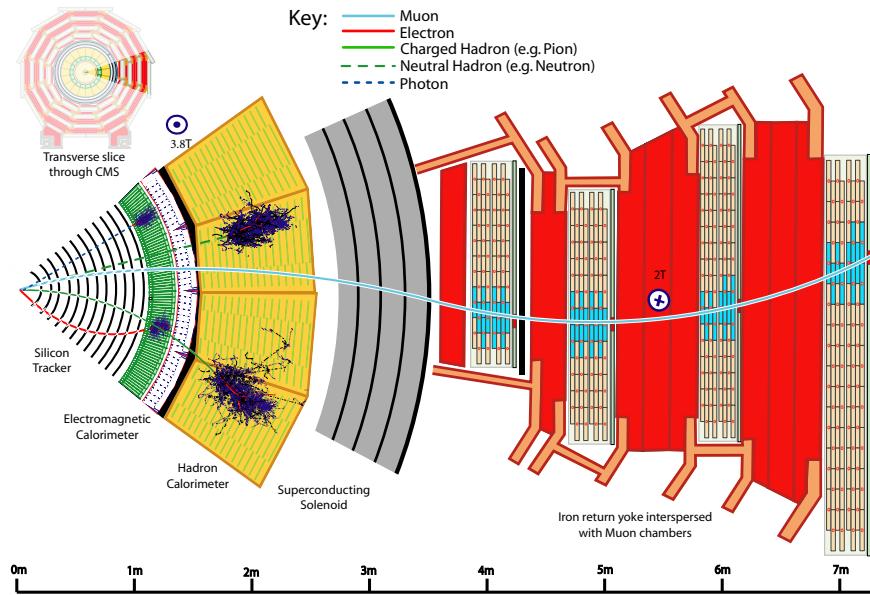


Figure 3.6: A transverse slice through the barrel section of the CMS detector with the main subsystems and components visible. Several particles produced at the primary vertex and their interactions with the detector are also depicted. Figure obtained from Ref. 128.

### 3.2.2 Data acquisition and triggering

With the enormous collision rate at the CMS interaction point, acquiring data requires some thought and ingenuity. Today’s electronics cannot handle the bandwidth from recording every single collision,  $\mathcal{O}(1 \text{ petabyte/s})$ . As such, a *trigger* is used to select the events that may be of use to analysers. CMS incorporates a two stage trigger [31]: the Level-1 Trigger (L1T), implemented in the detector hardware; and the High-Level Trigger (HLT) [47], a software farm to further reduce the events selected at Level-1. The trigger is part of the larger data acquisition (DAQ) system. An intricate network of custom electronics and commercial processors—a union of hardware, firmware, and software—are interconnected by multi-gigabit links to record the products of the highest-energy manmade collisions on Earth.

#### 3.2.2.1 The Level-1 Trigger

The Level-1 Trigger is a set of algorithms (a trigger menu) implemented in custom hardware designed to reduce the event rate from 40 MHz to a maximum of 100 kHz. FPGA and ASIC chips contain the algorithms in firmware, with timing systems synchronised with the LHC clock. When a collision occurs, particles interact with the detector and hits are registered by the components.

Coarsely-segmented data is read out from the ECAL and HCAL through a two-layer Calorimeter Trigger. These are arrays of custom processors located at Point 5. Layer-1 receives the calorimeter data from upwards of one thousand fibre optic links, each with multi-gigabit throughput. The information from the two subsystems are combined into calorimeter towers, and some simple position- and energy-dependent calibrations are applied. The data from Layer-1 is then transmitted to Layer-2, again over many high-bandwidth optical links. Here, physics object candidates are identified: jets,  $e$ ,  $\gamma$ , and  $\tau$ .<sup>3</sup> Additional calibrations are applied to them (for example, in Chpt. 3.2.4), and simple pileup subtraction is performed. Energy sums are also calculated at this stage, such as  $E_T^{\text{miss}}$ ,  $H_T$ , and  $H_T^{\text{miss}}$ .

In parallel, the various subdetectors in the muon chambers pass information through successive stages, and are then combined with some of the calorimeter data in a sorting/merging/isolation layer. The output from this layer is combined with the remaining information within Layer-2 in the Global Trigger, a series of  $\mu$ TCA boards with FPGAs. The trigger menu lives here, and all of the data gathered—object candidates, energy sums, beam conditions—is provided to it. These triggers may be dependent on the presence of a single object or the number of objects of a single class (e.g., one muon, two jets), multiple classes

---

<sup>3</sup>An overview of the latest Run-2 algorithms for object identification can be found in Ref. 148.

of object (cross triggers), the energy sums, the topologies of objects, and more. The latency given to make a decision on whether to keep or reject an event is 4  $\mu\text{s}$ . A diagram of this data flow is given in Fig. 3.7.

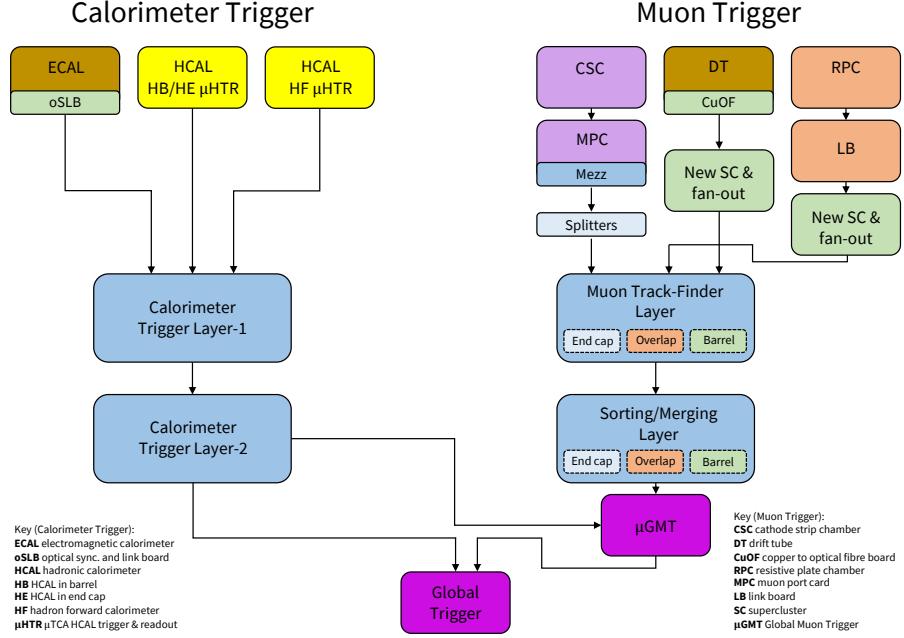


Figure 3.7: A summary of the CMS Level-1 Trigger data flow from the hits recorded by the subsystems to the Global Trigger. Figure reproduced from Ref. 75.

A small subset of data from the L1T is diverted for monitoring purposes. During data taking periods, CMS has a plethora of members that participate in the supervision and maintenance of the experiment. At Point 5, shifters monitor the systems in real time, collecting information from various sources. Data acquisition, data quality, access to the experimental cavern, the L1T, and the HLT are among them. Experts in these, and more specific subsystems, are on call on a rotating period. My responsibilities have included in-person shifts for the Level-1 Trigger and on call expertise for Layer-2 of the Calorimeter Trigger.

### 3.2.2.2 The High-Level Trigger

Events that pass a logical OR of the triggers at Level-1 are transmitted to the HLT. The higher resolution data collected at the point of collision is available, along with information from the tracker. Populated with Intel Xeon processors (high core count CPUs), approximately 22,000 cores are available (by the end of Run-2) to process the data sent from the L1T. In order to avoid a backlog, a 100 kHz input rate allows a HLT node  $\sim$ 220 ms to make a decision.

High-level software in the cmssw environment (written in C++ and Python) is executed. A larger and more complex trigger menu is available, including the possibility of analysis-specific triggers (such as those that target VBF topologies in Chpt. 6). Complex variables such as  $\alpha_T$  [129] can even be calculated and triggered on.

Physics objects are reconstructed further, with algorithms such as anti- $k_T$  leveraged to cluster jets [41]. Using a distance parameter—related to the radius of the cone surrounding the hadrons—of 0.4 denotes a jet as AK4, while fat jets are clustered with the value of 0.8 and are known as AK8. Additional classification algorithms are applied to objects, such as the DEEPCSV neural network [132] to identify  $b$ -jets. A global event reconstruction from the PARTICLE FLOW algorithm (PF) [62, 128] is performed as well. Some of these algorithms can be computationally expensive. Consequently, only approximations/parametrisations are used at HLT level. The full-scale versions are re-run on the retained events in later stages of postprocessing.

The HLT reduces the event rate from the maximum 100 kHz input substantially to around 1 kHz. The data stream of  $\mathcal{O}(6 \text{ GB s}^{-1})$  is then subject to further processing before the analysts access it. As well as the data being stored on networked hard drives at sites across the globe, backups are made to magnetic tape for long term storage. As with the Level-1 Trigger, some data is redirected for monitoring, object calibrations, and alignment of detector components.

### 3.2.2.3 Data reduction and compression

Several data tiers are used in CMS to serve different purposes. All sorted in ROOT file containers,<sup>4</sup> RAW data is repacked straight from prompt reconstruction. The files are then converted to the RECO tier where objects are fully reconstructed. Files are large (around 1.3–1.4 MB per event) and usually only kept for a short period of time for detector-related studies. The subset designed for analysis is designated Analysis Object Data (AOD). However, the level of compression is not ideal, especially with the volume of data amassed in Run-2. This is why miniAOD was developed. By *slimming* the trees in the files (removing unnecessary branches) and using smaller numeric data types, the footprint is roughly 10–15 % that of AOD.

An even more drastic reduction was made possible by the introduction of the nanoAOD tier. The methods for shrinking AOD to miniAOD are applied more aggressively, and a flat tree structure (with only leaves, and no nested levels) is utilised for simpler access to data. This allows for much smaller file sizes  $\mathcal{O}(1\text{--}2 \text{ kB/event})$ . Exporting to other data structures

---

<sup>4</sup>Event and ancillary data are stored in tree structures (otherwise known as *ntuples*), with *branch* and *leaf* levels that contain more specific quantities such as muon  $p_T$  or jet  $\eta$ .

for integration with external libraries, such as `numpy` and `pandas`, is also achievable. While nanoAOD is unable to cover all analyses because of its reduced event content, it is suitable in the search for invisibly decaying Higgs bosons.

### 3.2.3 Simulating CMS data

Data recorded by CMS is paramount for analyses searching for new physics. However, simulated samples are also of high importance. Events for specific processes are generated using Monte Carlo (MC) random sampling, and the output datasets are often collectively referred to by the method—*Monte Carlo* or *MC*. The datasets are often generated with large numbers of events to minimise the associated statistical uncertainty. MC samples are useful in a variety of cases: understanding the kinematics of signal processes in searches for new physics, modelling background processes that can mimic signal, and comparisons to data for validation purposes.

A matrix element generator such as `MADGRAPH` [14] or `POWHEG` [85, 116] models the hard scattering process, usually at leading order (LO) but sometimes at higher orders. Events then pass through a hadroniser (usually `PYTHIA` [141] in CMS) to model hadronisation of quarks and gluons, sometimes known as the *parton shower*—the softer radiation that accompanies the hard scatter. Jets are clustered here by, for example, the anti- $k_T$  algorithm. The particles are also run through a detector simulation that emulates the configuration and response of the detector from a given period. Material interactions and emulation of the triggers are included. `GEANT4` [10] provides this in CMS. Once the particles have been appropriately simulated, they are given the same postprocessing treatment as actual data, such as executing object-tagging algorithms so that the data and simulated samples are as comparable as possible.

### 3.2.4 Jet energy corrections in the Level-1 Trigger

Recording the properties of hadrons that are amalgamated into jets is not always consistent across the detector. While the components go through quality control, there is inevitably some variation in their performance. They can degrade at different rates. Some may also receive hits more often than others and be subject to greater radiation damage. As a result, non-uniformity of the detector response—as functions of  $p_T$  and  $\eta$ —must be compensated for. For jets, this comes in the form of jet energy corrections (JEC).

As outlined in Chpt. 3.2.2.1, the trigger primitives from the ECAL and HCAL enter Layer-1 of the Calorimeter Trigger, where coarse position- and energy-dependent calibrations are applied. Objects such as jets are initially identified in Layer-2, and preliminary calibrations

correct their energy. Disregarding these, even at this early stage in the data acquisition workflow, can affect the efficiency and rate of the Level-1 Trigger. It is therefore important to re-derive the calibrations regularly, since the configuration of the detector and beam conditions change over the lifetime of the experiment.

When a new round of calibrations are derived, there are many steps before this one. Preceding it, Layer-1 experts calculate their scale factors for the calorimeter towers. Once performed, the jet energy corrections are then derived in cmssw. From early 2017 to mid-2018, I was responsible for procuring the JEC. The repository is accessible at <https://github.com/eshwen/L1JetEnergyCorrections>.

### 3.2.4.1 The procedure

QCD multijet MC datasets with a large  $p_T$  range are required to derive the Layer-1 and Layer-2 calibrations. Corresponding jets in data for this process can often be mismeasured (so providing good calibrations in the most difficult scenario is a good test), and MC events contain *truth-level* information from the generator. Ntuples are made from these which have the Layer-1 corrections applied. Referring to the processing chain in Chpt. 3.2.3, the jets these calibrations are derived for are post-hadronisation, but before interaction with the detector. They are referred to as *Level-1 (L1) jets*. The reference jets directly from the generator (GenJets as colloquialised in CMS) are important for matching to the L1 jets to ensure jets from the parton shower—that have no well-defined source—are not mistakenly being used.

The reference and L1 jets are matched using the variable  $\Delta R$  (see Eq. 3.6). The algorithm matches the jets by inspecting each L1 jet in descending  $p_T$  and searching for a reference jet with  $\Delta R < 0.25$ . If there is more than one match, the reference jet with the smallest  $\Delta R$  is taken. Then the next L1 jet (and so on) follows the same procedure, with the previous reference jet removed from the matching collection.

The pairs of jets are categorised into sixteen bins of  $|\eta|$ , the highest granularity available since the calibrations must run quickly on hardware. Each bin is then analysed in turn. Within each  $|\eta|$  bin, the jet pairs are subdivided into bins of the transverse momentum of the reference jet ( $p_T^{\text{ref.}}$ ). The bin widths, like  $|\eta|$ , are variable. The ratio of the transverse momentum of the L1 jet ( $p_T^{\text{L1}}$ ) to  $p_T^{\text{ref.}}$  is taken for each pair of jets. The metric for measuring the detector response is the mean of these ratios:

$$(3.9) \quad r_j = \left\langle p_T^{\text{L1}} / p_T^{\text{ref.}} \right\rangle$$

The reciprocal of  $r_j$  vs.  $p_T^{\text{L1}}$  is inspected and a correction curve is fitted. A gaussian captures

the peak at low  $p_T$  and the following equation is used for the tail:

$$(3.10) \quad p_T^{L1, \text{corr.}} = p_T^{L1} \cdot \left( p_0 + p_1 \cdot \text{erf} \left( p_2 \cdot \left( \log_{10} p_T^{L1} - p_3 \right) \right) \right. \\ \left. + p_4 \cdot \exp \left( p_5 \cdot \left( \log_{10} p_T^{L1} - p_6 \right)^2 \right) \right)$$

The starting parameters for the function may not be adequate for all cases, so they are often tuned to capture the low- $p_T$  spike and high- $p_T$  plateau (see Fig. 3.8 for an example). The value of this fit function in each  $p_T^{\text{ref.}}$  bin is exported.

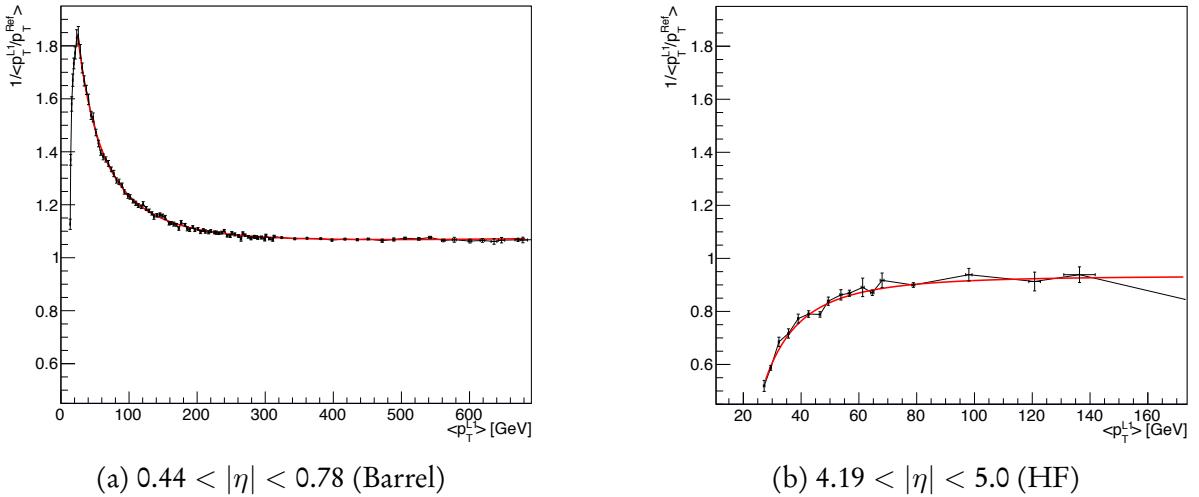


Figure 3.8: Examples of correction curves used to calibrate the jet energies in two  $|\eta|$  bins. The reciprocal of the response is plotted against the  $p_T$  of the Level-1 jet, and a complex function (Eq. 3.9) fits the points. These plots are from the jet energy corrections performed on 2018 QCD Monte Carlo.

Once all  $|\eta|$  bins have been inspected, the calibrations are consolidated in several forms. A machine-readable lookup table is included in the firmware of the Layer-2 hardware, so that the corrections are applied in the trigger. A version is added to the Level-1 Trigger packages in CMSSW so that the next steps in the calibration chain can utilise them.

A closure test is conducted to validate the corrections that were produced. The MC ntuples are regenerated with the JEC applied. Jet matching is performed and the calibrations are checked. Many diagnostic and performance plots are produced to ensure the calibrations are functioning as expected. These can be inclusive of the number of pileup interactions, or split into ranges to see if the calibrations differ between them. Examples of these are Fig. 3.9 showing scatter plots of  $p_T^{\text{ref.}}$  vs.  $p_T^{L1}$  before and after JEC are applied, and Fig. 3.10 illustrating the response.

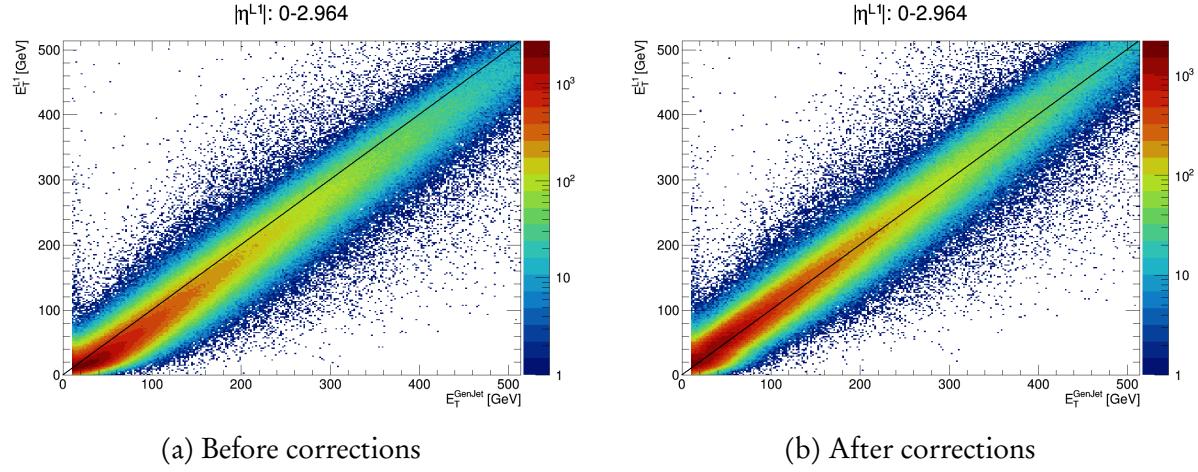


Figure 3.9: The energies of matched pairs of jets from simulation in the entire barrel and end cap, in the pileup 40–50 range, before and after jet energy corrections have been applied. An equivalent plot using jets from LHC data is expected to look similar after applying these calibrations.

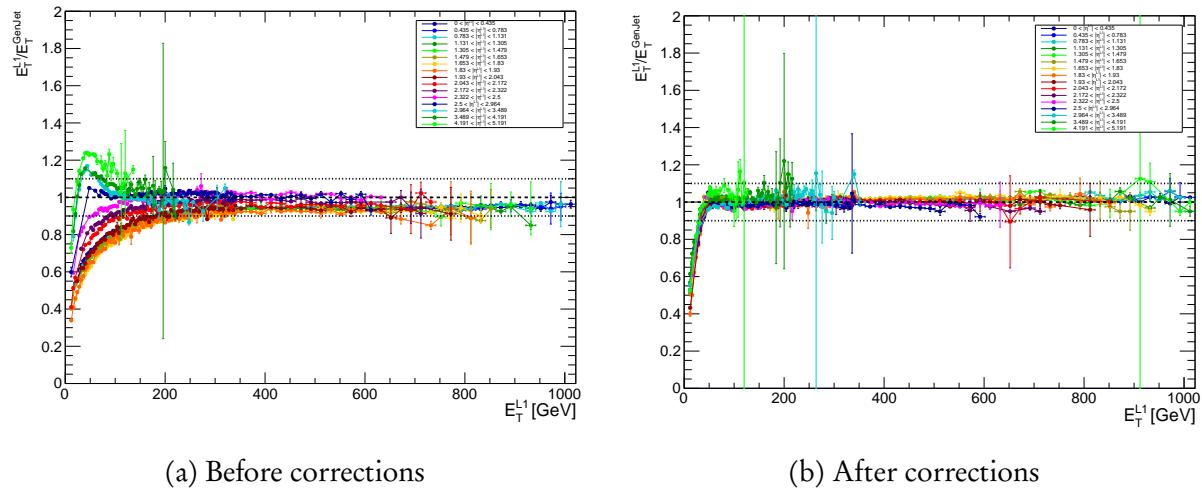


Figure 3.10: The response curves in each  $|\eta|$  bin as a function of  $p_T^{L1}$ , in the pileup 40–50 range, before and after JEC are applied. Note that in panel b, the  $x$ -axis is the corrected  $p_T^{L1}$ .

After calibrations, the energies of the generator and L1 jets are much more comparable and have a smaller spread. The jet energy corrections perform well, as intended, across wide ranges of pileup and jet  $p_T$ , providing confidence for accurate calibrations in the calorimeter when data are collected.

## DEFINITIONS OF PHYSICS OBJECTS AND OBSERVABLES

You know when grown-ups tell you everything's going to be fine, but you really think they're lying to make you feel better? Everything's going to be fine.

— The Eleventh Doctor, *Doctor Who*

This chapter first discusses some prevalent variables used in the analyses presented in this thesis. Accordingly, it is practical to consolidate their definitions here. The rest of the chapter is devoted to classification of the physics objects that inform the work in the subsequent chapters.

### 4.1 Ubiquitous observables and quantities in collider physics

#### 4.1.1 Transverse momentum ( $\vec{p}_T$ )

In the LHC (or any other hadron collider), the longitudinal momenta of the proton constituents that collide are typically unknown. However, the momentum transverse to the beam is zero before the collision, and must be zero afterward after due to momentum conservation. This is why the transverse momentum of a particle or physics object ( $\vec{p}_T$  for the vector quantity,  $p_T$  for its magnitude) is a useful variable in an analysis.

### 4.1.2 $H_T$

For analyses focused on hadronic objects—such as branches of supersymmetry where a large jet multiplicity is expected—it is practical to define the total amount of hadronic activity in an event. The scalar sum of the transverse momentum of the jets is symbolised as  $H_T$ :

$$(4.1) \quad H_T \equiv \sum_{\text{jets}} p_T$$

Typically, a lower limit on the  $p_T$  is enforced when calculating the  $H_T$ , so jets below this threshold do not factor into the sum. This is to avoid low momentum jets attributed to pileup events (see Chpt. 3.1.3) and soft radiation.

### 4.1.3 Missing transverse momentum ( $\vec{p}_T^{\text{miss}}$ )

The missing transverse momentum  $\vec{p}_T^{\text{miss}}$  is defined as the negative vector sum of the  $\vec{p}_T$  of all identified particles in an event. It is a term often interchangeable with missing transverse energy (MET,  $\vec{E}_T^{\text{miss}}$ ). Undetected particles from neutrinos or dark matter, or mismeasured kinematic properties of identified particles, will introduce an imbalance in the vector sum of the  $\vec{p}_T$ . Hence, the  $\vec{p}_T^{\text{miss}}$  will be non-zero. Formally,

$$(4.2) \quad \begin{aligned} \vec{p}_T^{\text{miss}} &\equiv - \sum_i^{N_{\text{particles}}} \vec{p}_{T,i} \\ p_T^{\text{miss}} &\equiv \left| \vec{p}_T^{\text{miss}} \right| \end{aligned}$$

The hadronic-only counterpart to this variable,  $\vec{H}_T^{\text{miss}}$ , is the negative vector sum of the jet transverse momenta in an event:

$$(4.3) \quad \begin{aligned} \vec{H}_T^{\text{miss}} &\equiv - \sum_j^{N_{\text{jets}}} \vec{p}_{T,j} \\ H_T^{\text{miss}} &\equiv \left| \vec{H}_T^{\text{miss}} \right| \end{aligned}$$

As with  $H_T$ , the  $H_T^{\text{miss}}$  is often calculated with a lower limit on the jet  $p_T$ .

### 4.1.4 Transverse mass ( $m_T$ )

The transverse mass  $m_T$  of a particle—or the collective products of a decay—is the component of its invariant mass in the transverse plane. In a hadron collider context, this observable is typically used when one of the decay products is invisible or unidentified, i.e., as  $\vec{p}_T^{\text{miss}}$ . When

searching for semi-visible jets, the  $m_T$  of the dijet system is very reliable in recovering the  $Z'$  mass. Einstein's equation  $E^2 = (pc)^2 + (m_0c^2)^2$  can be reformulated for each daughter particle in a decay as

$$(4.4) \quad E_T^2 = m_0^2 + p_T^2$$

where  $E_T$  is the transverse energy. The transverse mass of the system is then

$$(4.5) \quad \begin{aligned} m_T^2 &= (E_{T,1} + E_{T,2})^2 - (\vec{p}_{T,1} + \vec{p}_{T,2})^2 \\ &= m_1^2 + m_2^2 + 2(E_{T,1}E_{T,2} - \vec{p}_{T,1} \cdot \vec{p}_{T,2}) \end{aligned}$$

In the case of one unidentified daughter, the  $\vec{p}_T^{\text{miss}}$  may substitute a particle in the first line of Eq. 4.5. If one also treats the daughters as massless, i.e., if  $p_T \gg m_0$  like with boosted leptons, the formula approximates to

$$(4.6) \quad m_T^2 = 2p_T p_T^{\text{miss}} (1 - \cos(\Delta\phi))$$

where  $\Delta\phi = \phi - \phi(p_T^{\text{miss}})$ , and  $p_T$  and  $\phi$  are from the identified daughter.

## 4.2 Classification of analysis-level physics objects

The physics that occurs at the high energies the LHC is capable of is not simple. That makes the recording of particles and events overall a non-trivial undertaking. Based on complex hardware and software, research, and the experience of previous experiments, establishing physics objects and variables from an analyst's standpoint is thankfully straightforward. The basic definitions are already included into the ROOT files, with ancillary selections on top for the analysis-level objects. All of the descriptions that follow apply both to data and simulated samples, unless otherwise stated.

Several working points for the identification of many objects are available from the respective Physics Object Groups (POGs), trading efficiency for purity the tighter the restrictions are. *Veto*, *loose*, *medium*, and *tight* criteria exist for most objects. Particles are denoted with a subscript that signifies the working point. The requirements of which are elaborated upon in the ensuing subsections.

### 4.2.1 Jets

In the nanoAOD data tier, a jet is defined as an AK4-clustered, charged hadron-subtracted object reconstructed with the PARTICLE FLOW algorithm, with jet energy corrections applied

and a  $p_T > 15 \text{ GeV}$  prerequisite. For the analysis, the *latest* JEC and jet energy resolution (JER) treatments are applied to both data and simulation to determine the  $p_T$ . Those with  $p_T > 30 \text{ GeV}$  are kept. The JER is worse in data than in simulation, so jets in the latter are smeared to describe the former more accurately.

The tight, with or without an additional lepton veto, PF identification requirements are also mandatory to establish a sufficient degree of purity with a very high efficiency ( $> 98\%$  in all  $|\eta|$  regions). These are documented in Tabs. 4.1, 4.2, and 4.3. For all years, both the identification and background rejection efficiencies are above 98 % in the barrel and end caps.

A jet suspected to originate from a  $b$ -quark, or “ $b$ -jet,” is described by a nanoAOD AK4 jet with corrected  $p_T > 30 \text{ GeV}$  and within the acceptance of the tracker:  $|\eta| < 2.4$  in 2016, and  $< 2.5$  onward. It is also required to satisfy the medium working point (a mis-tag rate of 1 %) from the DEEPCSV algorithm [132].

#### 4.2.1.1 Jet identification criteria in 2016

The tight and tight + lepton veto ID specifications are detailed in Tab. 4.1. Specifically to this year, all jets are required to fulfill the criteria for the tight pileup ID to ensure they do not originate from pileup vertices. The algorithm that gives the pileup ID is based on a boosted decision tree trained on 2016 samples. It incorporates tracker information, jet topology, and constituent multiplicity as discriminating variables.

Criterion	$ \eta  \leq 2.4$	$2.4 <  \eta  \leq 2.7$	$2.7 <  \eta  \leq 3.0$	$ \eta  \geq 3.0$
Neutral hadron fraction	$< 0.90$	$< 0.90$	$< 0.98$	—
Neutral EM fraction	$< 0.90$	$< 0.90$	$> 0.01$	$< 0.90$
# constituents	$> 1$	$> 1$	—	—
Muon energy fraction	$< 0.80^*$	$< 0.80^*$	—	—
Charged hadron fraction	$> 0$	—	—	—
Charged multiplicity	$> 0$	—	—	—
Charged EM fraction	$< 0.99 \text{ (} 0.90^* \text{)}$	—	—	—
# neutral particles	—	—	$> 2$	$> 10$

Table 4.1: The requirements for a jet to pass tight identification for data taken in 2016, and to Monte Carlo events emulating that year. Information taken from Ref. 56.

\* Only applies to the tight + lepton veto ID.

### 4.2.1.2 Jet identification criteria in 2017

The tight and tight + lepton veto ID conditions are detailed in Tab. 4.2. Jets with  $p_T < 50 \text{ GeV}$  are required to meet the tight pileup ID criteria. Any jets with a raw (uncorrected) transverse momentum  $p_T^{\text{raw}} < 50 \text{ GeV}$  within the region  $2.65 < |\eta| < 3.139$  are vetoed due to noise in the ECAL end cap.

Criterion	$ \eta  \leq 2.4$	$2.4 <  \eta  \leq 2.7$	$2.7 <  \eta  \leq 3.0$	$ \eta  \geq 3.0$
Neutral hadron fraction	$< 0.90$	$< 0.90$	—	$> 0.02$
Neutral EM fraction	$< 0.90$	$< 0.90$	$> 0.02, < 0.99$	$< 0.90$
# constituents	$> 1$	$> 1$	—	—
Muon energy fraction	$< 0.80^*$	$< 0.80^*$	—	—
Charged hadron fraction	$> 0$	—	—	—
Charged multiplicity	$> 0$	—	—	—
Charged EM fraction	$< 0.80^*$	—	—	—
# neutral particles	—	—	$> 2$	$> 10$

Table 4.2: The requirements for a jet to pass tight identification for data taken in 2017, and to Monte Carlo events emulating that year. Information taken from Ref. 57.

\* Only applies to the tight + lepton veto ID.

### 4.2.1.3 Jet identification criteria in 2018

The tight and tight + lepton veto ID specifications are detailed in Tab. 4.3. As in 2017, jets with  $p_T < 50 \text{ GeV}$  must fulfill the tight pileup ID conditions.

Criterion	$ \eta  \leq 2.6$	$2.6 <  \eta  \leq 2.7$	$2.7 <  \eta  \leq 3.0$	$3.0 <  \eta  \leq 5.0$
Neutral hadron fraction	$< 0.90$	$< 0.90$	—	$> 0.20$
Neutral EM fraction	$< 0.90$	$< 0.99$	$> 0.02, < 0.99$	$< 0.90$
# constituents	$> 1$	—	—	—
Muon energy fraction	$< 0.80^*$	$< 0.80^*$	—	—
Charged hadron fraction	$> 0$	—	—	—
Charged multiplicity	$> 0$	$> 0$	—	—
Charged EM fraction	$< 0.80^*$	$< 0.80^*$	—	—
# neutral particles	—	—	$> 2$	$> 10$

Table 4.3: The requirements for a jet to pass tight identification for data taken in 2018, and to Monte Carlo events emulating that year. Information taken from Ref. 58.

\* Only applies to the tight + lepton veto ID.

### 4.2.2 Boosted decays of top quarks and electroweak bosons

Many traditional methods have been developed to tag jets originating from heavy particles such as top quarks, vector bosons, and Higgs bosons. Though, there are limitations with all of these, and in the era of machine learning a plethora of new tools have arisen. One such algorithm—DEEPAK8—has been developed by physicists in CMS to classify jets that originate from numerous heavy particle decays. In the analysis in search of invisibly decaying Higgs bosons, specific interest is paid to those coming from top quarks and vector bosons in order to assist the categorisation of the  $t\bar{t}H$  and  $VH$  signals. The design and optimisation of the algorithm is described in Ref. 60.

The algorithm operates on AK8 jets. Following the same reconstruction and identification prescription as AK4 jets, they differ only in the cone size and  $p_T$  requirement. Hadrons are clustered with the anti- $k_T$  algorithm using a radius parameter of 0.8 instead of the usual 0.4, and must possess a  $p_T > 170 \text{ GeV}$ . The corrections and uncertainties from JER and JES are also propagated identically.

The ensuing selections are applied to AK8 jets to classify them, as recommended by the DEEPAK8 developers:

- Boosted top:
  - $p_T > 400 \text{ GeV}$
  - $105 < m_{\text{SD}} < 210 \text{ GeV}$
  - $t$  vs. QCD discriminator score with the tight working point (1 % mis-tag rate)
- Boosted  $V$ :
  - $p_T > 200 \text{ GeV}$
  - $65 < m_{\text{SD}} < 105 \text{ GeV}$
  - $W$  vs. QCD discriminator score with the tight working point (1 % mis-tag rate)

Orthogonality is ensured between the two classifications from the soft drop mass  $m_{\text{SD}}$  [110] window. The quantity is corrected for pileup contributions using the PUPPI algorithm [34]. Tight working points that effectively separate the jets from QCD background ensure a high purity collection. Boosted  $V$  jets are not further separated into  $W$ - and  $Z$ -tagged since the mass degeneracy makes this inherently difficult. The  $W$  vs. QCD discriminator is therefore used to tag both.

### 4.2.3 Muons

The reconstruction and identification criteria for muons is described in significant detail in Ref. 133. From there, a loose muon  $\mu_{\text{loose}}$  satisfies the loose muon identification requirements. Additionally,  $|\eta| < 2.4$ ,  $p_T > 10 \text{ GeV}$ , and the loose working point of the PARTICLE FLOW relative isolation parameter (Eq. 5.4 in Ref. 128)  $I_{\text{PF}} < 0.25$  must be satisfied. A tight muon  $\mu_{\text{tight}}$  meets the tight muon ID requirements from Ref. 133,  $|\eta| < 2.4$ ,  $p_T > 20 \text{ GeV}$ , and the tight working point for the relative isolation ( $I_{\text{PF}} < 0.15$ ).

The relative isolation parameter is the ratio of the sum of track, HCAL and ECAL energies within  $\Delta R < 0.4$  of the muon to the particle's momentum. The loose and tight working points are designed to give efficiencies of 98 % and 95 %, respectively.

### 4.2.4 Electrons

Electrons are labelled as veto ( $e_{\text{veto}}$ ) or tight electrons ( $e_{\text{tight}}$ ), depending on which POG ID requirements are fulfilled. For 2016–18, version 2 of the cut-based identification scheme (tuned using 2017 samples), is applied. These are distinguished for the barrel region in Tab. 4.4, and end cap region in Tab. 4.5. The veto and tight ID specifications are designed for average efficiencies of 95 % and 70 %, respectively.

Criterion	Veto	Tight
Full $5 \times 5 \sigma_{i\eta i\eta}$	$< 0.0126$	$< 0.0104$
$ \Delta\eta_{\text{in}}^{\text{seed}} $	$< 0.00463$	$< 0.00255$
$ \Delta\phi_{\text{in}} $	$< 0.148$	$< 0.022$
$H/E$	$< 0.05 + \frac{1.16}{E_{\text{SC}}} + \frac{0.0324\rho}{E_{\text{SC}}}$	$< 0.026 + \frac{1.15}{E_{\text{SC}}} + \frac{0.0324\rho}{E_{\text{SC}}}$
Relative isolation with $A_{\text{eff.}}$	$< 0.198 + \frac{0.506}{p_T}$	$< 0.0287 + \frac{0.506}{p_T}$
$ E_{\text{ECAL}}^{-1} - p_{\text{Tracker}}^{-1} $	$< 0.209$	$< 0.159$
Expected missing inner hits	$\leq 2$	$\leq 1$
Pass conversion veto	yes	yes

Table 4.4: Criteria for defining the veto and tight electron IDs in the barrel region (supercluster  $|\eta| \leq 1.479$ ). Information taken from Ref. 54.

The variable  $\sigma_{i\eta i\eta}$  is the energy-weighted standard deviation of the ECAL crystal  $\eta$ , centred on the local energy maximum, in this case using the full  $5 \times 5$  crystal information from the calorimeter tower.  $\Delta\eta_{\text{in}}^{\text{seed}}$  and  $\Delta\phi_{\text{in}}$  are the differences in pseudorapidity and azimuthal angle, respectively, between the supercluster (seed in the case of  $\Delta\eta$ ) and the track.  $H/E$  is the ratio of the candidate's central energy deposit in the HCAL to the ECAL. The relative

Criterion	Veto	Tight
Full $5 \times 5 \sigma_{i\eta i\eta}$	$< 0.0457$	$< 0.0353$
$ \Delta\eta_{\text{in}}^{\text{seed}} $	$< 0.00814$	$< 0.00501$
$ \Delta\phi_{\text{in}} $	$< 0.19$	$< 0.0236$
$H/E$	$< 0.05 + \frac{2.54}{E_{\text{SC}}} + \frac{0.183\rho}{E_{\text{SC}}}$	$< 0.0188 + \frac{2.06}{E_{\text{SC}}} + \frac{0.183\rho}{E_{\text{SC}}}$
Relative isolation with $A_{\text{eff.}}$	$< 0.203 + \frac{0.963}{p_{\text{T}}}$	$< 0.0445 + \frac{0.963}{p_{\text{T}}}$
$ E_{\text{ECAL}}^{-1} - p_{\text{Tracker}}^{-1} $	$< 0.132$	$< 0.0197$
Expected missing inner hits	$\leq 3$	$\leq 1$
Pass conversion veto	yes	yes

Table 4.5: Criteria for defining the veto and tight electron IDs in the end cap region (supercluster  $|\eta| > 1.479$ ). Information taken from Ref. 54.

isolation parameter is computed in the same way as muons (Eq. 5.4 in Ref. 128), but uses a cone of  $\Delta R < 0.3$  instead of 0.4. The isolation threshold takes the effective area  $A_{\text{eff.}}$  into account. Along with  $\rho$ —a parameter in the FASTJET package [40] used for jet finding—these are used for estimating the contamination from pileup in an event.  $E_{\text{SC}}$  is the energy of the supercluster, and is used to scale the isolation requirement. From material interactions within the tracker, photons may convert into  $e^+ e^-$  pairs. A dedicated veto aims to mitigate this effect.

Common to all electrons in the analysis, a cut is placed of  $|\eta| < 2.5$  so they are reconstructed with tracker information. To further utilise the subdetector, impact parameter conditions on the transverse and longitudinal directions ensure they originate sufficiently close to the primary vertex:  $d_0 < 0.05$  cm and  $dz < 0.1$  cm in the barrel, with  $d_0 < 0.1$  cm and  $dz < 0.2$  cm in the end cap. The  $e_{\text{veto}}$  and  $e_{\text{tight}}$  sets are only separated by the identification prerequisites above, and transverse momentum. A veto electron requires  $p_{\text{T}} > 10$  GeV, and a tight electron  $p_{\text{T}} > 40$  GeV.

#### 4.2.5 Photons

Photons are identified in a similar manner to electrons—using version 2 of the cut-based scheme for 2016–18, optimised with 2017 Monte Carlo. Loose  $\gamma_{\text{loose}}$  and medium photons  $\gamma_{\text{medium}}$  are defined which fulfill the corresponding working point criteria for identification. For inclusivity, the latter is used instead of the tight working point. These are detailed in Tab. 4.6 and Tab. 4.7 for the barrel and end cap regions, respectively. The loose photon ID specification is designed to be 90 % efficient, and the medium ID 80 %. In the barrel and end caps, the loose ID rejects 86 % and 77 % of background, respectively, while the medium ID

removes 89 % and 82 %, respectively. Both collections must also pass the electron conversion veto to ensure they originate promptly from the primary vertex.

Criterion	Loose	Medium
$H/E$	$< 0.04596$	$< 0.02197$
$\sigma_{inj\eta}$	$< 0.0106$	$< 0.01015$
$\rho$ -corrected PF charged hadron isolation	$< 1.694$	$< 1.141$
$\rho$ -corrected PF neutral hadron isolation	$< 24.032 + X + Y$	$< 1.189 + X + Y$
$\rho$ -corrected PF photon isolation	$< 2.876 + Z$	$< 2.08 + Z$

Table 4.6: Criteria for defining the loose and medium photon IDs in the barrel region of the detector ( $|\eta| \leq 1.479$ ). The factors  $X = 0.01512 p_T$ ,  $Y = 2.259 \times 10^{-5} p_T^2$ , and  $Z = 0.004017 p_T$  in the isolation criteria are used to scale the threshold. Information taken from Ref. 55.

Criterion	Loose	Medium
$H/E$	$< 0.0590$	$< 0.0326$
$\sigma_{inj\eta}$	$< 0.0272$	$< 0.0272$
$\rho$ -corrected PF charged hadron isolation	$< 2.089$	$< 1.051$
$\rho$ -corrected PF neutral hadron isolation	$< 1.922 + X + Y$	$< 2.718 + X + Y$
$\rho$ -corrected PF photon isolation	$< 4.162 + Z$	$< 3.867 + Z$

Table 4.7: Criteria for defining the loose and medium photon IDs in the end cap region of the detector ( $|\eta| > 1.479$ ). The factors  $X = 0.0117 p_T$ ,  $Y = 2.3 \times 10^{-5} p_T^2$ , and  $Z = 0.0037 p_T$  in the isolation criteria are used to scale the threshold. Information taken from Ref. 55.

The variables  $\sigma_{inj\eta}$  and  $\rho$  follow the same definitions as in Chpt. 4.2.4. The  $\rho$ -corrected isolation is calculated by taking the maximum of zero, and the difference between the PARTICLE FLOW isolation and possible contamination from pileup:

$$(4.7) \quad I_{\text{PF}}^{\text{corr.}} = \max(I_{\text{PF}} - \rho A_{\text{eff.}}, 0)$$

As well as the ID, the two collections diverge further due to  $\eta$  and  $p_T$  requirements. A loose photon is characterised by  $|\eta| < 2.5$  and  $p_T > 15 \text{ GeV}$ , while a medium photon must be central ( $|\eta| < 1.442$ ) and possess  $p_T > 230 \text{ GeV}$ .

### 4.2.6 Tau leptons

Tau leptons  $\tau$  can decay hadronically and present a non-negligible background in searches for jets and  $p_T^{\text{miss}}$ . The only purpose they serve in the  $H \rightarrow$  invisible analysis is to veto events

that contain them. These are denoted as *very loose* taus  $\tau_{v.\,loose}$ . Version 2 of the multivariate identification algorithm is applied in all years, and tau leptons must pass the very loose criteria. Dedicated multivariate discriminators against electrons and muons are provided and employed, passing the loose and very loose working points, respectively. Kinematically, a  $\tau_{v.\,loose}$  also requires  $p_T > 20\,\text{GeV}$ ,  $|\eta| < 2.3$ , and  $dz < 0.2\,\text{cm}$ .

#### 4.2.7 Overlap removal

Particle detection is not a straightforward process. As such, there may be particles identified close together in an event either due to real physics or misreconstruction. To ensure only well-measured and correctly-identified objects are used in analysis, particles considered to be overlapping are removed—or *cross cleaned*—between the different classes. This is performed in the following order.

Muons are first identified, followed by electrons. Cross cleaning is performed against the  $\mu_{loose}$  collection of objects, so electrons are not counted in the event if they are within  $\Delta R < 0.3$  of one. Photons are next in the list. Both the  $\gamma_{loose}$  and  $\gamma_{medium}$  collections are cross-cleaned against  $\mu_{loose}$  and  $e_{veto}$  objects in the event with  $\Delta R < 0.3$ . The same collections and separation are used to remove overlap with the set  $\tau_{v.\,loose}$  leptons. Finally, jets are identified, with those in the AK4 collection requiring a separation of  $\Delta R > 0.4$  from any electron, muon, and photon defined previously.

#### 4.2.8 Revised energy sums

Most CMS analyses, including those presented in this thesis, define the missing transverse momentum with the *type-I correction*: corrections to the JER and JES are applied as in Chpt. 4.2.1, and so the corrected jet  $\vec{p}_T$  are propagated to the calculation of  $\vec{p}_T^{\text{miss}}$ . Uncertainties are calculated in the same manner as jets.

In 2017, significant noise in the ECAL end cap affected jets and  $\vec{p}_T^{\text{miss}}$ , leading to potentially large energy mismeasurements. These quantities were recomputed after reconstruction, and carried through the analysis.

The scalar sum  $H_T$  and negative vector sum of hadronic transverse momentum  $\vec{H}_T^{\text{miss}}$  are broadly defined in Chpts. 4.1.2 and 4.1.3, respectively. They are calculated from the analysis-level AK4 jets with  $p_T > 30\,\text{GeV}$ , as detailed in Chpt. 4.2.1.

Regarding the  $\vec{p}_T^{\text{miss}}$  and  $\vec{H}_T^{\text{miss}}$ , in most cases the magnitudes ( $p_T^{\text{miss}}$  and  $H_T^{\text{miss}}$ , respectively) are the important quantities used in the event selection and other aspects of the analysis.

## SIMULATION STUDIES OF SEMI-VISIBLE JETS

Of darkness visible so much be lent, as half to show,  
half veil, the deep intent.

— Alexander Pope

**A**ccess to the dark side of the universe is under a complex lock, seemingly unbreakable judging by the efforts of the past. A portal bridging the visible world to a realm of dark QCD, however, may be discoverable at the LHC. This chapter is dedicated to simulation studies in the context of a novel search for semi-visible jets, characterised by hadronic final states and moderate missing transverse momentum.

### 5.1 Analysis summary

The analysis in search for semi-visible jets is formed of a collaboration amongst the University of Bristol, University of Maryland, University of Rochester, Universität Zürich, and the Fermi National Accelerator Laboratory (FNAL). The  $s$ -channel production mode with the full Run-2 dataset of CMS is the subject of the current analysis, at the time of writing. A dijet final state with moderate  $p_T^{\text{miss}}$  is expected. The transverse mass of the dijet system  $m_T$ —calculated with Eq. 4.5—is used as the search variable. Confirmed by the distributions in this chapter for simulated signal, a peak close to  $m_{Z'}$  can be captured. With the inclusion of the  $p_T^{\text{miss}}$ , the resolution of  $m_T$  is better than simply the dijet mass ( $m_{jj}$ ) which would otherwise be used to recover the peak. The authors of Ref. 70 also demonstrate that sensitivity

is stronger with  $m_T$  over  $m_{jj}$ . Novel techniques are utilised, such as the development of a boosted decision tree to tag semi-visible jets and distinguish them from mismeasured QCD jets. A complete description of the  $s$ -channel analysis as of September 2019 can be found in Ref. 119.

My contribution is predominantly the development and understanding of signal simulation in `MADGRAPH5_aMC@NLO` [15] for both the  $s$ -channel and  $t$ -channel aspects of the model. As such, the remainder of the chapter is focused on their discussion. The code developed by myself to perform the simulation is being used for the future analysis of the  $t$ -channel mode as well as a low-mass boosted  $Z'$  search.

## 5.2 Signal simulation overview

In the main analysis, `PYTHIA8` [141] is employed to generate the  $s$ -channel signal as it can parametrise all the relevant aspects of the model in a simple manner with high signal efficiency. `MADGRAPH` is the often-preferred generator as it can handle more complex models and decays. A higher degree of customisation and tuning of parameters is also possible. For the  $s$ -channel mode, modelling of the hard process should be equivalent between `PYTHIA` and `MADGRAPH` at leading order (LO). In the former the properties, interactions, and decays of the dark sector particles are implemented via the Hidden Valley module, available from `PYTHIA` 8.226. Samples generated with `MADGRAPH` are hadronised by `PYTHIA` as part of the full simulation chain within `CMSSW`. Jet matching and filter efficiencies may noticeably reduce the final number of events. The  $t$ -channel model is possible to parametrise in `PYTHIA`, but due to its complexity `MADGRAPH` is favoured. Hadronisation, however, is still performed by `PYTHIA`. In all cases, signal samples are generated at LO.

All the details for  $s$ -channel signal generation with `PYTHIA8` is given in Chpts. 5.3 and 5.4. This avenue was developed primarily by collaborators at FNAL. An equivalent implementation from myself where the hard scatter is modelled in `MADGRAPH5_aMC@NLO`, as both an alternative and cross-check to the `PYTHIA` version, is chronicled in Chpt. 5.5. A description of the  $t$ -channel process is also present in the section. The parameters defining the shower are synchronised between myself and the FNAL colleagues. Comparisons between the `PYTHIA` and `MADGRAPH` implementations are discussed in Chpt. 5.6.

For all spectra, no additional object corrections or selections are performed and are direct from the simulation pipeline. Reconstructed jets satisfy the tight identification criteria given in Tabs. 4.1, 4.2, or 4.2 depending on the data taking year. The missing transverse momentum  $p_T^{\text{miss}}$  contains the type-I correction as outlined in Chpt. 4.2.8. Similarly, the rest of the objects

in the ntuples contain at least minimal quality assurances prior to any analysis-level selection.

The distributions that pepper the remainder of the chapter are generated with 2016 detector conditions and reconstruction. As the analysis began in 2017, the 2016 dataset was the only complete one intending to be used. Since it was the most mature and had been studied in depth by the Collaboration, the simulation should be highly representative of what could be observed in the data from that period. Selecting one year also ensures consistency across all of the plots even though they should not significantly change year-to-year.

## 5.3 Generation in PYTHIA

The Hidden Valley module allows for simulating  $q\bar{q} \rightarrow Z' \rightarrow \chi\bar{\chi}$ , where the  $Z'$  acts as intended—a vector portal between the visible and dark sectors. Since a small percentage of them are expected to decay visibly, a branching ratio to each of the six SM quarks is set to 0.003. The remaining fraction of 0.982 decay to  $\chi\bar{\chi}$ , where  $\chi$  is a Hidden Valley particle charged only under that gauge group. The masses of the  $Z'$  and  $\chi$ , and the narrow width of the  $Z'$  for resonant production can be given. Showering in the dark and visible sectors is then performed as in Chpt. 5.4. For both the complete event generation, and for hadronisation in the case of MADGRAPH generation, PYTHIA 8.230 is used.

## 5.4 Showering in PYTHIA

Once the hard process has been simulated, showering of the dark and visible particles is performed. This can be separated into three separate steps. The *parton shower* itself is where undecayed free quarks and gluons fragment and radiate due to their high energies. At *hadronisation*, the showered partons coalesce into composite hadrons due to colour confinement. Finally, the *underlying event* is the simulation of the previous two steps for any additional, softer collisions that occur between constituents of the incident particles, separate from the hard interaction. Often, the sum of all of these steps is just referred to as the *parton shower* or *hadronisation*.

### 5.4.1 Hadronisation of dark quarks

Dark quarks hadronise into one of two species of dark meson (as the number of dark flavours  $N_f = 2$ ) that correspond to particles from the Hidden Valley module:  $\pi_{\text{dark}}$  and  $\rho_{\text{dark}}$ , that are pseudoscalar and vector, respectively. Each possess flavour-diagonal and off-diagonal

variants. They are generated probabilistically with a ratio of 1:3. Dark hadrons are set to decay invisibly with a branching fraction  $r_{\text{inv}}$ . Remaining decay modes are to SM quarks via a virtual  $Z'$  since it is the leptophobic portal between the dark and visible sectors. Decays of  $\rho_{\text{dark}}$  are democratic, i.e., with equal probability to accessible SM quark-antiquark pairs.<sup>1</sup> Specifically,  $\rho_{\text{dark}}$  particles with  $m_{\text{dark}} > 2m_b$  have a 1/5 probability to decay to a  $u, d, c, s$ , or  $b$ . For  $2m_c < m_{\text{dark}} < 2m_b$ ,  $b$  quarks are excluded and the branching ratio increases to 1/4 for the remaining species. The  $c$  quark is removed—and branching ratio modified—in the same manner for  $m_{\text{dark}} < 2m_c$ .

The decays of the  $\pi_{\text{dark}}$  mesons, however, are through a mass insertion. They couple to the longitudinal component of the  $Z'$ , assumed to arise from a leptophobic Higgs sector in an analogous manner to the electroweak bosons. Running quark masses are accounted for, as calculated in Ref. 80, since the decays are produced at the Higgs mass scale. Branching fractions to the SM quarks are therefore based on the squares of the *running* masses over the *pole* masses. A graphical representation of the mass insertion decay is provided in Fig. 5.1.

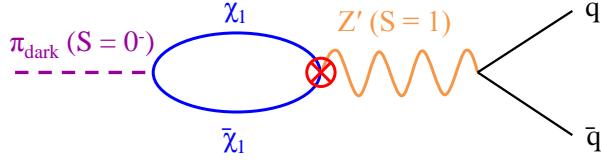


Figure 5.1: A diagram of the mass insertion decay of  $\pi_{\text{dark}}$  mesons in the  $s$ -channel semi-visible jet model.

The dark confinement scale  $\Lambda_{\text{dark}}$  and running dark coupling  $\alpha_{\text{dark}}$  are encoded based on the dark hadron mass as given below. PYTHIA is now aware of the energy scales by which to hadronise and decay the dark sector particles. Final state dark radiation is also permitted, with a minimum  $p_T$  of  $1.1\Lambda_{\text{dark}}$  as imposed by PYTHIA.

#### 5.4.1.1 Simpler parametrisation of $\alpha_{\text{dark}}$

For simplicity, the values of  $\alpha_{\text{dark}}$  used in the scan of parameter points are based on the value of  $\Lambda_{\text{dark}}$  that maximises the dark hadron multiplicity in an event. The dependence is very small on  $m_{Z'}$ , and instead modelled as a function of  $m_{\text{dark}}$  by performing a fit. The value of  $\Lambda_{\text{dark}}$  that leads to the largest number of dark hadrons ( $\Lambda_{\text{dark}}^{\text{peak}}$ ) is

$$(5.1) \quad \Lambda_{\text{dark}}^{\text{peak}} = 3.2m_{\text{dark}}^{0.8}$$

<sup>1</sup>Top quarks are excluded in all cases, since in the scan of model parameters where the greatest sensitivity is predicted, the dark mesons are always too light to decay on shell to a  $t\bar{t}$  pair.

which gives the parameter  $\alpha_{\text{dark}}^{\text{peak}} = \alpha_{\text{dark}}(\Lambda_{\text{dark}}^{\text{peak}})$  calculated according to Eq. 2.12. Variations on  $\alpha_{\text{dark}}$  are calculated by scaling it by a factor 0.5 and 1.5, yielding  $\alpha_{\text{dark}}^{\text{low}} = 0.5\alpha_{\text{dark}}^{\text{peak}}$  and  $\alpha_{\text{dark}}^{\text{high}} = 1.5\alpha_{\text{dark}}^{\text{peak}}$ . For the purposes of emphasizing the effect of the choice of  $\alpha_{\text{dark}}$  on the kinematics of the signal, an additional variation was introduced:  $\alpha_{\text{dark}}^{\text{v. high}} = 2\alpha_{\text{dark}}^{\text{peak}}$ . The advantage of denoting the dark coupling strength with “peak,” “low,” and “high,” enables a more intuitive understanding of the parameter over a numerical value. Eq. 5.1 also means that  $\Lambda_{\text{dark}}$  (and therefore  $\alpha_{\text{dark}}$ ) can be easily varied as a function of  $m_{\text{dark}}$  for each configuration of parameters in a scan of the model. Fig. 5.2 illustrates the evolution of  $\Lambda_{\text{dark}}$  as a function of  $m_{\text{dark}}$  for each value of  $\alpha_{\text{dark}}$  that is considered.

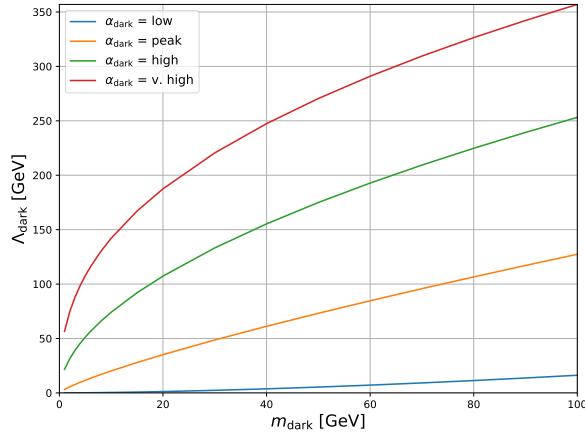


Figure 5.2: The dependence of the dark force scale  $\Lambda_{\text{dark}}$  on the dark hadron mass  $m_{\text{dark}}$  for each value of  $\alpha_{\text{dark}}$  used in the thesis.

### 5.4.2 Jet clustering

The softer, visible radiation is simulated, and the clustering of SM hadrons into jets happens in the hadroniser. The algorithm logic is known as “sequential recombination,” where particles are successively combined until certain criteria are met. In PYTHIA, clustering stops above the merging scale known as the  $q_{\text{Cut}}$ , which is set to 125 GeV. For events generated in MADGRAPH and interfaced with PYTHIA for the parton shower, it is recommended that  $q_{\text{Cut}} > 1.1x_{\text{qcut}}$ . The  $x_{\text{qcut}}$  denotes the merging scale in MADGRAPH. Referring to the transverse momentum, the symbol  $k_{\text{T}}$  is traditionally used in naming algorithms such as the anti- $k_{\text{T}}$  algorithm. For consistency within this thesis, however,  $p_{\text{T}}$  will be used.

For given particles  $i$  and  $j$ , distance between them  $d_{ij}$ , and to the beam  $d_{iB}$ , are calculated

as in Ref. 124:

$$(5.2) \quad d_{ij} = \min(p_{Ti}^{2k}, p_{Tj}^{2k}) \frac{\Delta R_{ij}}{R^2}, \\ d_{iB} = p_{Ti}^{2k}$$

where, in the rapidity-azimuthal ( $y\text{-}\phi$ ) plane,  $R$  is the radius of the cone—defined as 1.0 to match the merging parameters used in MADGRAPH.  $\Delta R_{ij}$  is the separation between  $i$  and  $j$ ,<sup>2</sup> and  $k$  defines the algorithm choice:  $k = -1$  in the anti- $k_T$  algorithm,  $k = 0$  in the Cambridge-Aachen algorithm, and  $k = 1$  in the  $k_T$  algorithm. The latter is chosen for the same reason as the cone radius size.

If  $d_{ij} < d_{iB}$ , the particles  $i$  and  $j$  are combined, replacing the individual constituents in the list of inputs. Otherwise, particle  $i$  is designated as a jet and removed from list of inputs. For all combinations of particles remaining in the input list, the distances are recalculated until all objects possess a  $p_T$  above the merging scale. Once the algorithm has finished, all that remains are fully-clustered jets. Matching is performed between the clustered jets and original partons to avoid double counting. Events with insufficiently-matched jets are rejected. A minimum number of jets may also be specified and events with fewer than this are rejected. A larger  $r_{\text{inv}}$  tends to reduce the merging and matching efficiencies since more energy is locked in the dark sector and fewer jets are clustered above the merging scale.

### 5.4.3 Filtering events

Two filters are implemented in PYTHIA that reject events with unrealistic decays: a  $Z_2$  symmetry in the model requires invisibly decaying dark hadrons to produce the dark matter particles in pairs, and an invisibly decaying  $Z'$  must do so into a  $\chi\bar{\chi}$  pair. Coupled with the efficiency of the jet matching and clustering algorithms—which only significantly affects events generated externally and decayed with PYTHIA—there are multiple sources of inefficiency in the generation. Tuning the merging scale and matching parameters based on  $r_{\text{inv}}$  may yield in more effective generation. Aggregated over all sources, efficiency on the order of 15–20 % are expected for most MADGRAPH samples.

## 5.5 Generation in MADGRAPH

The MADGRAPH5\_AMC@NLO 2.6.0 event generator is able to simulate the hard scatter for both the  $s$ - and  $t$ -channel UV completions of the model, i.e., the decay of the  $Z'$  into  $\chi\bar{\chi}$  in the

---

<sup>2</sup>This is calculated with Eq. 3.6 where  $\eta$  is replaced with  $y$  from Eq. 3.5.

$s$ -channel mode, and the exchange of the  $\Phi$  to transform  $q\bar{q}$  into  $\chi\bar{\chi}$  in the  $t$ -channel mode. Up to two additional standard model quarks or gluons may accompany the final state for either mode. The particles, couplings, and other parameters required to describe the models are defined using the FEYNRULES package [13]. Of the four principal free parameters from Chpt. 2.4.1, the mediator mass and  $m_\chi$  are defined at this stage.

The  $s$ -channel process is a modified version of the spin-1 “DMsimp” class of simplified dark matter models [25]. The  $t$ -channel description is completely custom. Both were initially acquired by the analysis team from one of the authors of Ref. 70 at <https://github.com/smsharma/SemivisibleJets>. At the time, it contained only one set of mass points. For the work presented here, this has been greatly expanded allowing any combination of  $Z'$ ,  $\Phi$ , and  $\chi$  masses and couplings. All signal samples produced with MADGRAPH that feature in this thesis are done so with <https://github.com/eshwen/SemivisibleJets>.

In addition to the hard process, the full cmssw simulation pipeline can also be executed. This enables hadronisation and decays of both the SM and dark quarks in PYTHIA (as detailed in Chpt. 5.4). The remaining free parameters of the model— $\alpha_{\text{dark}}$  and  $r_{\text{inv}}$ —are specified there. Subsequently, detector conditions and material interactions can be modelled in the 2016, 2017, and 2018 data taking periods of CMS. The parton distribution functions used in centrally-produced CMS simulation for each data taking period are also applied to these semi-visible jet samples for consistency.

Common to both the  $s$ - and  $t$ -channel signal models generated for these studies, the couplings between the mediator and SM quarks  $g_q$ —also the mediator and dark quarks  $g_\chi$ —are set to 1.0. Coupling strengths are furthermore independent of the dark and SM quark flavours. The number of dark flavours  $N_f$  and dark colours  $N_c$  each have values of 2. Both affect the dark force scale as per Eq. 2.12. Individually, former parameter allows for the production of two types of dark hadron, described in Chpt. 5.4, while the latter sets the dark colour composition of the dark gluon  $g_{\text{dark}}$ .

### 5.5.1 MADGRAPH run settings

Somewhat decoupled from the parameters of the signal models themselves, there are many important, adjustable settings controlled by the run card that affect the behaviour of MADGRAPH. The most notable of which is the parton/jet merging scale, denoted as the `xqcut`. MADGRAPH has difficulty simulating soft, collinear physics. To avoid this, a threshold may be placed on the `xqcut` to only generate sufficiently energetic events. Between two partons  $i$

and  $j$ , the  $k_T$  is given as

$$(5.3) \quad k_T^2 = 2 \min(p_{Ti}, p_{Tj}) [\cosh(\eta_i - \eta_j) - \cos(\phi_i - \phi_j)]$$

where  $\eta$  and  $\phi$  are the pseudorapidity and azimuthal angle, respectively. Complete definitions are presented in Chpt. 3.2.1.1. Events with  $k_T < \text{xqcut}$  are not simulated. In essence, the  $\text{xqcut}$  is a measure of the required separation between partons in the event. A value of 100 GeV is set in the analysis for consistency with Ref. 70. This interpretation of a merging scale is similar, but not quite the same, as PYTHIA's discussed in Chpt. 5.4.

The choice of parton distribution function also lives in the run card. As stated above, this changes to reflect the data taking period that is emulated. Preferences for particle separation, momentum and direction cuts, beam energy, and the QCD renormalisation and factorisation scales are some of the additional settings that can be customised.

### 5.5.2 $s$ -channel

In the description of the  $s$ -channel model, the leptophobic  $Z'$  is a spin-1 boson with vector couplings to the dark and SM quarks. A width  $\Gamma_{Z'} = 10$  GeV is specified; small enough to allow its resonant production. The dark quarks themselves are Dirac spinors with dark colour quantum numbers. Real and complex scalar dark quarks also exist in the files detailing the model, but are decoupled and not generated in this implementation. Consideration of these supplementary species of particle are beyond the scope of this thesis.

Cross sections for this process are calculated at NLO as a function of  $m_{Z'}$ , assuming resonant production and a universal quark coupling. Events in simulation are weighted by the cross section multiplied by integrated luminosity of the full Run-2 dataset for an estimation of the production rate over this era of the LHC.

It is important to understand the influence each free parameter of the model exerts on the kinematics of the final state. In which observables the variations of a given parameter manifest can aid in disentangling potential signal from background. The sensitivity, or lack thereof, of a parameter on the search variable also creates a sensible range for which to constrain the set of signal samples that must be generated. Signal models are given a label as  $\text{SVJ}_{-}\langle m_{\text{mediator}} \rangle_{-}\langle m_{\text{dark}} \rangle_{-}\langle r_{\text{inv}} \rangle_{-}\langle \alpha_{\text{dark}} \rangle$ . A benchmark point was chosen in the analysis with  $m_{Z'} = 3000$  GeV,  $m_{\text{dark}} = 20$  GeV,  $r_{\text{inv}} = 0.3$ , and  $\alpha_{\text{dark}} = \text{peak}$ , labelled as  $\text{SVJ\_3000\_20\_0.3\_peak}$ . The choice is motivated by Ref. 69 and considerations of the other attributes of the signal provided in this chapter. Fig. 5.3 shows the  $m_T$  in the case where a single free parameter is varied with respect to the benchmark point. A glimpse into the sensitivity of the parameter on the kinematics is evident.

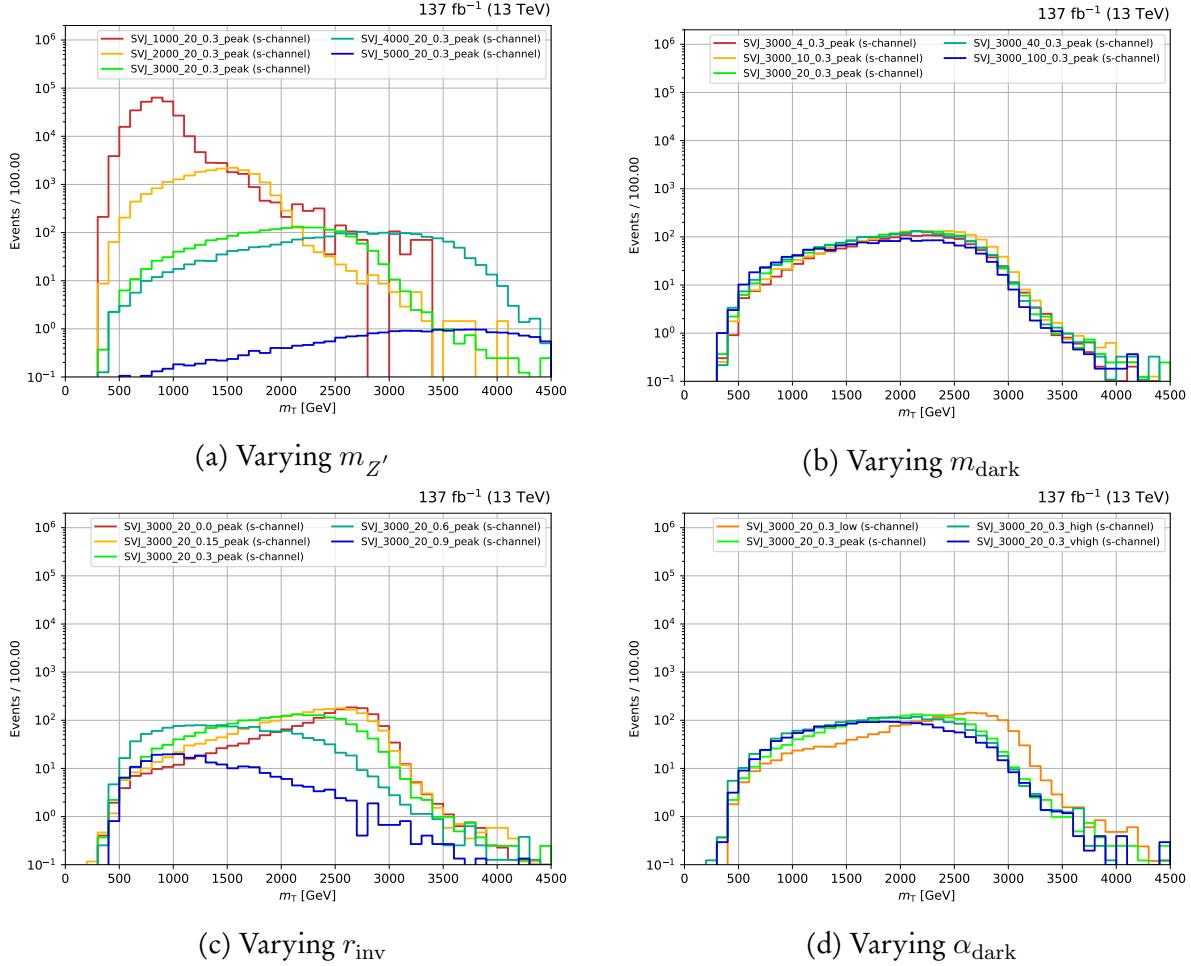


Figure 5.3: Distributions of the transverse mass of the dijet system  $m_T$  for  $s$ -channel semi-visible jet samples emulating the 2016 data taking period. In each panel, one of the free parameters of the model is varied with respect to the benchmark point SVJ\_3000\_20\_0.3\_-peak (bright green line).

Perhaps expectedly,  $m_{Z'}$  appears to have the largest influence on the kinematics of the signal. Both the normalisation and shape of the distributions are significantly affected. A lower  $m_{Z'}$  recovers the peak near the mediator mass more effectively. Surprisingly, varying the dark hadron mass does little to the distribution. Though, the transverse mass is more sensitive to the clustered jets and  $p_T^{\text{miss}}$  as a whole rather than the shower constituents themselves. The momentum of the jets will also be highly correlated to  $m_{Z'}$  since  $m_{\text{dark}}$  is usually small in comparison, boosting the dark hadrons significantly. Varying  $r_{\text{inv}}$  shows an interesting trend. A smaller value reshapes the  $m_T$  distribution toward the  $m_{Z'}$  peak. Since the visible jets in the system are naturally resolved better than the  $p_T^{\text{miss}}$ , recovering  $m_{Z'}$  is an easier task. Similarly to  $m_{\text{dark}}$ , the choices of  $\alpha_{\text{dark}}$  featured only have a small impact on the  $m_T$  spectrum. As

with  $r_{\text{inv}}$ , a smaller value leans more toward the  $m_{Z'}$  peak. This can only happen to a point, however, because a model with an  $\alpha_{\text{dark}}$  value too small refuses to shower in PYTHIA.

For both  $m_{\text{dark}}$  and  $\alpha_{\text{dark}}$ , checking many other observables leads to the same conclusion given above: the kinematics of the model do not appear particularly sensitive to changes in those parameters. This can be a positive consequence as it reduces the phase space required in the search for semi-visible jets and permits optimisation of the analysis for  $m_{Z'}$  and  $r_{\text{inv}}$ .

### 5.5.3 $t$ -channel

For the  $t$ -channel model, several pseudoscalar bifundamentals  $\Phi$  couple to fermionic dark quarks. While only one flavour of mediator exists for  $s$ -channel process, numerous exist for  $t$ -channel. Each flavour of  $\Phi$  couples to a given SM quark flavour, and is charged under electromagnetism with a value matching that of its coupled quark. The cross sections for this process are calculated by MADGRAPH at LO, varying most strongly as a function of  $m_{\Phi}$ . A small dependence on  $m_{\chi}$  is observed, and naturally no effect from  $\alpha_{\text{dark}}$  or  $r_{\text{inv}}$  since they only act at the hadronisation step.

The  $t$ -channel aspect of the signal has not been studied in as great a detail as  $s$ -channel. It also exhibits less distinguishing kinematic characteristics, making it more difficult to determine the phase space and discriminating variables to separate it from the standard model background. Nevertheless, simple distributions with a range of parameter points are presented. These are intended to demonstrate how it would manifest in the LHC, and how different the signature is from the  $s$ -channel model. Motivated by the sensitivities demonstrated by the theorists in Ref. 70, the benchmark point SVJ\_2000\_20\_0.5\_peak is chosen. As with the  $s$ -channel UV completion, each free parameter is varied individually with respect to the benchmark point, the results of which for the  $m_T$  distribution are shown in Fig. 5.4.

Analogously to the  $s$ -channel signal, the kinematics of the  $t$ -channel aspect of the model are dominated by the mediator mass and invisible fraction. The transverse mass exhibits a peak, but consistently at 700–900 GeV, seemingly regardless of  $m_{\Phi}$ . Shallower tails beyond the peak are seemingly observed for lower  $m_{\Phi}$ , although the normalisation from the cross section may be partially responsible. Though this distribution is not as distinguishing as with  $s$ -channel since there is no resonance or equivalent kinematic edge, it is more discriminatory than other variables investigated such as  $p_T^{\text{miss}}$  or  $H_T$ . Lack of statistical precision dominates the curves at higher  $m_T$  for larger  $m_{\Phi}$ . Accounting for this, however, and observing the general trend does reveal shapes comparable to smaller  $m_{\Phi}$ . For low  $m_{\Phi}$ , a reasonably large fraction of signal remains at high  $m_T$ . This is true in all panels of Fig. 5.4, suggesting a cut could be placed on the observable to reject background if not used as the search variable.

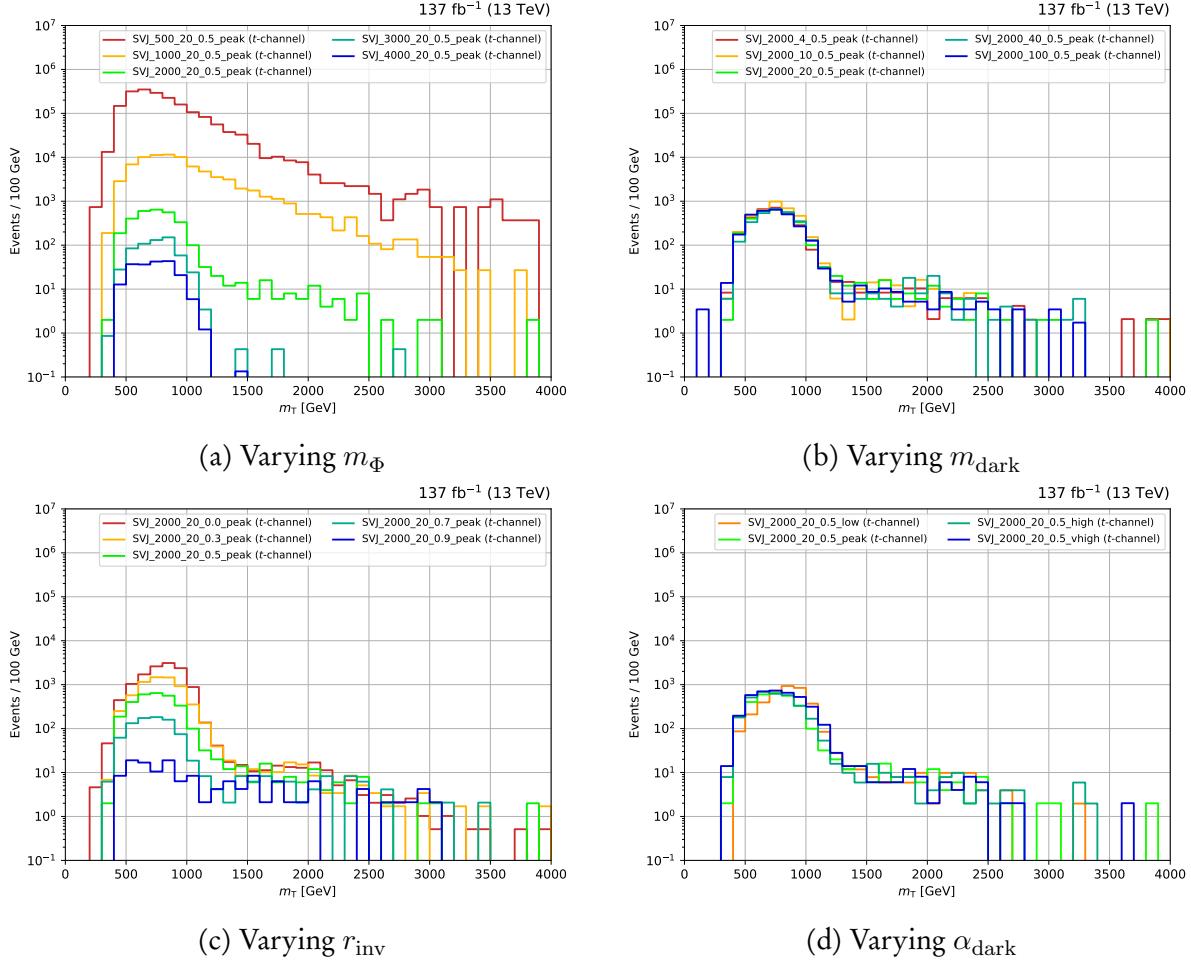


Figure 5.4: Distributions of the transverse mass of the dijet system  $m_T$  for  $t$ -channel semi-visible jet samples emulating the 2016 data taking period. In each panel, one of the free parameters of the model is varied with respect to the benchmark point  $SVJ\_2000\_20\_0.5\_peak$  (bright green line).

$m_{\text{dark}}$  and  $\alpha_{\text{dark}}$  have little effect on the  $m_T$ . Within statistical fluctuations, the distributions for varying  $m_{\text{dark}}$  look identical. It is difficult to tell if  $\alpha_{\text{dark}}$  has much effect on the kinematics overall, since the values used to vary the parameter are based on the shower of the  $s$ -channel signal. Reevaluation may be required in the  $t$ -channel case. The  $\alpha_{\text{dark}}^{\text{low}}$  curve is peaked at slightly higher  $m_T$  than the others. With a lower threshold for dark hadronisation, a greater number of dark hadrons would likely be produced meaning a greater number that decay into visible hadrons. Since measuring the visible aspect of an event is more accurate than the  $\vec{p}_T^{\text{miss}}$ , a more resolved peak would be observed.

The impact of  $r_{\text{inv}}$  on the  $m_T$  is similar to the  $s$ -channel signal insofar that the resolution of the bump is highly dependent on it. The peak near 1 TeV is recovered more effectively

with a lower invisible fraction. Again, this is likely due to the enhanced resolution of jets over  $\vec{p}_T^{\text{miss}}$ . The feature of identifying the peak at higher  $m_T$  with smaller  $r_{\text{inv}}$  may also be for the same reason.

## 5.6 PYTHIA-MADGRAPH comparisons for *s*-channel signal

PYTHIA8 is described as a multipurpose generator, meaning it has the flexibility to simulate many kinds of processes from QCD and electroweak SM, SUSY, technicolour, and leptoquarks. Its status as the de facto standard for the parton shower and hadronisation means interfacing between that step and the hard process is simple. The use of Sudakov form factors and resummation also yield realistic jet structure. However, it is difficult to simulate processes more complex than LO, and support for generating new models must sometimes be specifically added to the program.

MADGRAPH5\_amc@NLO, on the other hand, is an automatic matrix element generator. It calculates the scattering matrix element for each subprocess for a given final state,<sup>3</sup> with Feynman rules to generate each diagram within a subprocess. Integration is performed over the phase space for each subprocess to give the cross section. Generator- and phase space-level cuts are easy to implement, so that only signal in the phase space regions of interest are simulated. Systematic uncertainties from the generation step are also simple to extract in the situations they are important to an analysis. High order processes, such as NLO and NNLO, can be simulated. While extremely taxing computationally, it is a straightforward implementation on the user side.

To verify that both PYTHIA and MADGRAPH are suitable generators, 100,000 events were generated for a selection of parameter points with each program independently. Six models encompassing a sufficient range of  $m_{Z'}$ ,  $m_{\text{dark}}$ , and  $r_{\text{inv}}$  were chosen: SVJ\_1000\_20\_0.3\_peak, SVJ\_3000\_20\_0.1\_peak, SVJ\_3000\_20\_0.5\_peak, SVJ\_3000\_20\_0.9\_peak, SVJ\_3000\_50\_-0.3\_peak, and SVJ\_4000\_20\_0.3\_peak. For all the simulation steps after the initial generation (described in more detail in Chpt. 3.2.3), identical implementations were used. This was to compare the differences due only to the generator. No analysis-level selections were applied either, for the same purpose.

Below are some of the important and insightful distributions for the *s*-channel production mode:  $m_T$ , the number of jets (with the large cone size equivalent to AK8 jets in the anti- $k_T$  algorithm), the ratio of  $p_T^{\text{miss}}$  to  $m_T$ , and the minimum azimuthal angle between the  $p_T^{\text{miss}}$

---

<sup>3</sup>With these semi-visible jets models, around 50 subprocesses exist for the *s*-channel mode, while approximately 820 exist for *t*-channel.

## 5.6. PYTHIA-MADGRAPH COMPARISONS FOR S-CHANNEL SIGNAL

and two leading jets  $\Delta\phi_{\min}(j_{1,2}, p_T^{\text{miss}})$ . The ratio between the MADGRAPH-generated and PYTHIA-generated sample for a given model is also shown. For readability, plots are split into two sets comprising three models each. Fig. 5.5 showcases SVJ\_1000\_20\_0.3\_peak, SVJ\_-3000\_20\_0.1\_peak, and SVJ\_3000\_20\_0.9\_peak, while Fig. 5.6 shows SVJ\_3000\_20\_0.5\_peak, SVJ\_3000\_50\_0.3\_peak, and SVJ\_4000\_20\_0.3\_peak.

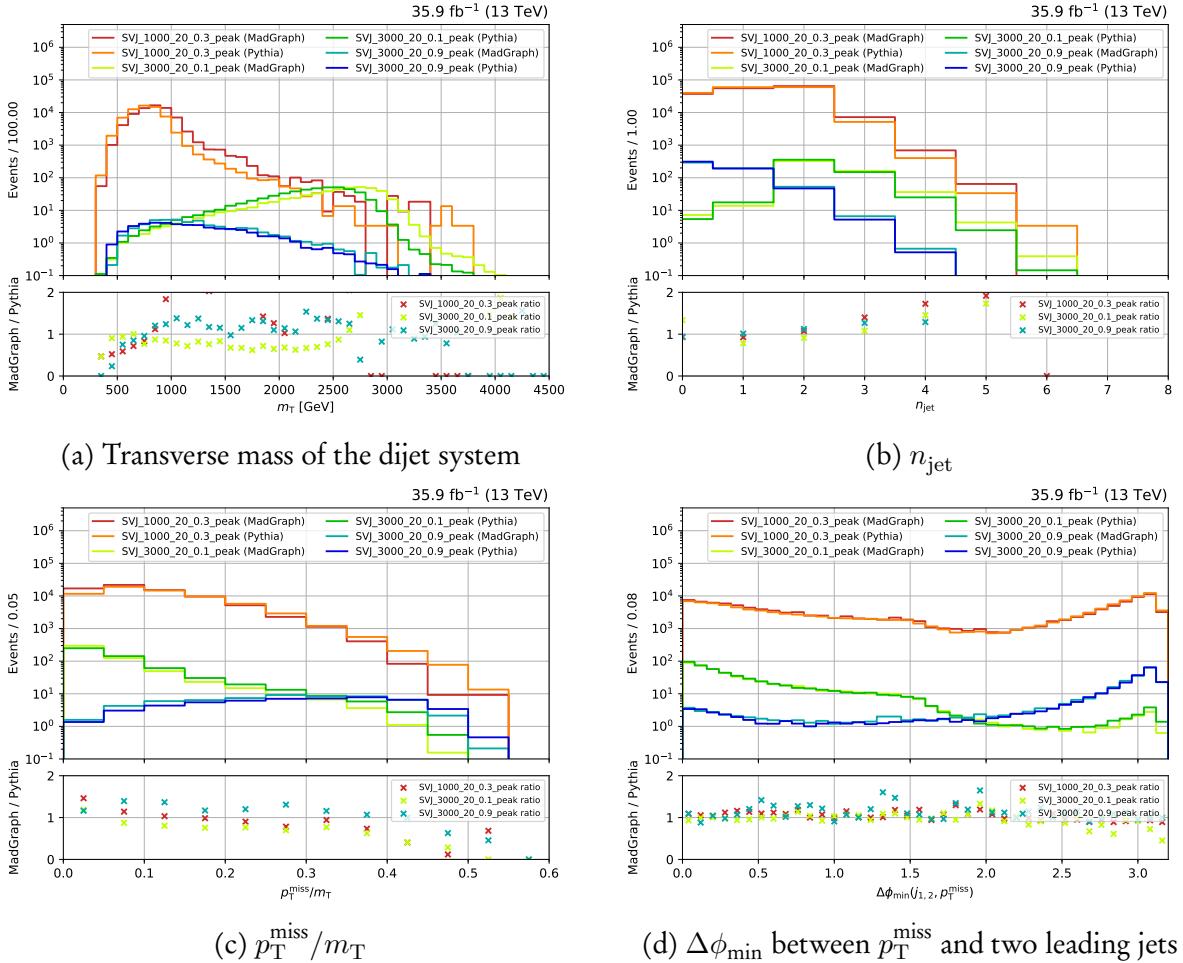


Figure 5.5: Distributions of several observables for the models SVJ\_1000\_20\_0.3\_peak, SVJ\_-3000\_20\_0.1\_peak, and SVJ\_3000\_20\_0.9\_peak. Generation in MADGRAPH5\_amc@NLO is compared to PYTHIA8, with the ratios between them for each model displayed in the respective subplot.

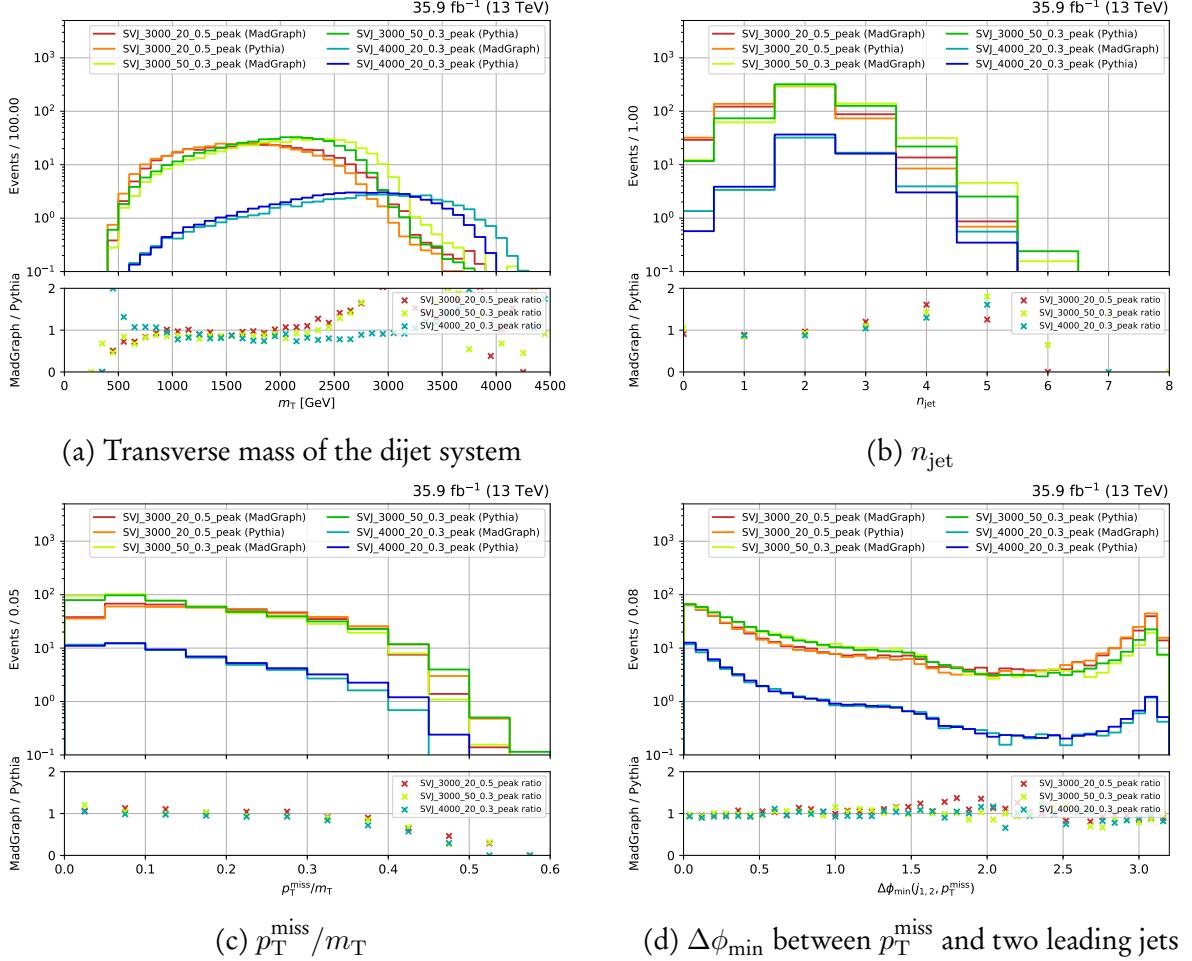


Figure 5.6: Distributions of several observables for the models `SVJ_3000_20_0.5_peak`, `SVJ_3000_50_0.3_peak`, and `SVJ_4000_20_0.3_peak`. Generation in `MADGRAPH5_amc@NLO` is compared to `PYTHIA8`, with the ratios between them for each model displayed in the respective subplot.

In general, the distributions above—as well as the others investigated such as  $p_{\text{T}}^{\text{miss}}$ ,  $H_{\text{T}}^{\text{miss}}$ , and jet  $p_{\text{T}}$ —show reasonable agreement between the two generators. In particular, position variables such as  $\eta$  and  $\phi$  of the jets and energy sums harmonise well. This is to be expected since those variables can be largely model-independent. A noticeable difference is in the  $m_{\text{T}}$  spectrum, where the shapes look similar but shifted to higher energies for the `MADGRAPH` samples. This may be attributed to the `xqcut`, or differences in the QCD scale. Tuning the `xqcut` in `MADGRAPH` or `qCut` in `PYTHIA` may be able to reduce the gap between them, especially if a relationship between the two merging scales can be derived for this model. For both generators, the peak close to the  $Z'$  mass can be recovered if the invisible fraction is small enough. This makes the transverse mass a more appealing search variable.

Both the  $n_{\text{jet}}$  and  $p_T^{\text{miss}}/m_T$  show trends in the MADGRAPH/PYTHIA ratio. A higher jet multiplicity is seen in the MADGRAPH samples, possibly due to the additional jets allowed in the matrix element calculation. For lower values of  $r_{\text{inv}}$ , a peak is seen at  $n_{\text{jet}} = 2$ , as expected given the theory. More jets may be generated from initial or final state radiation, or the dark shower prompting a large-width leading or subleading semi-visible jet causing one to be clustered as two. The larger invisible content of the high  $r_{\text{inv}}$  samples naturally shows a lower number of jets; fewer are likely to pass the qCut threshold since a larger momentum fraction of each semi-visible jet is invisible.

Larger values of  $r_{\text{inv}}$  in Fig. 5.5 manifest in an interesting  $\Delta\phi_{\min}$  spectrum. A double peak structure is demonstrated in samples with lower values. In the case of dijet events, an imbalance in the invisible content of each semi-visible jet may cause the  $p_T^{\text{miss}}$  to be aligned with the one containing the largest invisible fraction. In other scenarios, one of the two semi-visible jets may not be clustered if its invisible content is too large (i.e., if its visible content is too small). Then, in these single jet events, the  $p_T^{\text{miss}}$  would appear to recoil from the only clustered jet. Higher values of  $r_{\text{inv}}$  would cause this to happen more frequently, mimicking a WIMP signature. Failing to cluster a semi-visible jet may also result in the muddied  $m_T$  distribution as the  $\vec{p}_T^{\text{miss}}$  may not be resolved well enough.

For all of these comparisons, statistical limitation is an issue as it is only feasible to generate a rather limited sample size. Since those from PYTHIA have, by design, a higher jet matching and clustering efficiency, statistical power suffers partially for the MADGRAPH samples. Higher efficiencies are usually achieved by samples with a lower  $r_{\text{inv}}$ —expected since jet matching depends solely on the visible portion of the event.

In summary, an investigation has been performed into the generation and comprehension of the  $s$  and  $t$  production channels of semi-visible jets. Progress has been made into understanding how simulated events would manifest in the detector, how variations of the important free parameters in the model affect the kinematics, and which variables would be sensitive in a search for them. The mediator mass influences the model’s observable characteristics the most, affecting both the shapes of distributions and cross section significantly. Variables such as the transverse mass of the dijet system are able to emphasize the peak at the mediator mass in the  $s$ -channel mode, and to a lesser degree in the  $t$ -channel mechanism. Effective discrimination between signal and the dominant backgrounds is expected. Comparisons between the MADGRAPH5\_amc@NLO and PYTHIA8 event generators were also conducted for several combinations of parameters in the  $s$ -channel aspect of the model. In most cases, reasonable agreement was found, with the major differences appearing to stem from the modelling of jets. Either would be suitable for use in the main analysis.



## SEARCH FOR INVISIBLY DECAYING HIGGS BOSONS

Invisibility—there are things we can't see now, that are there, that are embedded, that it really takes time in order to be able to see. There are many ghosts that are lurking around and lingering through us that takes the technology of another generation or so in order to uncover and show what those stains and strains and perceived flaws really we're building towards.

— Lynn Hershman Leeson

Particles that escape the detector unseen in any experiment make them, by design, notoriously difficult to search for. The Higgs boson is particularly challenging with its small production rate at the LHC and a commensurate prediction of the invisible state branching ratio. As described in Chpt. 2.3, the leading experimental estimates are still far higher than the standard model's value. For the best chance of observing this decay, the inclusion of all of the Higgs boson's production modes is a necessity.

### 6.1 Overview of the analysis

The analysis discussed in the remainder of the chapter is a dedicated search for invisibly decaying Higgs bosons in hadronic final states, incorporating the  $t\bar{t}H$  and  $VH$  channels. In contrast to many of the previous analyses that separately set limits on  $\mathcal{B}(H \rightarrow \text{invisible})$

by reinterpretation, this approach has many benefits. A simultaneous search for several production modes allows the construction of search regions to target each one, embedding orthogonality to avoid overlap between them. Data and simulation samples, recipes for corrections and systematic uncertainties, the analysis framework, and results can all be shared to provide a cohesive and consistent environment from which to perform the analysis. As well as streamlining the process of the final combination over each production mechanism, communication when establishing the analysis ensures each can cover as much phase space as possible without the trouble of overlap or contamination.

This analysis was performed in collaboration with CMS, specifically by researchers from the University of Bristol, Laboratoire Leprince-Ringuet, Imperial College London, and Boston University.

### 6.1.1 Analysis strategy

The analysis targets  $t\bar{t}H$  and  $VH$  as they are novel, dedicated searches. Given the LHC is a *hadron* collider, hadronic final states are naturally chosen for this analysis. Missing transverse momentum from the purely-invisible decay of the Higgs—and hadronic activity from the particles associated with the production mechanism—constitute the final state, often known as a “jets +  $\vec{p}_T^{\text{miss}}$ ” search. In lieu of this, the dominant background processes from the standard model include QCD with high jet multiplicity, invisible decays of the  $Z$  boson, and those where the leptons from the decay are “lost” ( $\ell_{\text{lost}}$ ) from misidentification or are outside the bounds of the detector acceptance. The latter is predominantly populated by  $t\bar{t}$  and leptonically decaying  $W$  bosons. To accurately estimate their presence in the signal region of the analysis, dedicated methods are employed. Control regions separated by lepton and photon requirements predict the lost lepton and  $Z \rightarrow \nu\bar{\nu}$  processes. Sidebands to the signal region, where one or more selections otherwise designed to reject QCD multijet events are inverted, give rise to phase spaces enriched in them. A data-driven approach utilises these to predict the multijet background in the signal region.

Definitions of the physics objects used analysis-wide have already been discussed in Chpt. 4.2. An outline of the contents of this chapter is as follows. A brief summary of the software packages appears in Chpt. 6.2. The data from CMS and from simulation are then reported in Chpt. 6.3. Application of the event selection in Chpt. 6.4 is intended to separate signal from background and discard poorly measured events. Many cuts are universal, though some are category- or region-specific to mitigate problems found only there, or to further distinguish signal and background. Categorisation of the remaining events to highlight the

production modes is next illustrated in Chpt. 6.5. The binning scheme and definition of the signal region are also included.

To extract an upper limit on  $\mathcal{B}(H \rightarrow \text{invisible})$  in the absence of an excess in data, a fit of the signal and background to data is performed with the  $\text{CL}_s$  method [120]. A description of the model is given in Chpt. 6.6. Estimation of the dominant background processes using the aforementioned control regions and sidebands are described in Chpt. 6.7. Corrections to simulation, designed to model the data more accurately, and associated systematic uncertainties are discussed in Chpt. 6.8. The culmination of all of the previous sections is in the presentation of the results in Chpt. 6.9. Upper limits are provided for the  $t\bar{t}H$ - and  $VH$ -targeting categories, and for the combination over the full Run-2 dataset. Interpretations of the results of the analysis in simplified dark matter scenarios close the chapter in Chpt. 6.10.

## 6.2 Only tools and forces: software and toolkits

Analysing high energy physics data is a long and complex task with many stages that must be stitched together. In this analysis, the first step involves a light skim—or reduction of events—of the remotely-available datasets, and is effectuated with nanoAOD-tools.<sup>1</sup> Operating on the nanoAOD data tier, the repository contains centrally-maintained corrections and systematic uncertainties related to physics objects. Custom modules are applied on top. Processing the datasets on the Worldwide LHC Computing Grid, output is stored on networked university storage elements for improved performance of the later stages of the analysis.

Skimmed data is analysed predominantly using the FAST set of tools, developed by colleagues at the University of Bristol. A suite of packages harmoniously work together to run most components of an analysis. The event selection and categorisation, as well as many studies and measurements of the data are conducted at this stage. Use of vectorisation and industry-standard Python libraries such as numpy and pandas allow complex and efficient processing with simple syntax. Visualisation can also be achieved with interfaces to matplotlib.

Output from the previous stage is processed through the fit of signal and background to data. Specification of all aspects are handled with the HiggsAnalysis-CombinedLimit package.<sup>2</sup> A plethora of diagnostic information is available for understanding the effects of systematic uncertainties and other aspects of the analysis on the results.

---

<sup>1</sup>See <https://github.com/cms-nanoAOD/nanoAOD-tools> for the original fork of the repository.

<sup>2</sup>Official documentation: <https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/>.

## 6.3 Data and simulation

A summary of the datasets comprising CMS data, and simulated signal and background samples is given below.

### 6.3.1 Data

The data collected by the CMS experiment that is used in the analysis corresponds to an integrated luminosity of  $137.2 \text{ fb}^{-1}$ , and recorded from 2016–2018. A breakdown by year is given in Tab. 3.1. At over three times the volume at  $\sqrt{s} = 13 \text{ TeV}$  analysed in previous searches, there is potential to substantially lower the ceiling of  $\mathcal{B}(H \rightarrow \text{invisible})$ . Data is split into *primary datasets* that are grouped by the class of HLT path that an event triggered. The signal region, QCD sidebands, and muon control regions use the primary dataset composed of  $p_T^{\text{miss}}$  and  $H_T^{\text{miss}}$  combination triggers, where muons are excluded from the sums. The datasets made use of in the electron and photon control regions consist of triggers based on the properties of those objects. These were separate for 2016 and 2017, then merged for 2018.

Some serious issues arose during Run-2 that had to be mitigated after the events were reconstructed. With the high collision rate at CMS, in order to properly correlate the hits in the subdetectors with the correct particles and events they are attributed to, the timing infrastructure must be very precise. Timing scans are conducted frequently during runs to correct or compensate for any shifts that may occur. However, in 2016 and 2017, the gradual timing drift in the ECAL was not correctly propagated to the Level-1 trigger primitives. This resulted in a significant fraction of them in the forward direction (as the effect increases with  $\eta$ ) being mistakenly associated to the previous bunch crossing, known as *prefiring*. One rule at Level-1 forbids two consecutive bunch crossings from firing signals to the trigger system. But, a consequence of this—in addition to not finding the primitive in the nominal bunch crossing—is that events can be vetoed if a significant amount of ECAL energy is found in the  $2 < |\eta| < 3$  part of the end cap. It is observed only in data, and so a correction is applied to MC (see Chpt. 6.8.6) to account for the effect.

In 2018, a sector of the HCAL end cap lost power, leaving the approximate area bound by  $\eta \in [-3.0, -1.4]$  and  $\phi \in [-1.57, -0.87]$  inoperative for the remainder of the year. Of the  $59.7 \text{ fb}^{-1}$  recorded and certified for use in 2018,  $38.6 \text{ fb}^{-1}$  were affected. With this section missing, it was more difficult to accurately record the properties of jets, and therefore reliably calculate the  $\vec{p}_T^{\text{miss}}$ . Studies performed showed an excess in data in the affected area of the  $\phi(\vec{p}_T^{\text{miss}})$  distribution of most regions and categories. To mitigate this issue, the cuts in Chpt. 6.4.3 were added to the preselection.

### 6.3.2 Simulated signal processes

All Monte Carlo datasets in this analysis were produced centrally as outlined in Chpt. 3.2.3. The signal samples are generated at NLO with the POWHEG generator interfaced with PYTHIA8. When reweighting events for their cross section, they are specified at the highest accuracy available as given in Tab. 2.3. The samples are produced with a Higgs boson mass of  $m_H = 125$  GeV and assumes a 100 % branching ratio to invisible states. In each of the  $VH$  channels, the vector boson decays to  $q\bar{q}$ . Simulation of the  $ggH$  and VBF processes are included as their high cross sections and somewhat similar topologies would lead to a degree of overlap with the  $t\bar{t}H$ - and  $VH$ -tagged events.

When simulating 2016 data, the tuning of the shower parameters in PYTHIA was labelled CUETP8M1 [104] (the standard at the time), whilst for 2017 and 2018 datasets was replaced by the newer `cp5` version [135].

### 6.3.3 Simulated background processes

Monte Carlo datasets for many processes are required in the analysis to accurately represent the standard model background. Every one uses PYTHIA8 to hadronise the events generated by the hard scatter. The underlying event tune applied in the hadroniser follows the same prescription as with signal—CUETP8M1 for simulating 2016 while `cp5` is used thereafter. The only exception is the  $t\bar{t} + \text{jets}$  background, for which *all* samples were generated with the `cp5` tune. They are described as follows:

- $t\bar{t} + \text{jets}$ : The QCD-induced process presents a major background, largely in the  $t\bar{t}H$  channel. Large  $p_T^{\text{miss}}$  can arise in lost lepton scenarios. The three channels (hadronic, semi-leptonic, and dileptonic) are modelled with POWHEG at NLO accuracy.
- $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ : From the presence of neutrinos, the genuine  $\vec{p}_T^{\text{miss}}$  from this channel is a formidable background in the signal region. The datasets are generated at LO accuracy with `MADGRAPH5_aMC@NLO`.
- $W(\rightarrow \ell\nu) + \text{jets}$ : A QCD-induced process, this decay is dominant in the single lepton control regions. If the lepton is lost, the  $p_T^{\text{miss}}$  can be inflated and lead to a significant contribution in the signal region. The datasets are produced at LO with `MADGRAPH5_aMC@NLO`.
- QCD multijet: The most common type of event produced in  $pp$  collisions at the LHC is several jets from QCD vertices. These typically do not have large  $p_T^{\text{miss}}$ , but its high cross section leads to a significant number of events with mismeasured momenta that

can artificially increase it. As such, they serve as a non-negligible background in the analysis. The datasets are produced at LO by `MADGRAPH5_AMC@NLO`.

- Multiboson: This process encompasses the production of two (diboson) or three (triboson) electroweak bosons. They may decay into charged leptons with or without neutrinos, but also hadronically—producing jets. A mixture of the two in an event will lead to similar signatures to that of the signal. The diboson processes are both generated and showered with `PYTHIA8` at LO. For triboson events, on the other hand, `MADGRAPH5_AMC@NLO` at NLO accuracy models the hard scatter.
- Electroweak  $V + 2$  jets: The production of electroweak bosons from an electroweak vertex is a minor background in the signal region. But as above, they may produce genuine  $p_T^{\text{miss}}$  from the decay of the  $V$ . The datasets are generated at LO with `MADGRAPH5_AMC@NLO`.
- Single top: Events with one final-state top quark are a subdominant background in the analysis, but are important to consider, especially in the  $t\bar{t}H$  channel. These electroweak-induced processes include  $s$ -channel and  $t$ -channel production where a *four-flavour scheme* [108] in the event generator is used for treatment of  $b$  quarks. This approach considers the  $b$  as massive, and as such, may only enter the final state. Associated production with a  $W$  boson (known as  $tW$ ) is also considered with a five-flavour scheme, i.e.,  $b$  quarks may be considered massless and can appear in both the initial and final states. For all of these channels, the events are produced at NLO. Modelling the hard scatter for the  $s$ -channel diagram is performed with `MADGRAPH5_AMC@NLO`. The  $t$ -channel and  $tW$  mechanisms are generated with `POWHEG`, with the former decaying the  $W$  exchanged by the initial state  $b$  and  $q$  with `MADSPIN` [22] to include spin correlations.
- $t\bar{t}X + \text{jets}$ : These are rare processes where a boson  $X$  ( $\gamma$ ,  $W$ ,  $Z$ ,  $H$ ) is produced in association with a  $t\bar{t}$  pair. Several combinations of the decays of the  $t\bar{t}$  and  $X$  are covered. All of the datasets are generated at NLO. The  $t\bar{t}\gamma + \text{jets}$  and  $t\bar{t}W + \text{jets}$  datasets use `MADGRAPH5_AMC@NLO`, and decay the particles with `MADSPIN`.  $t\bar{t}Z + \text{jets}$  also uses `MADGRAPH5_AMC@NLO` but without the aforementioned addition. Meanwhile,  $t\bar{t}H + \text{jets}$  (where the  $H$  decays to visible states) is generated with `POWHEG`.
- Drell-Yan ( $Z \rightarrow \ell\ell$ ) + jets: The Drell-Yan process occurs when the quark of one incident proton annihilates with the antiquark in the oncoming proton, producing a neutral vector boson (a  $Z$  in this case) from a QCD vertex. It subsequently decays to a lepton pair. So while absent in the signal region, it is dominant in the dilepton control regions for modelling the  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$  process. The datasets are generated at LO accuracy

with `MADGRAPH5_amc@NLO`, where a dilepton mass cut of 50 GeV is applied.

- $\gamma + \text{jets}$ : Events with photons are vetoed in the signal region. However, this is the largest contributor to the  $\gamma + \text{jets}$  control region that predicts—along with the dilepton control regions—the  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$  contribution to the signal region. These datasets are generated at NLO with `MADGRAPH5_amc@NLO` in bins of  $p_T^\gamma$ .

### 6.3.4 Cross section reweighting

Cross sections are specified at the highest order available, except for the LO  $V + \text{jets}$  processes that require reweighting, expounded in Chpt. 6.8.3. Since an arbitrary number of events can be generated for simulation, and a larger number of events gives higher statistical precision, events in these datasets need to be reweighted to normalise their presence in a given region or category. To first order, the weight applied is

$$(6.1) \quad w_\sigma = \frac{\sigma \mathcal{L}_{\text{int.}}}{N \varepsilon}$$

where  $\sigma$  is the cross section of the process at the order it was generated,  $\mathcal{L}_{\text{int.}}$  is the integrated luminosity of the LHC data it is being compared to,  $N$  is the number of events in the dataset before any analysis-level cuts are applied (or the sum of the generator weights), and  $\varepsilon$  is the filter efficiency.

## 6.4 Event selection

The event selection aims to strike a balance between rejecting as many background events while retaining as much signal as possible. The preselection, in Chpt. 6.4.1, is applied to data and simulation in all analysis regions and categories to do just that. Filters to reject potentially-mismeasured events and those that lead to incorrect  $\vec{p}_T^{\text{miss}}$  calculations are documented in Chpt. 6.4.2. A strategy to combat the HEM issue faced in 2018 (detailed in Chpt. 6.3.1) is given in Chpt. 6.4.3.

### 6.4.1 Preselection

The preselection is designed to discriminate between signal and background events, and is characterised by applying the following cuts:

- $p_T^{j_1} > 80 \text{ GeV}$
- $H_T > 200 \text{ GeV}$

- $H_T^{\text{miss}} > 200 \text{ GeV}$
- $p_T^{\text{miss}} > 200 \text{ GeV}$
- $H_T^{\text{miss}}/p_T^{\text{miss}} > 0.8$
- $H_T^{\text{miss}}/p_T^{\text{miss}} < 1.2$
- $n_{\tau_{v, \text{loose}}} = 0$

To ensure orthogonality with the phase space occupied by the VBF topology, the two leading jets must be within the acceptance of the tracker, and events must fail one or more of the following criteria:

- $p_T^{j_2} > 40 \text{ GeV}$
- $\eta_{j_1} \cdot \eta_{j_2} < 0$
- $p_T^{\text{miss}} \geq 250 \text{ GeV}$
- $|\Delta\eta(j_1, j_2)| > 1.0$
- $m_{jj} > 200 \text{ GeV}$
- $\Delta\phi(j_1, j_2) < 1.5$

### 6.4.2 Additional filters

Further selections are applied to filter poorly measured or mis-reconstructed events in both data and MC. These are applied to all years, regions, and categories unless stated otherwise.

A “muon jet filter” rejects events with mis-reconstructed muons by requiring all jets with  $p_T > 200 \text{ GeV}$  to have a muon energy fraction  $f_E^\mu < 0.5$  or  $\Delta\phi(j, p_T^{\text{miss}}) < \pi - 0.4$ .

Events containing a jet in the forward region with  $p_T > 50 \text{ GeV}$  are rejected since they are more susceptible to mismeasurement in that area of the detector. An additional benefit is the further separation of signal from background.

Charged ( $f_E^{h\pm}$ ) and neutral hadron energy fraction ( $f_E^{h0}$ ) requirements are applied to all jets via fulfillment of the tight jet ID criteria (see Chpt. 4.2.1). Furthermore, stricter selections are placed on the leading two jets as follows:

- $f_E^{h\pm}(j_1) > 0.1$
- $f_E^{h0}(j_1) < 0.8$
- $f_E^{h\pm}(j_2) > 0.1$
- $f_E^{h0}(j_2) < 0.8$

In the QCD sidebands, despite the requirement of  $p_T^{\text{miss}} > 200 \text{ GeV}$ , an excess in data was observed for events with low missing transverse momentum calculated from track momenta ( $p_{T,\text{trk.}}^{\text{miss}}$ ). This indicated a significant presence of neutral particles in such events, warranting a cut of  $p_{T,\text{trk.}}^{\text{miss}} > 80 \text{ GeV}$  in the signal region and sidebands.

A discrepancy between the directions of the  $\vec{H}_T^{\text{miss}}$  and  $\vec{p}_T^{\text{miss}}$  were found to be present in both the  $t\bar{t}H$  and  $VH$  categories, prompting the requirement  $\Delta\phi(\vec{H}_T^{\text{miss}}, \vec{p}_T^{\text{miss}}) < 0.5$  everywhere. In the  $t\bar{t}H$  categories, this was further tightened. Large differences between the azimuthal angle of the missing transverse momentum calculated with PARTICLE FLOW ( $\vec{p}_T^{\text{miss}}$ ), and either from track momenta ( $\vec{p}_{T,\text{trk.}}^{\text{miss}}$ ) or jets ( $\vec{H}_T^{\text{miss}}$ ) are indicative of an inconsistent event description. An elliptical cut is therefore placed in the plane of  $\Delta\phi(\vec{p}_{T,\text{trk.}}^{\text{miss}}, \vec{p}_T^{\text{miss}})$  and  $\Delta\phi(\vec{H}_T^{\text{miss}}, \vec{p}_T^{\text{miss}})$  for events in the signal region and sidebands for  $t\bar{t}H$ :

$$(6.2) \quad \sqrt{\Delta\phi(\vec{p}_{T,\text{trk.}}^{\text{miss}}, \vec{p}_T^{\text{miss}})^2 + 4 \cdot \Delta\phi(\vec{H}_T^{\text{miss}}, \vec{p}_T^{\text{miss}})^2} < 1.0$$

The filters described below were recommended by the MET POG [61] to remove events with potentially miscalculated  $\vec{p}_T^{\text{miss}}$ :

- Primary vertex filter to remove events failing vertex quality criteria
- Beam halo filter
- HCAL barrel and end cap noise filters
- Filter for dead cells in the ECAL when constructing trigger primitives
- Filter for low-quality PARTICLE FLOW muons

There are supplementary filters applied only to data. These are to generally mitigate ECAL end cap supercrystal noise, as well as crystals where losses of transparency would otherwise require large laser corrections.

### 6.4.3 Mitigating the HEM issue

During the 2018 data taking period, the HEM issue forced additional measures to be taken that suppressed its effect on the events. For the data recorded within affected period, region-dependent selections were updated to remove events that met the following criteria:

- o  $-1.8 < \phi(\vec{p}_T^{\text{miss}}) < -0.6$  in the signal region and sidebands
- o Any veto electron  $e_{\text{veto}}$  with  $p_T > 10 \text{ GeV}$ ,  $-3.0 < \eta < -1.4$ , and  $-1.57 < \phi < -0.87$  in the  $e + \text{jets}$  and  $ee + \text{jets}$  control regions

Events in simulation that met the same criteria were instead weighted by the integrated luminosity from 2018 that was not affected by the issue (i.e.,  $21.1 \text{ fb}^{-1}$ ) rather than the entire

year. It is applicable under the assumption that both data and simulation are distributed comparably in the  $\eta$ - $\phi$  portions of the detector when the issue was not present. Given geometric variables agree very well between data and simulation in the analysis, and corrections are implemented to further synchronise them, the assumption is valid.

The effect of the HEM issue can be seen in Fig. 6.1. The  $\phi(\vec{p}_T^{\text{miss}})$  distribution is shown in a sideband to the signal region defined by inverting the  $\tilde{\omega}_{\min}$  variable (see Chpt. 6.5.3.1) since it kinematically resembles the signal region, and the data–simulation discrepancy can be removed while still blind to the data in the signal region.

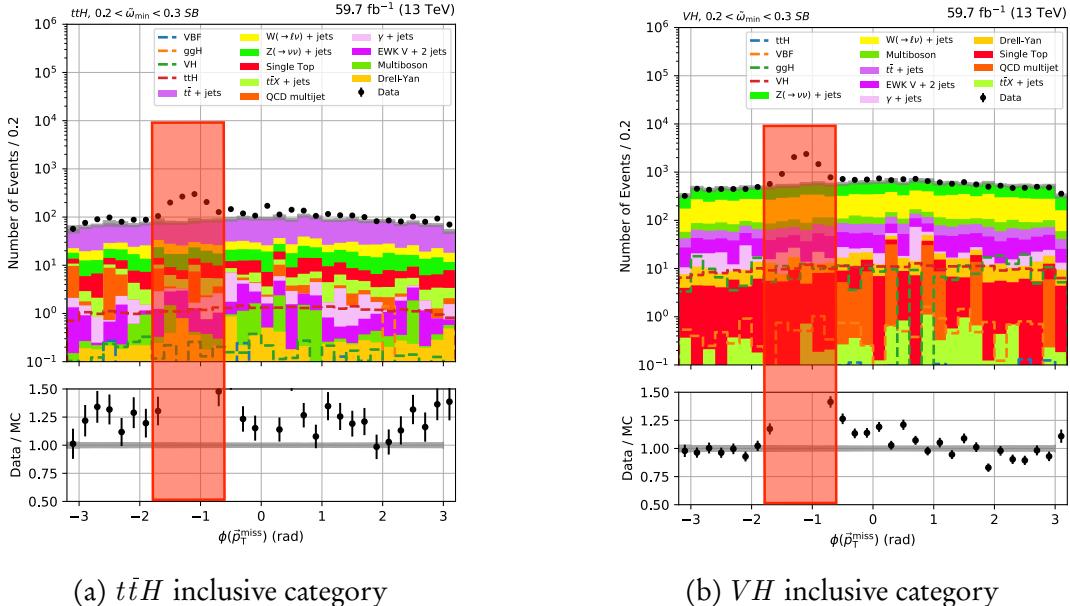


Figure 6.1: The azimuthal angle of the  $\vec{p}_T^{\text{miss}}$  inclusive in the  $t\bar{t}H$  and  $VH$  categories before applying the selections designed to mitigate the HEM issue in 2018. The loose  $\tilde{\omega}_{\min}$  sideband is used to demonstrate the effect. A red box encloses the sector that is removed by the selection applied in the signal region and sidebands.

## 6.5 Categorisation of the production modes

To extract maximum sensitivity to the  $H \rightarrow$  invisible decay from the analysis, events passing the selections in Chpt. 6.4 must be further subdivided. Categories are established to target each of the production modes. From first principles, the topologies outlined in Chpt. 2.3.1 provide an initial direction of the expected structure. Steps are taken to ensure the categories that target the production mechanism—and the categories that capitalise on specific topologies

## 6.5. CATEGORISATION OF THE PRODUCTION MODES

---

within the mechanism—are orthogonal. Additional cuts based on optimisation studies in Chpt. 6.5.3 are also implemented, devising the categories in Tab. 6.1.

Channel	Category	$n_{\text{jet}}$	$n_b$	$n_t$	$n_V$	QCD suppression	Background rejection
$t\bar{t}H$	2Boosted	$\geq 0$	$\geq 1$	2			
	1t1b	$\geq 3$	1	1	0		
	1t2b	$\geq 3$	$\geq 2$	1	0		—
	1W1b	$\geq 3$	1	0	1		
	1W2b	$\geq 3$	$\geq 2$	0	1	$\tilde{\omega}_{\min} > 0.3,$ $\Delta\phi_{\min} > 0.5$	
	5j1b	5	1	0	0		$\Delta\phi(b_1, \vec{p}_{\text{T}}^{\text{miss}}) > 1.0,$ $\Delta\phi(j_1, \vec{p}_{\text{T}}^{\text{miss}}) > \pi/2$
	6j1b	$\geq 6$	1	0	0		$\Delta\phi(b_1, \vec{p}_{\text{T}}^{\text{miss}}) > 1.0,$ $\Delta\phi(b_2, \vec{p}_{\text{T}}^{\text{miss}}) > \pi/2$
	5j2b	5	$\geq 2$	0	0		
	6j2b	$\geq 6$	$\geq 2$	0	0		
$VH$	2j0b	2	0	0	0		$m_{jj} \in [65, 105)$
	2j1b	2	1	0	0	$\tilde{\omega}_{\min} > 0.3,$ $\Delta\phi_{\min} > 0.5$	$m_{jj} \in [65, 105)$
	2j2b	2	2	0	0		$m_{jj} \in [65, 105)$
	1V	0	0	0	1		—

Table 6.1: Categorisation of the  $t\bar{t}H$  and  $VH$  production modes in the analysis. Each category highlights one of the possible final states of the mechanism, accounting for inefficiencies in object tagging or reconstruction.

In the table, the number of jets  $n_{\text{jet}}$  refers specifically to the number of AK4-clustered jets that do not overlap with a boosted  $t$  or  $V$  jet to avoid double counting objects, i.e., an AK4 jet within  $\Delta R < 0.8$  of a boosted object does not count toward the  $n_{\text{jet}}$  requirement in the category definition. This is not the case for  $b$ -tagged jets as boosted AK8 jets can conceal  $b$ -jets from the decay of the primary particle. The overlap removal between AK4 and AK8 jets only applies to the categorisation and does not affect selections made on jets in Chpt. 6.4. The number of boosted top quark- and vector boson-tagged jets are denoted as  $n_t$  and  $n_V$ , respectively.

For categorisation and the analysis altogether,  $n_{\text{jet}}$  and  $n_b$  are counted independently, so no overlap removal is performed. For example, the  $VH$  2j2b requires two jets, both of which are  $b$ -tagged.

Ancillary groupings of categories may be of interest: 2Boosted, 1t1b, 1t2b, 1W1b, and 1W2b target boosted decays of the top quark and may be collectively designated the “ $t\bar{t}H$  boosted” category; resolved decays are the focus of the remaining categories, appropriately named the “ $t\bar{t}H$  resolved” category; then, assembling the 2j0b, 2j1b, and 2j2b  $VH$  categories elicits the “ $VH$  resolved” moniker. Low yields may be observed in individual categories, so

having these grouped alternatives is useful to quickly inspect the wider effect of changes to the analysis.

Chpts. 6.5.1 and 6.5.2 explain the reasoning behind the jet-based requirements in Tab. 6.1. Optimisation of the category definitions for QCD multijet suppression and signal enhancement are described in Chpt. 6.5.3.

### 6.5.1 $t\bar{t}H$ categories

The categories comprising the  $t\bar{t}H$  channel are designed to capture both boosted and resolved topologies. The range of categories in the class are to account for inefficiencies in detection and tagging, i.e., events where only one of the top quarks or  $W$  bosons are identified. Requirements on the number of  $b$ -jets effectively distinguish background from signal.

For lower momentum events, the decay products of the top quarks are spread further apart, allowing individual jets to be reconstructed. The  $t\bar{t}H$  resolved categories, as in the boosted case, are intended to compensate for identification and/or reconstruction inefficiencies. In the 6j1b and 6j2b categories, while only six jets are expected from a  $t\bar{t}$  decay, initial and final state radiation may be present. The products of  $t\bar{t} + X$  events can also yield additional jets. While only two  $b$  quarks are expected in  $t\bar{t}H(H \rightarrow \text{invisible})$ , the  $b$ -jet requirement is left open-ended in the 5j2b and 6j2b categories. While this allows for mis-tagged  $b$ -jets to still enter the category, it is much more prevalent in true  $t\bar{t} (+X)$  decays than other processes.

### 6.5.2 $VH$ categories

As with  $t\bar{t}H$ , the categorisation of the  $VH$  channel aims to encapsulate boosted and resolved decays of the vector boson. In the resolved case, a dijet signature is expected. Both  $W$  and  $Z \rightarrow q\bar{q}$  populate the 2j0b category, while  $Z \rightarrow b\bar{b}$  ideally falls into 2j2b. Events with one of the two  $b$ -jets unidentified is the reasoning for the 2j1b category.

In high energy events, the decay products of the vector boson are merged into a single fat jet. Consequently, only a single  $V$ -tagged object is required in the 1V category and the soft drop mass window substitutes  $m_{jj}$  from its resolved counterparts.

### 6.5.3 Optimisation of the categories

While first principles are a good starting point to categorise events and accentuate the Higgs production modes, they allow much room for improvement. QCD multijet is still a prominent background, especially in the  $t\bar{t}H$  channel. Historically, variables such as  $\Delta\phi_{\min}^*$  and  $\alpha_T$  have been used to suppress it, as in Ref. 129.

### 6.5.3.1 Angular variables for QCD suppression

Recently, more elaborate variables have been developed to better remove multijet background events in analyses with hadronic final states [122]. For the analysis,  $\tilde{\omega}_{\min}$  was chosen to suppress said background. In Ref. 122, noticeable improvement is demonstrated over previous variables (particularly  $\Delta\phi_{\min}^*$ , which is used there as a benchmark for comparisons). Optimisation was performed for  $\tilde{\omega}_{\min}$  on a per-channel, over a per-category, basis to avoid excessive fine tuning for the many categories in the analysis.

The motivation behind the variable  $\tilde{\omega}_{\min}$  is to minimise the  $H_T^{\text{miss}}$  by varying the  $p_T$  of one of the jets in the event. For mismeasured jets, it is typically the magnitude rather than the direction of the transverse momentum that is affected. Hence, a jet whose  $p_T$  can be varied such that the  $H_T^{\text{miss}}$  can be greatly reduced suggests it was mismeasured. For each jet  $i$ , the ratio between its  $p_T$  and the  $H_T^{\text{miss}}$  is defined as  $f_i$ . The variable  $\sin \Delta\tilde{\phi}_i$  is the factor that scales the  $H_T^{\text{miss}}$  to its minimised value based on the varied  $p_T$  of jet  $i$ :

$$(6.3) \quad \sin \Delta\tilde{\phi}_i = \begin{cases} \sin \Delta\phi_i & \text{if } f_i + \cos \Delta\phi_i \geq 0, \\ \sqrt{1 + f_i^2 + 2f_i \cos \Delta\phi_i} & \text{otherwise} \end{cases}$$

where  $\Delta\phi_i$  is the azimuthal angle between jet  $i$  and the  $\vec{H}_T^{\text{miss}}$ . The factor is derived from geometrical arguments in the normalised  $p_T$  plane (where  $H_T^{\text{miss}} = 1$ ). Finally, the angular variable  $\tilde{\omega}_i$  is defined with respect to the aforementioned variables:

$$(6.4) \quad \begin{aligned} \tilde{\omega}_i &= \frac{\sin \Delta\tilde{\phi}_i}{f_i}, \\ \tilde{\omega}_{\min} &= \min_{i \in \text{jets}} \tilde{\omega}_i \end{aligned}$$

To reject mismeasured multijet events manifesting in jets aligned with the  $\vec{p}_T^{\text{miss}}$ , a simple cut of  $\Delta\phi_{\min}(j_{1,2,3,4}, \vec{p}_T^{\text{miss}})$ —contracted to  $\Delta\phi_{\min}$ —above 0.5 was first introduced. Correlations between  $\Delta\phi_{\min}$  and  $\tilde{\omega}_{\min}$  were found to be high at small values, though at the thresholds applied to the categories the correlation is weaker. As such, each variable is capable of suppressing multijet events with different characteristics.

Various metrics, or *figures of merit*, were considered for the threshold on which to cut in  $\tilde{\omega}_{\min}$ . In a counting experiment, a known Poisson-distributed background  $B$  yields a statistical uncertainty—assuming one standard deviation—of  $\sqrt{B}$ . A signal count  $S$  can then be statistically significant (i.e., unlikely to be a statistical fluctuation of the background) if it is greater than the uncertainty on the background. This somewhat simplistic method gives the standard deviation for the expected signal with respect to background, often just referred

to as the *expected significance*  $Z$ , as

$$(6.5) \quad Z_{\text{Poisson}} = \frac{S}{\sqrt{B}}$$

An estimate of the overall effect of systematic uncertainties on the background  $\sigma_{B, \text{syst.}}$  can also be incorporated, leading to

$$(6.6) \quad Z_{\text{Poisson}} = \frac{S}{\sqrt{B + (\sigma_{B, \text{syst.}} B)^2}}$$

Another figure of merit is to use the significance from an *Asimov dataset*—which replaces the ensemble of different datasets by a single representative one [72]. The median significance can then be extracted, colloquially ascribed the *Asimov significance*  $Z_{\text{Asimov}}$ . In many fields, including particle physics, a likelihood ratio is used for hypothesis testing. This is described in the context of the analysis in Chpt. 6.6. The asymptotic limit, where the sample size is large (as in the case for LHC data and simulation), can also be exploited as a property within the likelihood model. In this regime, the Asimov significance can be expressed as

$$(6.7) \quad Z_{\text{Asimov}} = \sqrt{2 \left( (S + B) \cdot \ln \left( 1 + \frac{S}{B} \right) - S \right)}$$

reducing to Eq. 6.5 for  $S \ll B$ . Including a systematic uncertainty expands Eq. 6.7 to

$$(6.8) \quad Z_{\text{Asimov}} = \sqrt{2 \left( (S + B)c_1 - \frac{B^2}{\sigma_{B, \text{syst.}}^2}c_2 \right)}, \text{ where}$$

$$c_1 = \ln \left( \frac{(S + B) \cdot (B + \sigma_{B, \text{syst.}}^2)}{B^2 + (S + B)\sigma_{B, \text{syst.}}^2} \right), \quad c_2 = \ln \left( 1 + \frac{\sigma_{B, \text{syst.}}^2 S}{B(B + \sigma_{B, \text{syst.}}^2)} \right)$$

While these are quantitative measures of the sensitivity with a given analysis configuration, they are only a guide to inform an analyst of a more specific area of phase space to consider, rather than providing a precise cut for the given variable. To derive the appropriate cut for each production mode, all simulated signal and background events were processed through the analysis. They were categorised according to Tab. 6.1. For each category, the significances for both methods were calculated, with and without an estimated 5 % systematic uncertainty on the background distribution. To estimate the significance for a production mode, the significances for each category were summed in quadrature. Fig. 6.2 demonstrates the results as a function of  $\tilde{\omega}_{\min}$  for the  $t\bar{t}H$  and  $VH$  categories with the 2017 datasets.

## 6.5. CATEGORISATION OF THE PRODUCTION MODES

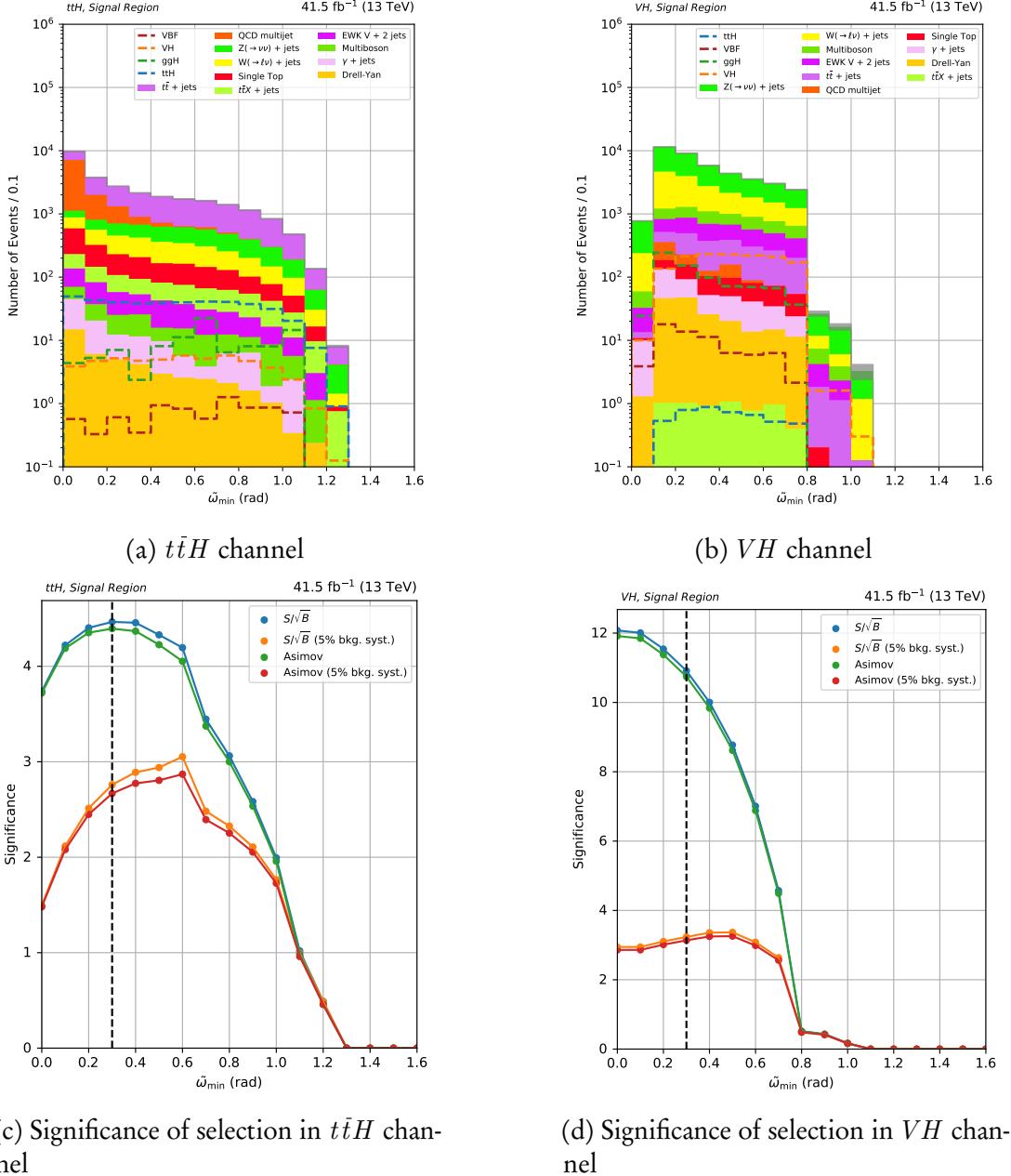


Figure 6.2: Top row: distributions of  $\tilde{\omega}_{\min}$  in the signal region inclusive in the  $t\bar{t}H$  and  $VH$  categories. Bottom row: the significance—from several figures of merit—if a cut is placed to the right of a given value. The black dotted line indicates the threshold used in the analysis. These are showcased after the analysis-level selection on signal and background simulation for the 2017 data-taking era.

The preselection from Chpt. 6.4.1 and  $\Delta\phi_{\min} > 0.5$  were applied to demonstrate the remaining multijet events that can be suppressed by  $\tilde{\omega}_{\min}$ , as well the improvements in significance.

It is obvious that the shapes and magnitudes of the significance distributions are sensitive to the inclusion of an estimated systematic uncertainty. Increasing its size affects the shape little, though impacts the magnitude more significantly. In bins with high background occupancy, even a small systematic uncertainty can wash out traces of signal. The thresholds chosen for  $\tilde{\omega}_{\min}$  do not necessarily maximise the significance—where the Asimov method with  $\sigma_{B,\text{syst.}} = 5\%$  was the leading choice—but to sufficiently separate signal and background, while not removing too many events.

### 6.5.3.2 Other optimisations

In the “background rejection” column of Tab. 6.1, selections were made to distinguish signal from leading backgrounds. The  $t\bar{t}H$  resolved categories employ requirements on the azimuthal angle between the missing transverse momentum and either the leading jet, leading  $b$ -jet, or subleading  $b$ -jet. These are specifically to differentiate  $t\bar{t}H$  signal from  $t\bar{t}$  background, as the former has been shown to exhibit larger separation between the jets and  $\vec{p}_T^{\text{miss}}$  than the latter. A window on the dijet mass is applied to the  $VH$  resolved categories to reconstruct the  $W$  and  $Z$  mass. The 15 GeV below  $m_W$  and above  $m_Z$  allow some leeway to accommodate jet resolution effects. A tight window also acts as an extra multijet background suppressor.

### 6.5.4 Binning

In addition to the categories in Chpt. 6.5, events are further separated into bins of  $p_T^{\text{miss}}$  as that distribution is expected to maximally differentiate signal and background in the fit. Since the number of events can significantly differ between categories in the different regions of the analysis, the number and widths of the bins are tuned to ensure sufficient statistical precision. These schemes are tied to the category, so are reflected in all regions such that the background estimation methods can operate on a bin-by-bin basis. The binning configuration is outlined in Tab. 6.2.

Channel	Category	$p_T^{\text{miss}}$ bins (GeV)
$t\bar{t}H$	2Boosted	[200, 300), [300, 400), [400, $\infty$ )
	1t1b	[200, 300), [300, 400), [400, $\infty$ )
	1t2b	[200, 300), [300, 400), [400, 600), [600, $\infty$ )
	1W1b	[200, 300), [300, 400), [400, $\infty$ )
	1W2b	[200, 300), [300, 400), [400, $\infty$ )
	5j1b	[200, 300), [300, 400), [400, $\infty$ )
	6j1b	[200, 300), [300, 400), [400, $\infty$ )
	5j2b	[200, 300), [300, 400), [400, $\infty$ )
	6j2b	[200, 300), [300, 400), [400, $\infty$ )
$VH$	2j0b	[200, 300), [300, 400), [400, $\infty$ )
	2j1b	[200, 300), [300, 400), [400, $\infty$ )
	2j2b	[200, $\infty$ )
	1V	[200, 300), [300, 400), [400, $\infty$ )

 Table 6.2: The binning scheme implemented to categorise events in terms of  $p_T^{\text{miss}}$ .

### 6.5.5 The signal region

The signal region is the area of phase space where the highest rate of invisible Higgs boson decays are expected, and the background to be reduced sufficiently that the potential presence of signal in data can be statistically verified. Only hadronic final states are permitted in the signal region. Events with leptons, photons and taus (defined in Chpt. 4.2) are vetoed.

Events must satisfy a logical OR of HLT cross-triggers for PARTICLE FLOW  $p_T^{\text{miss}}$  and  $H_T^{\text{miss}}$  calculated without muons (as they can also appear in the jet collection in CMS), and at least one jet fulfilling the tight ID criteria (see Chpt. 4.2). The triggers are illustrated in Tab. 6.3.

Year	$p_{T,\mu}^{\text{miss}}$ (GeV)	$H_{T,\mu}^{\text{miss}}$ (GeV)	$H_T$ (GeV)	$n_{\text{jet}}$ with tight ID
2016	90	90	—	$\geq 1$
	100	100	—	$\geq 1$
	110	110	—	$\geq 1$
	120	120	—	$\geq 1$
2017	120	120	—	$\geq 1$
	120	120	60	$\geq 1$
2018	120	120	—	$\geq 1$

Table 6.3: The trigger thresholds required for events to enter the signal region in each data taking year. Each quantity is computed with the PARTICLE FLOW algorithm at HLT level.

The relative composition in each category of the signal region (after the event selection) for signal and background are illustrated in Fig. 6.3, highlighting the effectiveness to which the categories capture the targeted topologies, and the dominant background processes, respectively. The  $t\bar{t}H$  categories, especially those designed to tag boosted top quarks and vector bosons are very pure in  $t\bar{t}H$ . Similarly, the  $VH$  categories predominantly consist of events from the production mode of focus. Contamination from the VBF process is present at a low level, most notably in the resolved categories given the similar topologies.  $ggH$  signal permeates several categories as the second largest component. This is likely due to its high cross section (more than ten times any other individual process) and its ability to mimic the final states of the other signals.  $t\bar{t} + \text{jets}$  is the dominant background for the  $t\bar{t}H$  categories as expected, with  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$  leading in the others.  $W(\rightarrow \ell\nu) + \text{jets}$  also contributes a reasonable fraction.

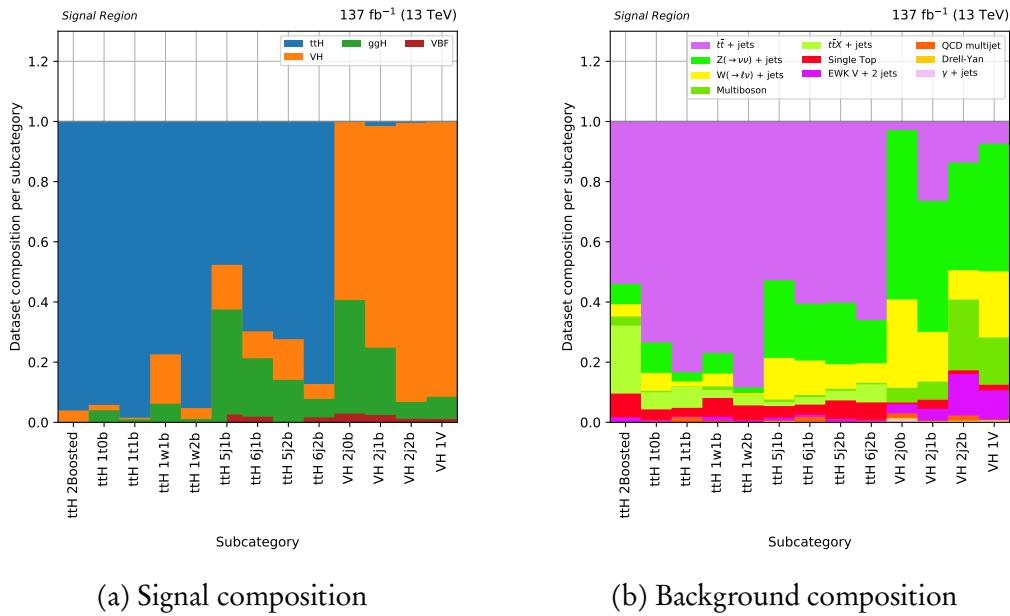


Figure 6.3: Composition of each category in the signal region for simulated signal and background processes after the analysis-level selection.

## 6.6 Statistical model for fit to data

A likelihood model is used in the fit to data, simultaneously over the signal and control regions to obtain the standard model expectation values as well as testing for signals of the  $H \rightarrow \text{invisible}$  decay. Events are categorised in two dimensions: the categories that target specific Higgs boson production modes and final state topologies as per Tab. 6.1, and in bins

of  $p_T^{\text{miss}}$  as per Tab. 6.2. The observed event counts from data in each category and  $p_T^{\text{miss}}$  bin are modelled as Poisson-distributed variables around the SM expectation with a potential contribution from signal (assumed to be zero in the null hypothesis). Expected event counts in the signal region are obtained from simulation, aided by corrections from the control regions and sidebands for electroweak and QCD multijet processes, respectively.

In the signal region, the background samples in Chpt. 6.3.3 are grouped into three processes for the purposes of background estimation in the fit:

- Lost lepton: Comprised of  $t\bar{t} + \text{jets}$ ,  $W(\rightarrow \ell\nu) + \text{jets}$ , single top,  $t\bar{t}\gamma + \text{jets}$ ,  $t\bar{t}W + \text{jets}$ ,  $t\bar{t}H + \text{jets}$ , electroweak  $W + 2 \text{jets}$ , and  $\gamma + \text{jets}$
- $Z \rightarrow \nu\bar{\nu}$ : Comprised of  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ , multiboson, electroweak  $Z + 2 \text{jets}$ ,  $t\bar{t}Z + \text{jets}$ , and Drell-Yan ( $Z \rightarrow \ell\ell$ ) + jets
- QCD: QCD multijet

Systematic uncertainties associated with simulation of both signal and background processes, discussed in Chpt. 6.8, are incorporated as nuisance parameters within the model. The likelihood function  $\mathcal{L}_{H \rightarrow \text{invisible}}$  can be summarised as

$$(6.9) \quad \mathcal{L}_{H \rightarrow \text{invisible}} = \mathcal{L}_{\text{SR}} \cdot \mathcal{L}_{\mu + \text{jets CR}} \cdot \mathcal{L}_{\mu\mu + \text{jets CR}} \cdot \mathcal{L}_{e + \text{jets CR}} \cdot \mathcal{L}_{\mu\mu + \text{jets CR}} \cdot \mathcal{L}_{\gamma + \text{jets CR}}$$

where the aim of the fit is to minimise  $-\ln \mathcal{L}_{H \rightarrow \text{invisible}}$ . The likelihood in a given region of the analysis may be written as multiple Poisson likelihoods, denoting  $\mathcal{P}(n|\lambda) \equiv \frac{e^{-\lambda}\lambda^n}{n!}$ . In the signal region,

$$(6.10) \quad \begin{aligned} \mathcal{L}_{\text{SR}}(r, a_{\ell_{\text{lost}}}, a_{Z \rightarrow \nu\bar{\nu}}, \rho) &= \prod_i \prod_{j(i)} \mathcal{P}(N_{\text{obs.}}^{i,j} | N_{\text{pred.}}^{i,j}), \text{ where} \\ N_{\text{pred.}}^{i,j} &= r \cdot s^{i,j} \cdot \rho_s^{i,j} \\ &+ b_{\ell_{\text{lost}}}^{i,j} \cdot a_{\ell_{\text{lost}}}^{i,j} \cdot \rho_{\ell_{\text{lost}}}^{i,j} \\ &+ b_{Z \rightarrow \nu\bar{\nu}}^{i,j} \cdot a_{Z \rightarrow \nu\bar{\nu}}^{i,j} \cdot \rho_{Z \rightarrow \nu\bar{\nu}}^{i,j} \\ &+ c_{\text{QCD}}^{i,j} \cdot \omega_{\text{QCD}}^{i,j} \end{aligned}$$

where the indices  $i$  and  $j$  refer to each category and  $p_T^{\text{miss}}$  bin, respectively,  $r$  is the unconstrained signal strength parameter, i.e.,  $\sigma \times \mathcal{B}(H \rightarrow \text{invisible})/\sigma_{\text{SM}}$ ,  $s$  is the signal expectation determined from simulation,  $\rho$  encodes the systematic uncertainties associated with simulation,  $b$  is the number of events from simulation,  $a$  is an unconstrained rate parameter connecting the signal and corresponding control regions,  $c$  is the predicted number of QCD multijet events, and  $\omega$  contains the uncertainties on those events. Similarly for the control

regions, the likelihood functions are

$$\begin{aligned}
 \mathcal{L}_{\mu + \text{jets CR}} &= \prod_i \prod_{j(i)} \mathcal{P}(N_{\text{obs., } \mu}^{i,j} | r \cdot s_{\mu}^{i,j} \cdot \rho_{s, \mu}^{i,j} + b_{\mu}^{i,j} \cdot a_{\ell_{\text{lost}}}^{i,j} \cdot \rho_{\mu}^{i,j}) \\
 \mathcal{L}_{\mu\mu + \text{jets CR}} &= \prod_i \prod_{j(i)} \mathcal{P}(N_{\text{obs., } \mu\mu}^{i,j} | r \cdot s_{\mu\mu}^{i,j} \cdot \rho_{s, \mu\mu}^{i,j} + b_{\mu\mu}^{i,j} \cdot a_{Z \rightarrow \nu\bar{\nu}}^{i,j} \cdot \rho_{\mu\mu}^{i,j}) \\
 (6.11) \quad \mathcal{L}_{e + \text{jets CR}} &= \prod_i \prod_{j(i)} \mathcal{P}(N_{\text{obs., } e}^{i,j} | r \cdot s_e^{i,j} \cdot \rho_{s, e}^{i,j} + b_e^{i,j} \cdot a_{\ell_{\text{lost}}}^{i,j} \cdot \rho_e^{i,j}) \\
 \mathcal{L}_{ee + \text{jets CR}} &= \prod_i \prod_{j(i)} \mathcal{P}(N_{\text{obs., } ee}^{i,j} | r \cdot s_{ee}^{i,j} \cdot \rho_{s, ee}^{i,j} + b_{ee}^{i,j} \cdot a_{Z \rightarrow \nu\bar{\nu}}^{i,j} \cdot \rho_{ee}^{i,j}) \\
 \mathcal{L}_{\gamma + \text{jets CR}} &= \prod_i \prod_{j(i)} \mathcal{P}(N_{\text{obs., } \gamma}^{i,j} | r \cdot s_{\gamma}^{i,j} \cdot \rho_{s, \gamma}^{i,j} + b_{\gamma}^{i,j} \cdot a_{Z \rightarrow \nu\bar{\nu}}^{i,j} \cdot \rho_{\gamma}^{i,j} + c_{\text{QCD}}^{i,j} \cdot \omega_{\text{QCD}}^{i,j})
 \end{aligned}$$

where the products over the indices  $i$  and  $j$  are the same as in Eq. 6.10. Signal contamination  $s$  is accounted for in all control regions. The rate parameters  $a$  are shared across the signal region and complementary control regions for the same categories and  $p_T^{\text{miss}}$  bins, i.e.,  $a_{\ell_{\text{lost}}}$  in the single lepton regions, and  $a_{Z \rightarrow \nu\bar{\nu}}$  in the dilepton and single photon regions. All non-multijet samples are grouped into a single process, represented by  $b$ .<sup>3</sup> QCD multijet MC is not included in any of the lepton control regions, but is estimated in the  $\gamma + \text{jets}$  control region from a photon purity measurement discussed in Chpt. 6.7.3.1. As in the signal region, the number of QCD events is given by  $c$ , with any associated systematic uncertainty as  $\omega$ .

Three types of fit to data are considered in the analysis: one involving only the control regions—the *control region-only* fit—where the likelihood function only comprises the terms in Eq. 6.11; a combined fit to the control regions, and the background simulation and data in the signal region, corresponding to the *background-only* hypothesis; and the fit that also includes signal simulation in the signal region, corresponding to the *signal plus background* hypothesis. The control region-only fit is used to judge the level to which the SM background prediction from those regions describes the observed data in the signal region, without the knowledge to constrain it directly. A metric for optimising the sensitivity of an analysis is the value of the expected upper limit on the signal strength parameter resulting from this fit. In the absence of a statistically significant excess of events in the signal region, the background-only fit constrains the upper limit on  $\mathcal{B}(H \rightarrow \text{invisible})$  expected in the SM, and the result of the signal plus background fit is the observed limit.

The  $\text{CL}_s$  method for setting an upper limit on  $r$  is used in the case no new physics is observed. It is often applied to set exclusion limits on non-negative parameters. Instead of simply comparing the  $p$ -value from the signal plus background (alternative) hypothesis  $p_{s+b}$ —

---

<sup>3</sup>In pre-fit or post-fit figures, the non-multijet background is labelled as “electroweak.”

obtained from the fit—to the threshold  $\alpha$ , both  $p_{s+b}$  and the  $p$ -value from the background-only (null) hypothesis  $p_b$  are involved. The alternative hypothesis is rejected, i.e., the signal model is excluded, where

$$(6.12) \quad \frac{p_{s+b}}{1 - p_b} \leq \alpha$$

As such, the upper limit on the signal strength parameter is given by the value of  $r$  in which the left-hand side of Eq. 6.12 equals  $\alpha$ . An asymptotic formula [72], whereby the ensemble of simulated datasets is replaced by the single, representative one to reduce computational expense, is applicable in the large sample limit and is incorporated into the model.

Statistical uncertainties from simulation are accommodated as a single nuisance parameter per bin based on Ref. 28. Above 10 weighted events, the uncertainty is profiled according to a Gaussian distribution centred on its value. Below that threshold, a Poisson distribution is instead invoked to provide stability in the fit for bins with small event counts. These methods are implemented, along with the likelihood function, in the `HiggsAnalysis-CombinedLimit` package.

## 6.7 Background estimation

Accurate estimation of the standard model background processes in the signal region is paramount to a search for new physics. Mismeasured backgrounds and uncertainties can wash out traces of signal and affect the fit to data. To aid in this endeavour, several regions of phase space aside from the signal region are explored in the analysis. The control regions and sidebands aid in the prediction of the SM electroweak and QCD multijet backgrounds, respectively. They are separated by High-Level Trigger requirements, and object selections in the control regions or kinematic selections in the sidebands.

The single lepton control regions are used to constrain, specifically, the lost lepton background in the signal region, arising primarily from  $t\bar{t}$  + jets and  $W(\rightarrow \ell\nu) + \text{jets}$  as a missed lepton would serve as a source of  $\vec{p}_T^{\text{miss}}$ . The dilepton and photon control regions predict the  $Z \rightarrow \nu\bar{\nu}$  background. The former parallels the decay as it is enriched in  $Z(\rightarrow \ell^+\ell^-) + \text{jets}$ —possessing the same kinematic properties while being much easier to detect, improving statistical accuracy. The latter region substitutes the  $Z \rightarrow \nu\bar{\nu}$  decay with a photon, permissible as long as the momentum of the photon is large compared to the  $Z$  mass.

These estimations rely on the distributions of data and Monte Carlo in those regions, controlled via rate parameters to connect them to the signal region. Sidebands to the signal region estimate QCD multijet contributions from data. The prediction in this case takes the

form of a *transfer factor* method—a constrained application of a rate parameter. Prediction of the background yields from each of these methods is done so within the fit and expressed in the likelihood functions. Further explanation of the mechanics is given in the remainder of this section. Fig. 6.4 illustrates the correspondence between the analysis regions and background predictions.

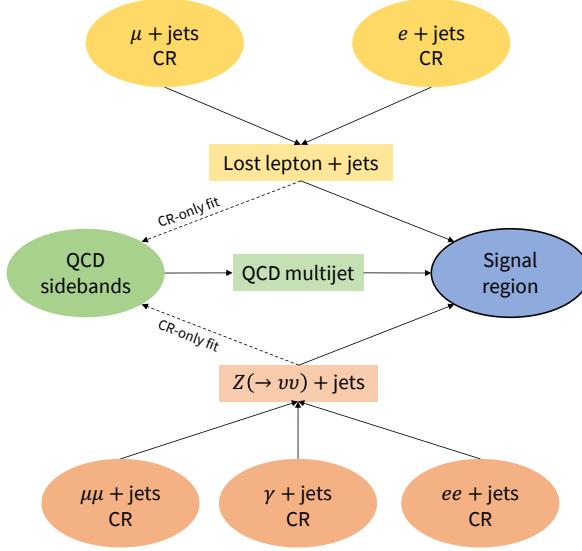


Figure 6.4: An infographic showcasing the role of each analysis region in the fit to data. The control regions predict the lost lepton and  $Z \rightarrow \nu\bar{\nu}$  backgrounds, and a control region-only fit informs the QCD multijet prediction that contributes to the eventual background determination in the signal region.

### 6.7.1 Control regions

Control regions serve two complementary purposes in many analyses: the prediction of certain backgrounds that dominate in the signal region, as a more accurate method than using the yields directly from Monte Carlo; and to validate the data, MC, and corrections applied. They are orthogonal to the signal region and to each other by way of lepton or photon requirements, and by triggers that may pertain to those objects. Control regions are designed to, ideally, be devoid of signal given their role as SM descriptors. Contamination is sometimes present, but at a very small level. Five control regions are used in the analysis:  $\mu + \text{jets}$ ,  $\mu\mu + \text{jets}$ ,  $e + \text{jets}$ ,  $ee + \text{jets}$ , and  $\gamma + \text{jets}$ . The criteria for the objects—which are defined explicitly in Chpt. 4.2—are summarised in the list below:

- $\mu + \text{jets}$ : one tight muon  $\mu_{\text{tight}}$  with  $p_T > 20 \text{ GeV}$  and a transverse mass (as calculated in Eq. 4.6) in the range  $50 < m_T^\mu < 110 \text{ GeV}$

- $\mu\mu + \text{jets}$ : one tight muon  $\mu_{\text{tight}}$  with  $p_T > 20 \text{ GeV}$ , and one loose muon  $\mu_{\text{loose}}$  with  $p_T > 10 \text{ GeV}$  that has opposite charge, with a combined invariant mass of  $60 < m_{\mu\mu} < 120 \text{ GeV}$  in the  $VH$  categories, while instead  $75 < m_{\mu\mu} < 105 \text{ GeV}$  in the  $t\bar{t}H$  categories. The leading muon is required to possess  $p_T > 110 \text{ GeV}$
- $e + \text{jets}$ : one tight electron  $e_{\text{tight}}$  with  $p_T > 40 \text{ GeV}$  and  $50 < m_T^e < 110 \text{ GeV}$
- $ee + \text{jets}$ : one tight electron  $e_{\text{tight}}$  with  $p_T > 40 \text{ GeV}$ , and one veto electron  $e_{\text{veto}}$  with  $p_T > 10 \text{ GeV}$  that has opposite charge, with a combined invariant mass of  $60 < m_{ee} < 120 \text{ GeV}$  in the  $VH$  categories, while instead  $75 < m_{ee} < 105 \text{ GeV}$  in the  $t\bar{t}H$  categories. The leading electron is required to possess  $p_T > 110 \text{ GeV}$
- $\gamma + \text{jets}$ : one medium photon  $\gamma_{\text{medium}}$  with  $p_T > 230 \text{ GeV}$

The event selection for the control regions mirrors the signal region almost exactly, with few exceptions that are detailed in Chpt. 6.4.2. The only other differentiation is the QCD suppression criteria from Tab. 6.1. These are applied in the signal region, but not to any of the control regions for the  $t\bar{t}H$  categories. This was to improve the statistical accuracy in the control regions, notably in the otherwise depleted dilepton regions, and was found to avoid compromising the data–simulation agreement and the shapes of the important distributions. For the  $VH$  categories, the QCD suppression criteria were included in the control regions as the shape of the  $p_T^{\text{miss}}$  distribution was found to be different depending on whether the criteria were included or not. The control regions in those categories were also not deficient in events, unlike in the case of  $t\bar{t}H$ .

The transverse mass cuts in the single lepton control regions reduce the effect of signal contamination, specifically from  $t\bar{t}H$  since its  $m_T$  generally eclipses that of  $t\bar{t}$ .<sup>4</sup> The trigger requirements for the  $\mu + \text{jets}$  and  $\mu\mu + \text{jets}$  control regions are the same as for the signal region (Tab. 6.3) since the same primary dataset is used.

In the  $\mu\mu + \text{jets}$  and  $ee + \text{jets}$  control regions, the  $m_{\ell\ell}$  window is halved for the  $t\bar{t}H$  categories compared to the others to reduce contamination from dileptonic  $t\bar{t}$  decays, granting a higher purity  $Z(\rightarrow \ell^+\ell^-) + \text{jets}$  region. The leading lepton  $p_T$  requirement is increased in these regions for all categories, also for purity purposes.

The  $e + \text{jets}$  and  $ee + \text{jets}$  control regions take advantage of the increased statistical power of two primary datasets in both 2016 and 2017, characterised by electron and photon triggers. In 2018, they were merged into a single  $e/\gamma$  primary dataset. Tabs. 6.4 and 6.5 elucidate how the trigger requirements are specified for each year. As before, each quantity and object is

---

<sup>4</sup>The  $m_T$  for a  $t\bar{t}$  event with  $\vec{p}_T^{\text{miss}}$  solely from the neutrino (in  $t \rightarrow bW$ ,  $W \rightarrow \ell\nu$ ) should be in a window around the  $W$  mass. Introducing additional decay products such as the Higgs boson increase it.

defined at HLT level. If an event aims to enter the  $e + \text{jets}$  or  $ee + \text{jets}$ , and is from the dataset of  $e$ -based triggers, the criteria for either of the two triggers for the given year in Tab. 6.4 must be satisfied. If an event aims to enter either control region and is from the primary dataset of  $\gamma$ -based triggers, the failure of both of the electron triggers in Tab. 6.4 and the passing of any of the photon triggers in Tab. 6.5 are required. This condition avoids double counting events that also appear in the electron trigger-based dataset. In 2018, any of the year’s triggers in Tabs. 6.4 and 6.5 may be satisfied since there is only one primary dataset. For Monte Carlo events, as the datasets are not categorised by trigger, they may pass any of the requirements for their respective year in either table.

Year	$E_{\text{T}, \text{SC}}^e$ threshold (GeV)	$e$ WP	Calorimeter ID	GSF track-SC matching
2016	27	Tight	—	—
	105	—	Very tight	Tight
2017	35	Tight	—	—
	115	—	Very tight	Tight
2018	32	Tight	—	—
	115	—	Very tight	Tight

Table 6.4: The trigger requirements for events to enter the  $e + \text{jets}$  or  $ee + \text{jets}$  control regions, if they originate from the dataset of  $e$ -based triggers. Selections are on the transverse energy of the supercluster  $E_{\text{T}, \text{SC}}$ , the working point of the candidate electron, ID of the candidate in the calorimeters, and the matching between the gaussian sum filter (GSF)-fitted track and supercluster, all at HLT level.

Year	$E_{\text{T}}^{\gamma}$ threshold (GeV)	$H/E$
2016	165	$< 0.1$
	175	—
2017	200	—
2018	200	—

Table 6.5: The trigger requirements for events to enter the  $e + \text{jets}$   $ee + \text{jets}$ , or  $\gamma + \text{jets}$  control region, if they originate from the dataset of  $\gamma$ -based triggers. Selections are on the transverse energy  $E_{\text{T}}$  of the candidate, and the ratio of the candidate’s central energy deposit in the HCAL to the ECAL ( $H/E$ ), all computed at HLT level.

For the  $\gamma + \text{jets}$  control region, data is used from CMS originating only from the dataset of  $\gamma$ -based triggers. Data and MC must satisfy any of the triggers from Tab. 6.5 for the respective year.

In each of the control regions, the  $\vec{p}_T^{\text{miss}}$  is recalculated without the objects used to define said region as a proxy for  $\vec{p}_T^{\text{miss}}$  in the signal region. Conditions in the event selection and binning that refer to  $\vec{p}_T^{\text{miss}}$  or  $p_T^{\text{miss}}$  use this recalculated quantity when applied to the control regions. Using the  $\mu + \text{jets}$  control region as an example, the new  $\vec{p}_T^{\text{miss}}$  is the vector sum of the old  $\vec{p}_T^{\text{miss}}$  and the tight muon's  $\vec{p}_T$ .

The relative composition of the background processes in each control region after the analysis-level selection is summarised in Fig. 6.5. As desired, the single lepton control regions are very similar and dominated by the  $t\bar{t}$  and  $W \rightarrow \ell\nu$  processes. The dilepton regions also possess a very similar composition to each other, enriched in Drell-Yan  $Z \rightarrow \ell\ell$ , and to a lesser extent  $t\bar{t}$  and  $t\bar{t}X$  which contain several dileptonic decay channels. In the  $\gamma + \text{jets}$  control region, expectedly the  $\gamma + \text{jets}$  process dominates virtually every category. QCD multijet events make up a consistent, small fraction of the events. Rather than the yields from MC as they would likely consist of jets misidentified as photons, its presence in this region is estimated from a data driven purity measurement explained in Chpt. 6.7.3.1.

### 6.7.2 Lost lepton background estimation

In order to predict the lost lepton background in the signal region, a freely floating rate parameter  $a_{\ell_{\text{lost}}}$  is introduced in the fit. It is shared across the single lepton control regions and signal region, where it scales the event count from simulation in these regions to obtain the expected values. Events are categorised and binned in the same manner in the single lepton control regions as in the signal region, such that the rate parameters—and therefore the prediction—is derived bin-by-bin. The predicted number of lost lepton events in Eqs. 6.10 and 6.11 is then simply  $b_{\ell_{\text{lost}}} \cdot a_{\ell_{\text{lost}}}$ .

Distributions of the single lepton control regions after the control region-only fit are illustrated in Figs. 6.6 and 6.7 in the  $t\bar{t}H$  and  $VH$  categories, respectively, for the 2017 dataset. Corresponding event counts, along with the rate parameters, can be found in Tab. A.3. Equivalent figures for the other data taking periods are accessible in Apps. A.1 and A.2 for the  $t\bar{t}H$  and  $VH$  categories, respectively.

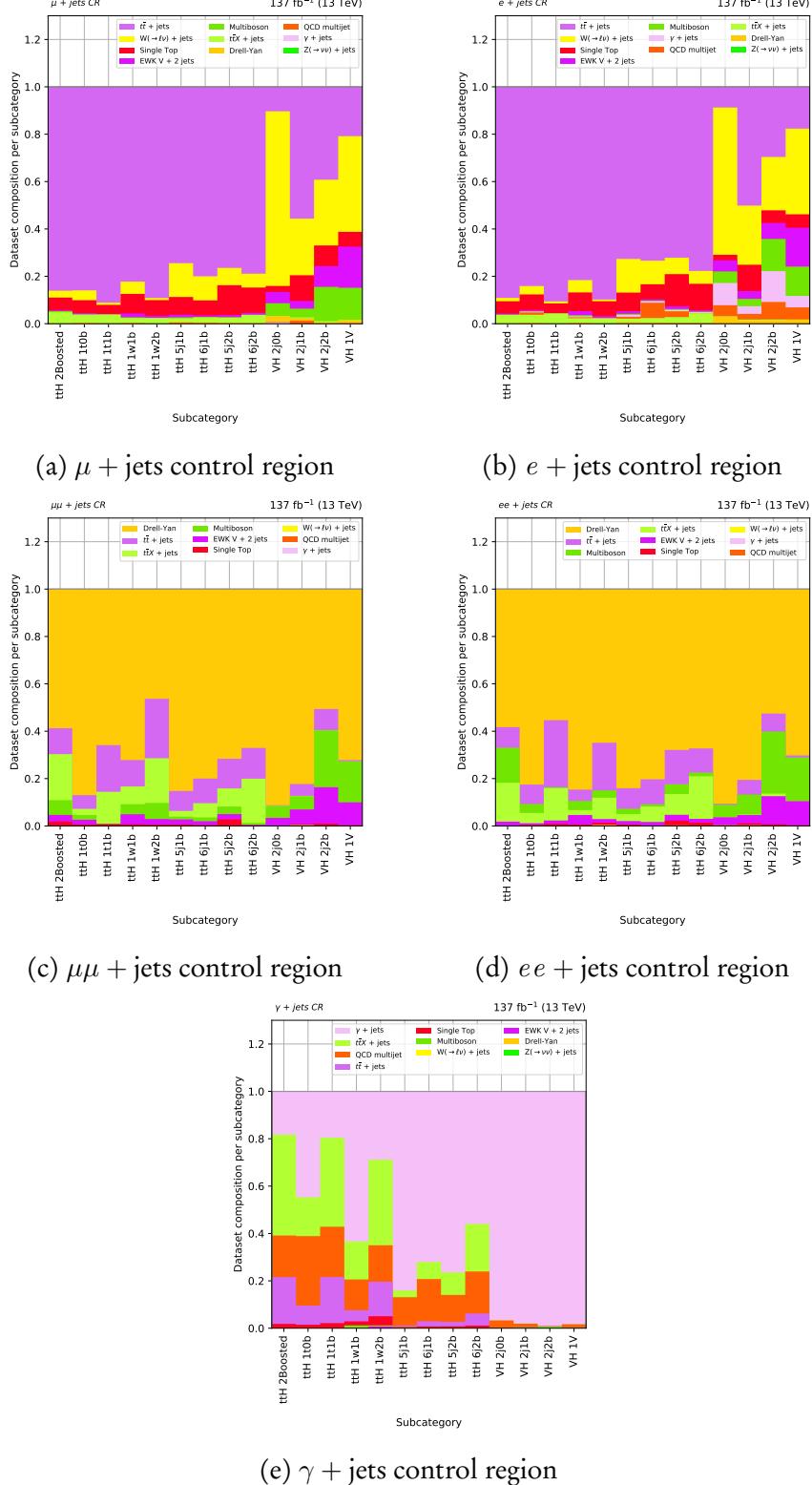


Figure 6.5: Composition of each category in each of the control regions for simulated background processes after the analysis-level selection.

## 6.7. BACKGROUND ESTIMATION

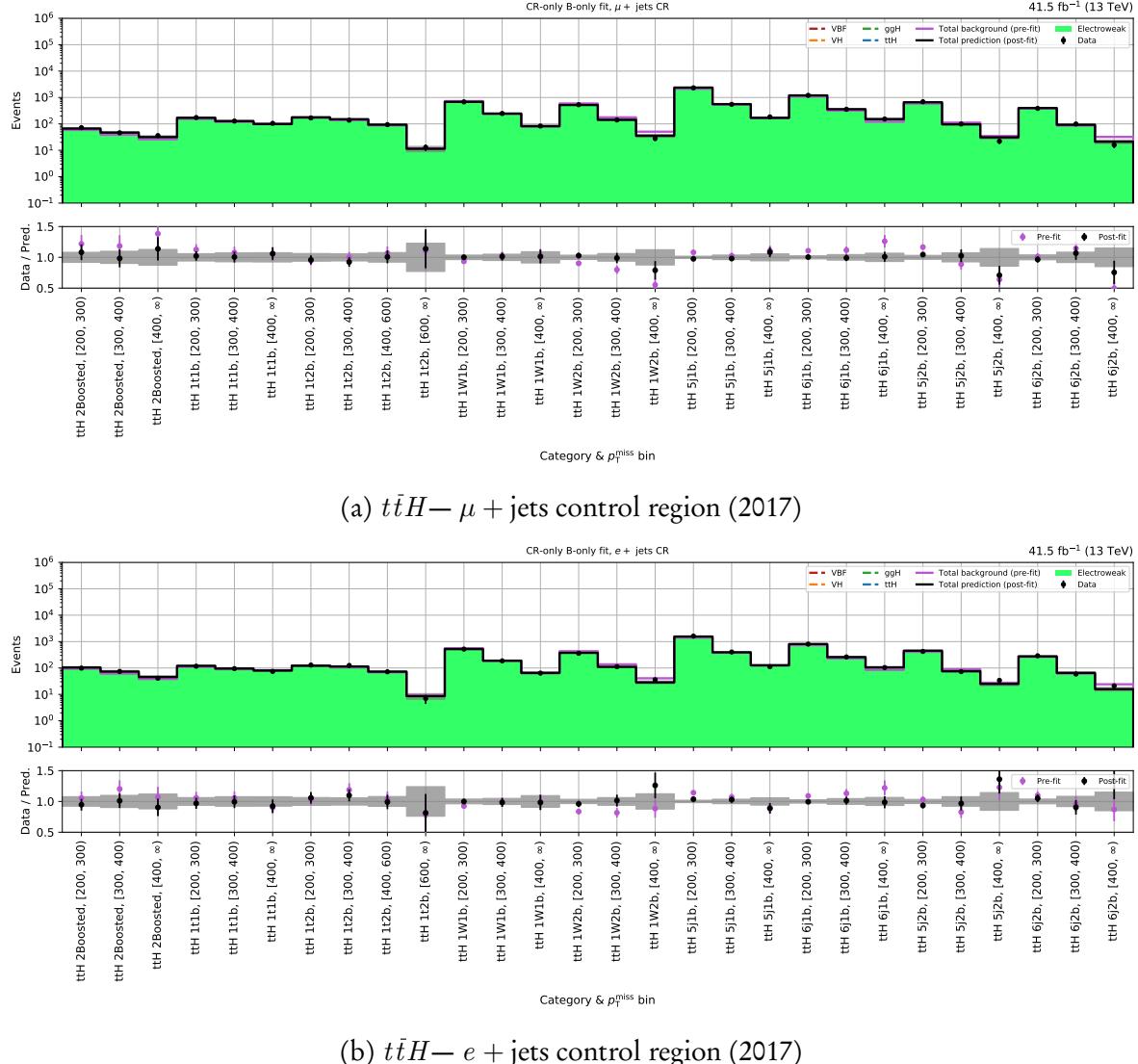


Figure 6.6: Post-fit yields for each category and  $p_T^{\text{miss}}$  bin in the single lepton control regions of the  $t\bar{t}H$  categories for the 2017 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subfigure.

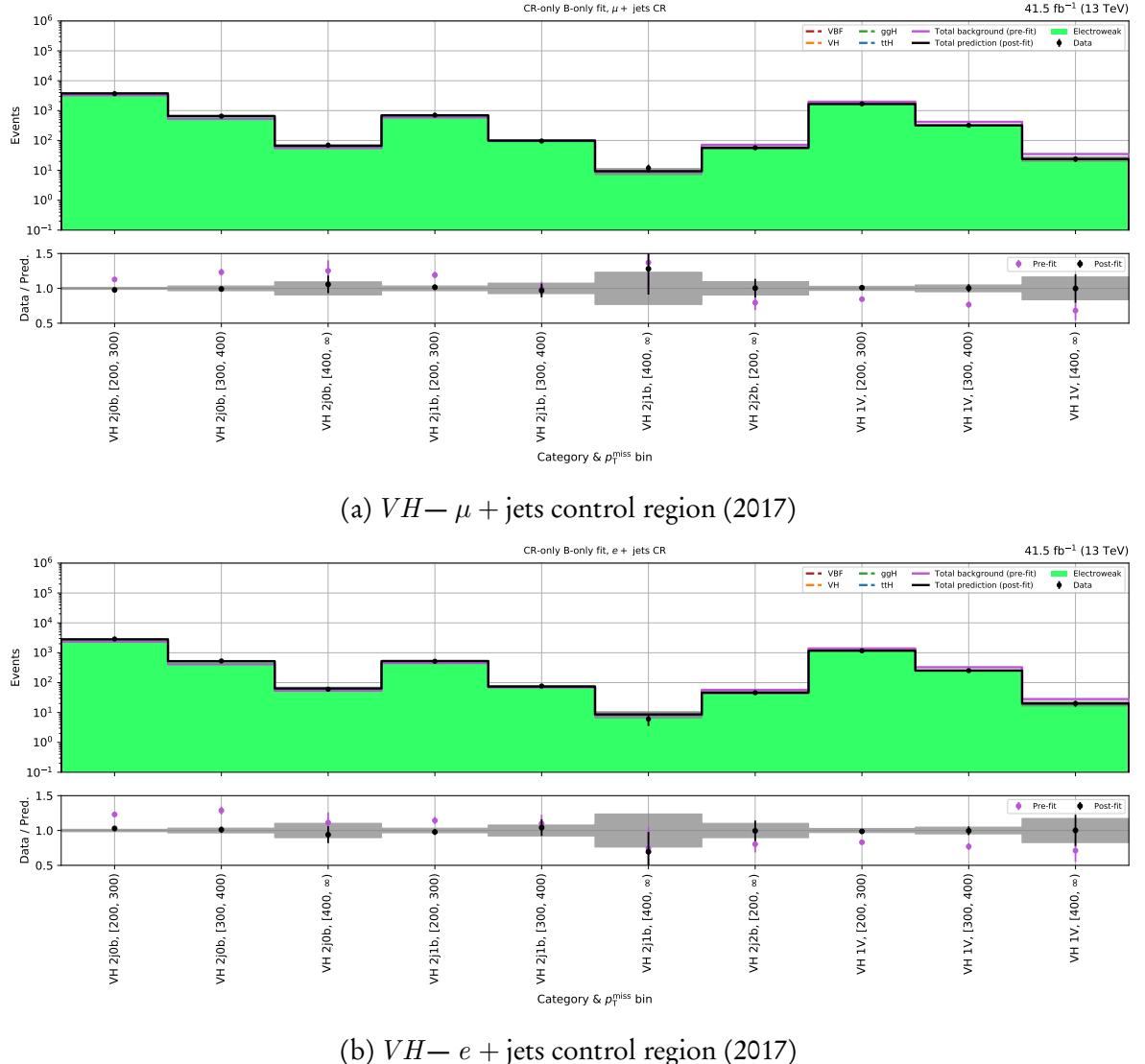


Figure 6.7: Post-fit yields for each category and  $p_T^{\text{miss}}$  bin in the single lepton control regions of the  $VH$  categories for the 2017 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subfigure.

The fit performs adequately in bringing the simulation in closer agreement to the data, despite large initial disagreement in some bins. All of the bins are sufficiently populated, such that the uncertainty on the prediction is small in most cases. The predicted SM background counts in the signal region from Tab. 6.10 further corroborates the results as the yields are close to, if not above than the data in several bins.

### 6.7.3 $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ background estimation

The irreducible background from invisibly decaying  $Z$  bosons is estimated in the same manner as the lost lepton background. A rate parameter  $a_{Z \rightarrow \nu\bar{\nu}}$  ties together the dilepton control regions, the photon control region, and the signal region. The yields from simulation are scaled by the best fit value of the parameter, obtained during the simultaneous fit to the analysis regions. As with the lost lepton background, event categorisation is the same in these regions, allowing the predictions to be derived independently for each category and  $p_T^{\text{miss}}$  bin.

For the  $t\bar{t}H$  categories, the  $\gamma + \text{jets}$  control region is not used for the  $Z \rightarrow \nu\bar{\nu}$  prediction. The region is leveraged by the  $VH$  categories, however. In addition to the electroweak background dominated by  $\gamma + \text{jets}$  simulation, the QCD multijet presence—consisting of jets misidentified as photons—in the photon control region is estimated by way of a data-driven purity measurement. This is detailed extensively in Chpt. 6.7.3.1. The results of a control region-only fit to data for the 2017 dataset are shown as an example in Chpt. 6.7.3.2.

#### 6.7.3.1 Photon purity measurement for the $\gamma + \text{jets}$ control region

Photons are reconstructed from clusters in the ECAL. They can usually be discriminated from other sources leaving ECAL deposits due to the properties of the deposits themselves, as well as the lack of other signatures that typically belong to other particles. However, this method is imperfect, and occasionally other particles will be incorrectly identified as photons (which are known as “fakes”). The leading sources of fake photons is from QCD multijet events where a jet is misidentified as such. Due to the high cross section of the process, even a small rate of fake photons becomes important to consider.

In order to separate real photons from fakes in the  $\gamma + \text{jets}$  control region, a purity measurement is performed. The purity is defined as the fraction of reconstructed photons that are from an isolated photon emerging from the hard scatter of the event, rather than a fake. The variable  $\sigma_{inj}$  is able to distinguish between real and fake photons with sufficient power. A peak with a hard cut off at  $\sigma_{inj} \approx 0.01$  is observed for real photons, while fakes

possess a less pronounced peak and much slower decline above that threshold. As such, a template fit is performed to the distribution in data to extract the purity.

As inputs to the fit, photons in data are selected by applying the medium identification requirements from Tab. 4.6 with the exception of  $\sigma_{in\eta}$  to observe the full range. Photons from  $\gamma + \text{jets}$  simulation are selected with the same criteria and are used to define the real photon template. A fake photon template is obtained from data by requiring at least one of the isolation criteria from the medium ID in Tab. 4.6 to be unfulfilled. This ensures the photons from this set do not overlap with the real photons from data.

The templates are derived in separate bins of photon  $p_T$ , and the purity measurement is performed separately for each data taking year. The following event selection is applied:

- Photon trigger requirement for the respective dataset from Tab. 6.5
- $p_T^{\text{miss}}$  filters from Chpt. 6.4.2
- $p_T^{\text{miss}} < 60 \text{ GeV}$
- At least one jet with  $p_T > 80 \text{ GeV}$  and  $|\eta| < 2.4$ , separated from photons with  $\Delta R > 0.4$
- If a second jet is present, it is required to have  $|\eta| < 2.4$  and also be separated from photons with  $\Delta R > 0.4$
- $H_T > 200 \text{ GeV}$

A combination of cuts are used to select photons appropriately and ensure the phase space resembles that of the  $\gamma + \text{jets}$  control region. The shapes of the real and fake templates are fit to the data using a likelihood function in the range  $0.004 \leq \sigma_{in\eta} \leq 0.02$ . Two examples, for the lowest and highest  $p_T$  bins from the 2017 dataset, are shown in Fig. 6.8.

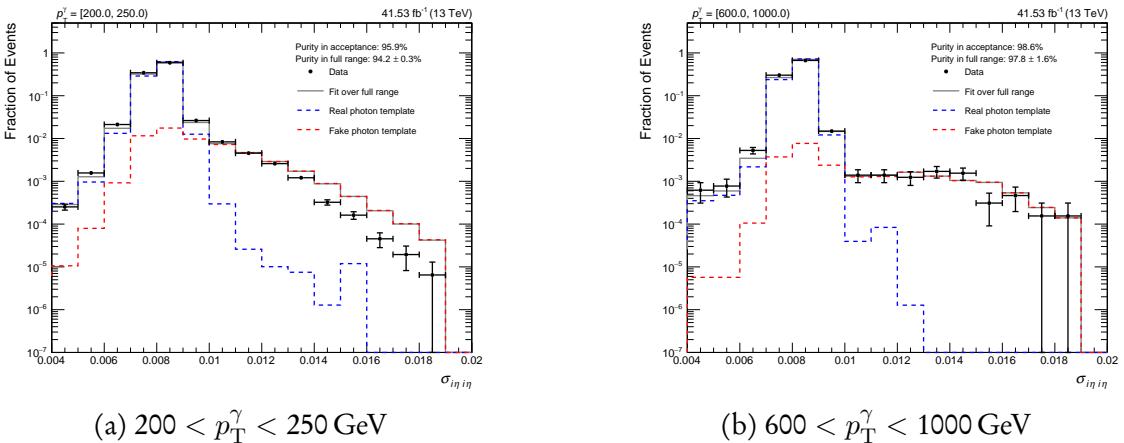


Figure 6.8: Fits of real and fake photon templates to data in the  $\sigma_{in\eta}$  spectrum. Two bins of  $p_T^\gamma$  are illustrated, demonstrating the different shapes of the templates and data, with the purity over the full range and within acceptance ( $\sigma_{in\eta} \leq 0.01$ ) noted.

It can be seen that the data resembles the real template at low values of  $\sigma_{inj\eta}$  with the fake template dominating at larger values. Each fit captures the features of both templates appropriately. By calculating the purity within acceptance (i.e.,  $\sigma_{inj\eta} \leq 0.01$  as given by the medium ID requirement), the *impurity* can be derived as a function of photon  $p_T$ . An exponential function is fitted to interpolate within the range that also serves to extrapolate above it. Fig. 6.9 illustrates the impurity versus photon  $p_T$  for each data taking year. A 25 % uncertainty around the fit is assumed to account for effects related to the binning of the  $\sigma_{inj\eta}$  distribution.

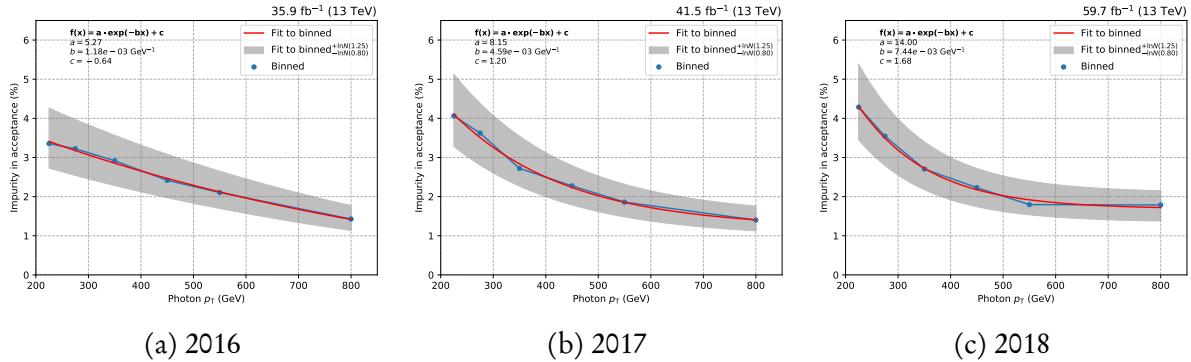


Figure 6.9: The fraction of impure photons in data as a function of  $p_T$  for each data taking year in Run-2. An exponential is fit to the binned data with a 25 % uncertainty assigned to account for binning effects.

The impurity expectedly decreases as a function of the photon  $p_T$ . Harder particles are usually reconstructed better in the detector with more isolated deposits in the calorimeters while maintaining excellent momentum resolution. Several factors affecting the data quality in 2017 and 2018 could have led to the larger fake rates in the distributions, such as pre-firing and the HEM issue. Few issues were present in 2016. Monte Carlo in each year may have also impacted the measurement as the modelling parameters often differ between years.

In the analysis, the purity measurement is used to estimate the QCD multijet background in the  $\gamma + \text{jets}$  control region, replacing the contribution from MC. For each event in data that enters the region, a QCD multijet pseudo-event is created with the same properties—notably  $p_T^{\text{miss}}$  and photon  $p_T$ . The value of the impurity calculated from the photon's  $p_T$  weights the event. The new QCD background is therefore generated with the same shape as the data and weighted to represent the rate of non-prompt photons. A 25 % uncertainty is attributed to the normalisation of the yield.

### 6.7.3.2 Results from a control region-only fit

Distributions of the dilepton and photon control regions from the control region-only fit are shown in Figs. 6.10 and 6.11 in the  $t\bar{t}H$  and  $VH$  categories, respectively, for the 2017 dataset. Corresponding event counts, along with the rate parameters, can be found in Tab. A.4. The predicted invisible  $Z$  background in the signal region, for comparisons to the observed data, is tabulated in Tab. 6.10. Equivalent figures for the other data taking periods are accessible in Apps. A.1 and A.2 for the  $t\bar{t}H$  and  $VH$  categories, respectively.

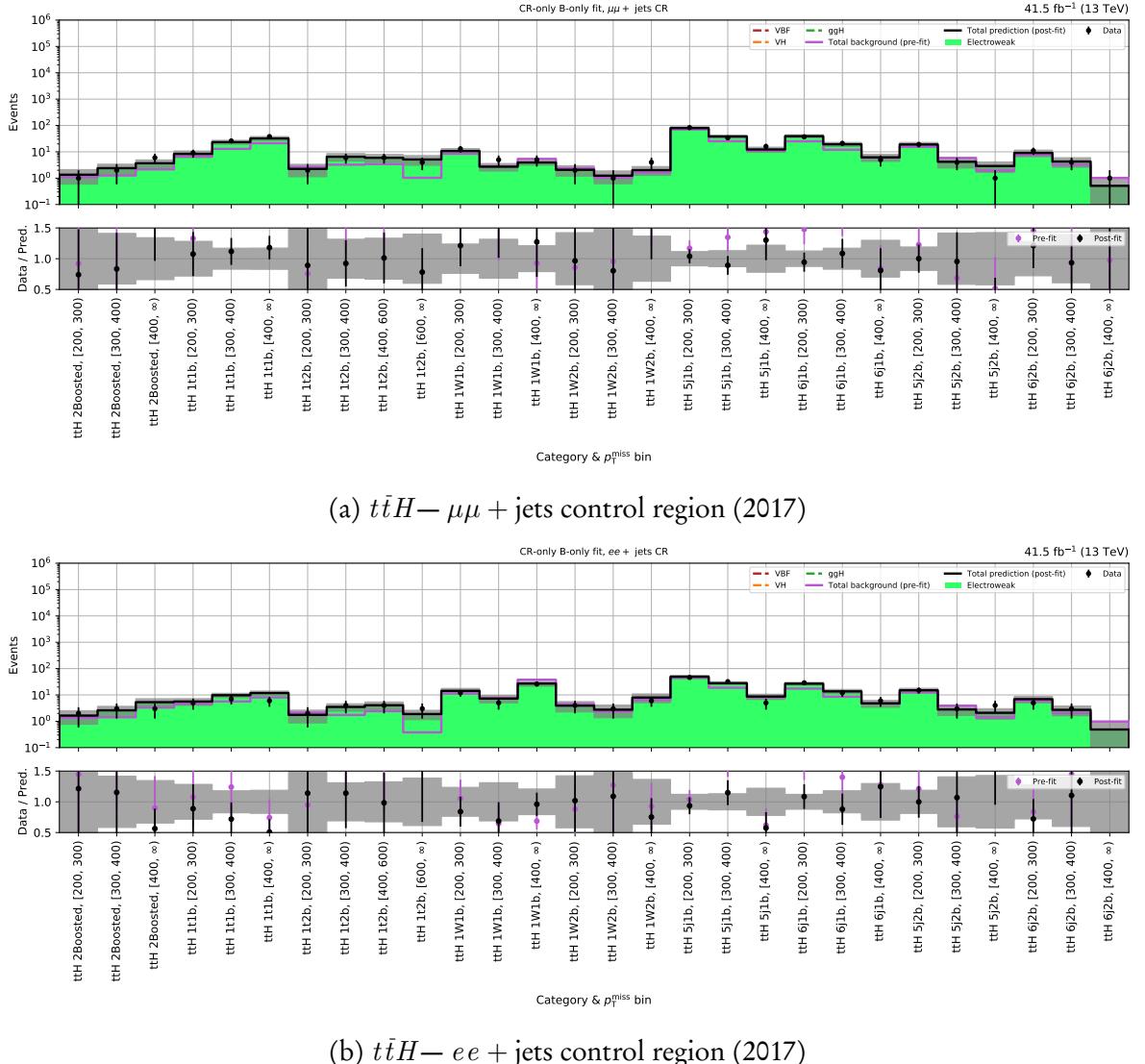


Figure 6.10: Post-fit yields for each category and  $p_T^{\text{miss}}$  bin in the dilepton control regions of the  $t\bar{t}H$  categories for the 2017 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subfigure.

## 6.7. BACKGROUND ESTIMATION

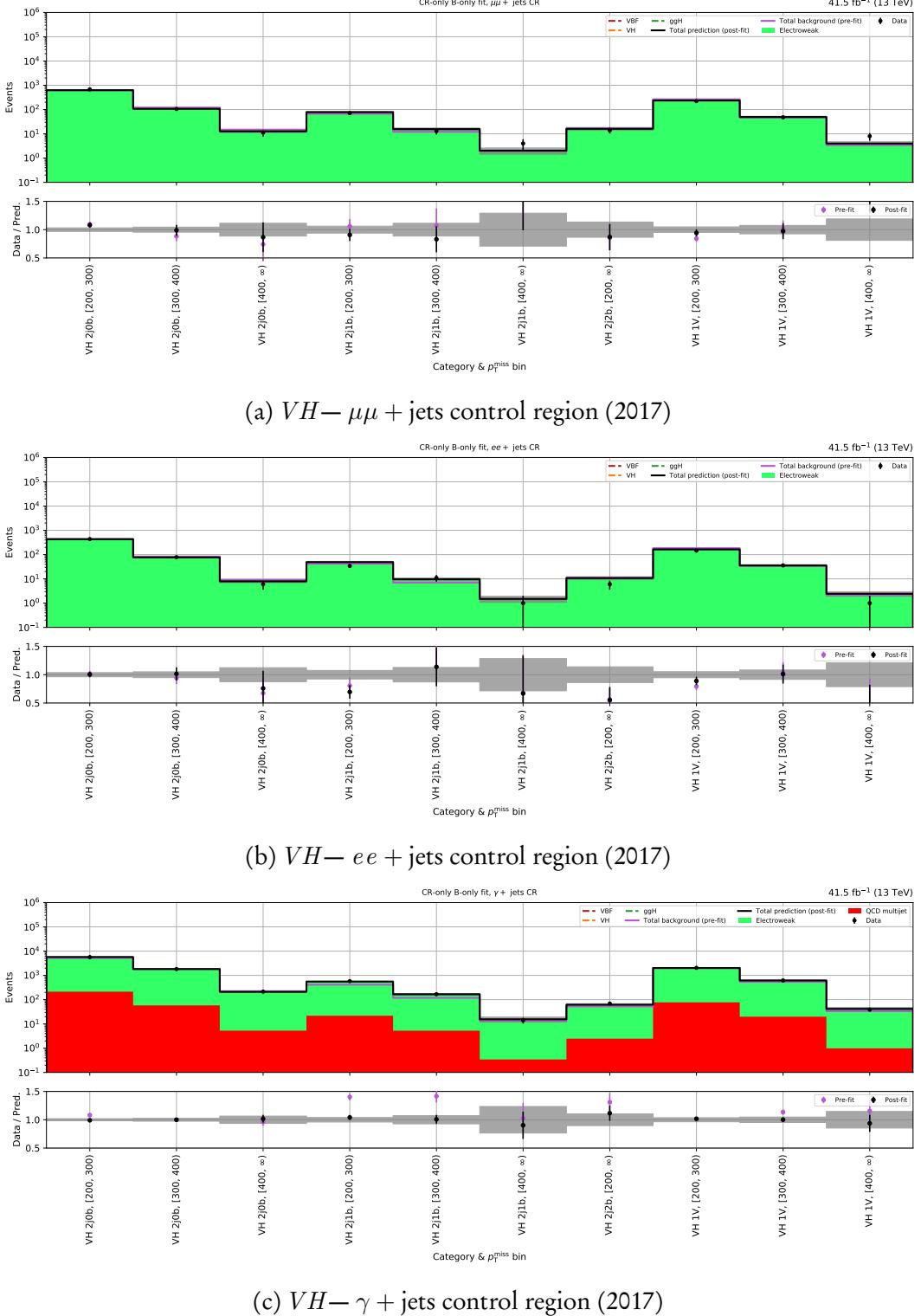


Figure 6.11: Post-fit yields for each category and  $p_T^{\text{miss}}$  bin in the dilepton and photon control regions of the  $VH$  categories for the 2017 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subfigure.

Despite the low yields in the  $t\bar{t}H$  categories, the fit is able to correct the simulated background closer to the observed data in most of the bins. A comparable result is achieved in the  $VH$  categories. The highly populated  $\gamma + \text{jets}$  control region adds a powerful additional constraint to the prediction; an extremely useful inclusion given  $Z \rightarrow \nu\bar{\nu}$  is comfortably the dominant background for the  $VH$  channel. With QCD contributing under 5 % to the total yield in the photon region, convergence of the fit is naturally driven by the electroweak processes.

### 6.7.4 QCD multijet background estimation

The effects of jet mismeasurements are difficult to quantify. With a final state of several jets in the QCD multijet process, little, or even no  $\vec{p}_T^{\text{miss}}$  is expected. Therefore, a single mismeasured jet will introduce artificial  $\vec{p}_T^{\text{miss}}$  in the direction of that jet. A low  $\Delta\phi_{\min}(j, \vec{p}_T^{\text{miss}})$  is therefore expected. Though it is not just this background that suffers as jets from “cleaner” processes may also be affected. But those with real  $p_T^{\text{miss}}$  in an event (e.g.,  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ ) are unlikely to be significantly influenced by one stray object. The enormous cross section of QCD multijet amplifies the problem, making the process as a whole more sensitive to, e.g., fluctuations in the calorimeter response that would affect the energy measurement.

Contributions to the signal region from QCD multijet events should be adequately suppressed by the analysis-level selection requirements. However, it is still a process that must be accurately accounted for considering its rate of production at CMS. A metric by which to estimate the number of events a dataset should require is by calculating the equivalent luminosity:

$$(6.13) \quad \mathcal{L}_{\text{eq.}} = \frac{N_{\text{events}}}{\sigma}$$

A general rule is that the equivalent luminosity of a given dataset should be comparable to, or even exceed, that of the data collected by the experiment. Since the QCD multijet process has a very large cross section, simulating the required number of events to match the luminosity of the data recorded during Run-2 is not feasible.

To estimate the presence of QCD in the signal region, a data driven approach is taken utilising the sidebands defined in Chpt. 6.7.4.1. Then, the application of the sidebands to derive the predicted event counts in the signal region is outlined in Chpt. 6.7.4.2. This is derived separately for each channel and data taking year.

#### 6.7.4.1 Sidebands to the signal region

Several kinematic cuts in the analysis are designed to reject QCD in the signal region. So by inverting and tightening one or more of them, it is possible to construct multijet-enriched regions known as *sidebands*. The sideband is defined for the  $t\bar{t}H$  channel by inversions on the angular variables  $\tilde{\omega}_{\min}$  and  $\Delta\phi_{\min}(j_{1,2,3,4}, \vec{p}_T^{\text{miss}})$ :

- $t\bar{t}H$  sideband:  $\tilde{\omega}_{\min} < 0.2, \Delta\phi_{\min}(j_{1,2,3,4}, \vec{p}_T^{\text{miss}}) \leq 0.5$

The signal region is defined by  $\tilde{\omega}_{\min} > 0.3$  and  $\Delta\phi_{\min}(j_{1,2,3,4}, \vec{p}_T^{\text{miss}}) > 0.5$  as in Tab. 6.1. Otherwise, the same selection applied to the signal region is also done so for the sidebands. Yields from all of the  $t\bar{t}H$  categories are combined to form the sidebands as statistical limitations would otherwise ensue.

Events with a  $VH$  topology, being characterised by a dijet pair or single fat jet recoiling from the  $\vec{p}_T^{\text{miss}}$ , naturally have a large  $\Delta\phi_{\min}(j, \vec{p}_T^{\text{miss}})$ . Obtaining a multijet enriched sideband by inverting this variable was therefore not possible. The sidebands inverted in  $\tilde{\omega}_{\min}$  also contained little QCD to extrapolate from. Therefore, a single sideband was created from the 2j0b category that inverts  $\Delta\phi_{\min}, \tilde{\omega}_{\min}$  (tightly, as with  $t\bar{t}H$ ), and the dijet mass requirement from Tab. 6.1:

- $VH$  sideband:  $\tilde{\omega}_{\min} < 0.2, \Delta\phi_{\min}(j_{1,2,3,4}, \vec{p}_T^{\text{miss}}) \leq 0.5, m_{jj} \notin [65, 105]$

The latter inversion populates the sideband sufficiently with multijet events. This sideband is utilised by all  $VH$  categories, as while the 1V category does not enforce the same dijet mass window, the soft drop mass requirement for the boosted  $V$ -tagged jet is a suitable proxy. Inclusion of  $b$ -jets in the 2j1b and 2j2b categories reduces the multijet content of potential sidebands significantly. This is the reason these categories also utilise the sideband constructed from 2j0b.

Fig. 6.12 illustrates the  $p_T^{\text{miss}}$  distributions in the  $t\bar{t}H$  and  $VH$  sidebands after the analysis-level selections with the full Run-2 dataset. Background estimation is performed separately for each year due to the differing running conditions, detector configuration, and the different effects or features seen in the data.

#### 6.7.4.2 Prediction in the signal region

A control region-only fit is performed without multijet simulation to extract the rate parameters ( $a$  from Eq. 6.11) that scale the electroweak backgrounds in the signal region. Applying these also to the corresponding backgrounds in the sidebands changes the event distribution and hence the data-MC agreement. The excess in data in a sideband is assumed to arise

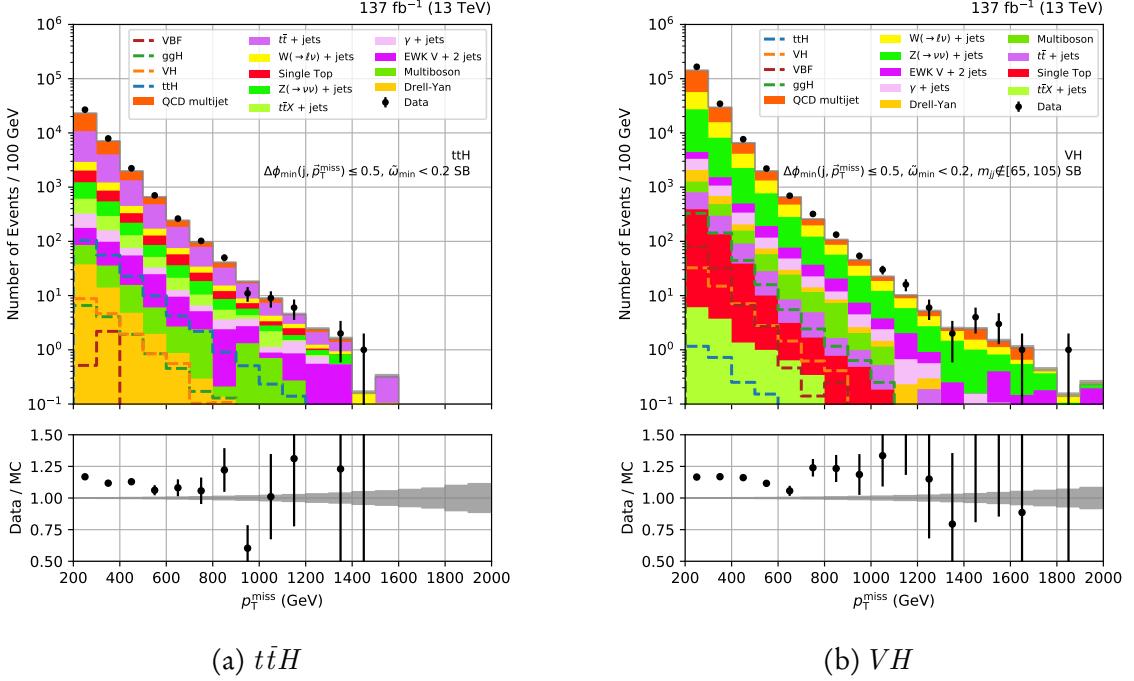


Figure 6.12: Data–simulation comparisons of the  $p_T^{\text{miss}}$  distribution in the  $t\bar{t}H$  and  $VH$  sidebands, aggregated over the full Run-2 dataset.

solely from multijet events, and as such the difference between data and the non-multijet background is attributed to QCD (denoted as  $N_{\text{SB}}^{\text{QCD}}$ ).

QCD in the signal region  $N_{\text{pred.}}^{\text{QCD}}$  (corresponding to  $c_{\text{QCD}}$  in Eq. 6.10) is predicted in each category and  $p_T^{\text{miss}}$  bin as follows:

$$(6.14) \quad N_{\text{pred.}}^{\text{QCD}}(\text{category}, p_T^{\text{miss}}) = N_{\text{SB}}^{\text{QCD}} \cdot \mathcal{T}_{\text{QCD}} \cdot f_c(\text{category}) \cdot f_p(p_T^{\text{miss}})$$

The distribution of the QCD background for each category and  $p_T^{\text{miss}}$  bin is extrapolated from the sidebands with the factors  $f_c$  and  $f_p$ .  $f_c$  is the fraction of non-QCD MC in a given category of the signal region, inclusive of  $p_T^{\text{miss}}$ . While  $f_p$  is the fraction of non-QCD MC in a given  $p_T^{\text{miss}}$  bin of the signal region, inclusive of category. Determining  $f_c$  and  $f_p$  from non-multijet simulation assumes the QCD background is distributed in the same proportion across categories as the the non-QCD background, though gives a more consistent description of the data in the sidebands than estimating the fractions from QCD MC. A 50 % systematic uncertainty following a log normal distribution is assigned to the predicted multijet event counts. Numerical values for the terms in Eq. 6.14 for deriving the QCD prediction are visible in App. A.4.

The term  $\mathcal{T}_{\text{QCD}}$  is the transfer factor relating QCD in the sideband to that in the signal

region. For  $t\bar{t}H$ , the transfer factor is inclusive over the categories, i.e.,

$$(6.15) \quad \mathcal{T}_{\text{QCD}} = \frac{N_{\text{MC, SR}}^{\text{QCD}}}{N_{\text{MC, SB}}^{\text{QCD}}}$$

To attempt to mitigate the statistical limitations of the QCD MC nominally used in the analysis, their event count was increased a hundredfold using a “smear and rebalance” method as performed in Ref. 138. These smeared samples replaced the QCD MC only for the derivation of  $\mathcal{T}_{\text{QCD}}$ . The  $VH$  categories have the benefit of a much larger sample size, both before and after smearing the MC. In this case, the transfer factor is derived bin-by-bin.

For both the  $t\bar{t}H$  and  $VH$  channels,  $N_{\text{SB}}^{\text{QCD}}$  and  $\mathcal{T}_{\text{QCD}}$  were derived from their respective sidebands. In  $t\bar{t}H$ ,  $f_c$  was determined from non-multijet simulation in the signal region. No category fraction was introduced in  $VH$  since all of those categories predict the QCD background from the inversion of the 2j0b category.

## 6.8 Corrections and systematic uncertainties for simulation

In order for simulated events, particularly SM background processes, to resemble LHC data as closely as possible, many corrections are applied. These are discussed in more detail in the subsequent sections. All corrections and associated systematic uncertainties are applied to all samples for all data taking years, unless stated otherwise, and summarised in Tab. 6.8.

A final event weight  $w_{\text{event}}$  is the product of the weights from all of the individual sources  $i$  that provide one:

$$(6.16) \quad w_{\text{event}} = \prod_i w_i$$

When representing these events in histograms, the yield in a given bin  $N_{\text{corr.}}$  is the sum of these event weights:

$$(6.17) \quad N_{\text{corr.}} = \sum_j^{N_{\text{MC}}} w_{\text{event}, j}$$

where  $N_{\text{MC}}$  is the number of unweighted, simulated events in the bin. In the data–MC distributions shown previously in this chapter, the statistical uncertainty ascribed to the MC yield in a bin is estimated from the unweighted number of events ( $k$  in a Poisson statistical treatment):

$$(6.18) \quad \Delta N_{\text{corr.}} = \pm \frac{N_{\text{corr.}}}{\sqrt{N_{\text{MC}}}}$$

While the statistical uncertainty for the number of events in data is simply the Poissonian error,

$$(6.19) \quad \Delta N_{\text{data}} = \pm \frac{N_{\text{data}}}{\sqrt{N_{\text{data}}}}$$

### 6.8.1 Efficiency of the triggers

Given the modelling of the Level-1 and High-Level Trigger systems in simulation is only an approximation of their behaviour with data, corrections must be made to unite them.

A dedicated control region is formed to assure orthogonality with the signal region and avoid unblinding of data, but with a similar kinematic selection to the analysis for accuracy in the desired phase space. Events are selected in data from the primary dataset consisting of muon-based triggers, and in simulation from the  $W(\rightarrow \ell\nu) + \text{jets}$  process at LO. Analysed events consist of those containing a single, tightly-isolated muon with  $p_T > 30 \text{ GeV}$  that triggers the HLT path for an isolated muon with  $p_T > 27 \text{ GeV}$ . This is to avoid biasing the efficiencies of the triggers defining the signal region. An offline selection is then applied to keep those with similar kinematic properties to those in the signal region:

- $p_T^{\mu} > 80 \text{ GeV}$
- $H_T > 200 \text{ GeV}$
- $H_T^{\text{miss}} / p_{T,\mu}^{\text{miss}} < 1.2$

where  $p_{T,\mu}^{\text{miss}} = |\vec{p}_T^{\text{miss}} + \vec{p}_T^{\mu}|$  to approximate the PF calculation at HLT level. Of these, events that additionally pass the online selection of the trigger requirements in the signal region (Tab. 6.3) are also recorded. The efficiency of a trigger—or collection of triggers— $\epsilon_{\text{trg}}$  is defined as

$$(6.20) \quad \epsilon_{\text{trg}} = \frac{s_{\text{offline}} \cap s_{\text{online}}}{s_{\text{offline}}}$$

where  $s$  denotes the set of events that have passed the selection given by its subscript. For simulation, events are weighted by cross section as per Eq. 6.1. The efficiencies are binned in two dimensions as a function of  $p_{T,\mu}^{\text{miss}}$  and  $H_T^{\text{miss}}$ . The weight applied to events in simulation  $w_{\text{trg}}$  is then simply the efficiency in data divided by the efficiency in simulation for the values of  $p_{T,\mu}^{\text{miss}}$  and  $H_T^{\text{miss}}$  in the event:

$$(6.21) \quad w_{\text{trg}}(p_T^{\text{miss}}, H_T^{\text{miss}}) = \frac{\epsilon_{\text{trg, data}}(p_T^{\text{miss}}, H_T^{\text{miss}})}{\epsilon_{\text{trg, MC}}(p_T^{\text{miss}}, H_T^{\text{miss}})}$$

Uncertainties are estimated using the Clopper-Pearson method [48] and impact the yields by up to 1–2 %. In the signal region, sidebands, and muon control regions, weights from

the efficiencies of the  $p_T^{\text{miss}} - H_T^{\text{miss}}$  cross triggers are used. The weights are provided for each data taking year in Fig. 6.13. In the electron and photon control regions, the efficiencies are calculated in the same manner, instead from the triggers in Tabs. 6.4 and 6.5.

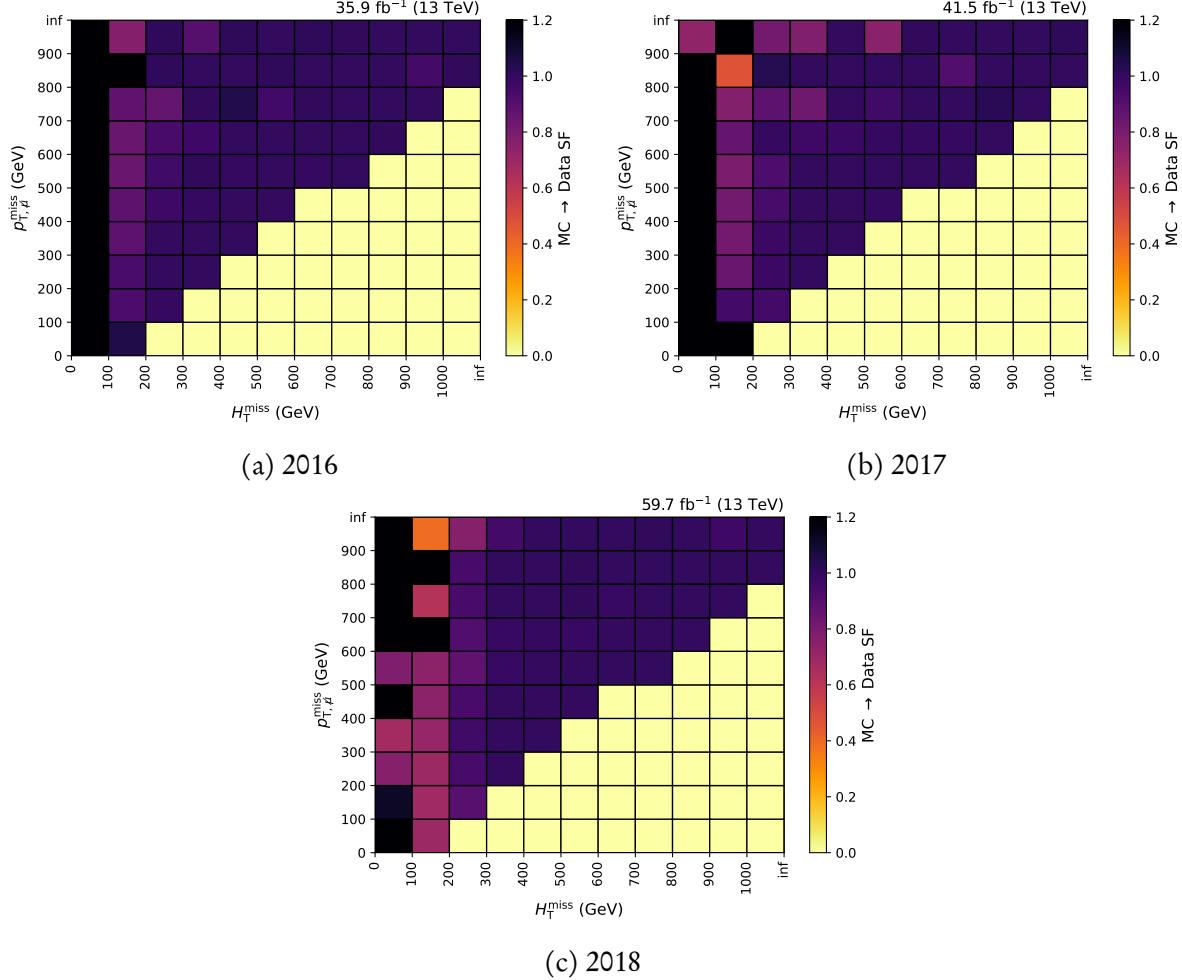


Figure 6.13: Scale factors accounting for the efficiencies of the HLT  $p_T^{\text{miss}} - H_T^{\text{miss}}$  cross triggers in each data taking year of Run-2. The momentum of the muon used to select events that compute the efficiency is added to the  $\vec{p}_T^{\text{miss}}$  (with a magnitude  $p_{T,\mu}^{\text{miss}}$ ) in order to approximate the calculation in the High-Level Trigger.

## 6.8.2 Pileup reweighting

Pileup interactions at the LHC are frequent (see Chpt. 3.1.3) and must be modelled appropriately in simulation. Simulated samples are generated with a certain distribution of the number of pileup interactions which usually does not match the data recorded by CMS. This is due to changing conditions in the beam over a period of data taking. In order to

make them comparable, the simulated events are reweighted; in this context it is known as *pileup reweighting*. ROOT files containing histograms of the number of pileup interactions from short runs in the LHC are available centrally and are used as the reference for which to reweight the simulated events.

Simulated events are nominally reweighted according to data, where the inelastic  $pp$  cross section is measured to be 69.2 mb. An uncertainty of  $\pm 4.6\%$  in the measurement is used to calculate the systematic uncertainty on this weight. These were found to be a 1–2 % effect in the signal and control regions. The pileup distributions, and therefore the weights, are different for each data taking year. However, the inelastic cross section and associated uncertainty are consistent for the entirety of Run-2.

### 6.8.3 Higher order corrections to $V + \text{jets}$ samples

Vector boson + jets processes present a sizeable background in the signal region for some of the categories, and are a large component of some of the control regions. At the dataset sizes required to provide sufficient statistical coverage of the analysis phase space, NLO MC samples are not available with the complete detector simulation applied. To circumvent this, LO samples are used primarily and reweighted at generator level on an event-by-event basis to NLO accuracy. This method is usually referred to as a “ $k$ -factor” correction, where  $k$  is the ratio of the NLO to LO cross section.

Separate NLO QCD and electroweak corrections are applied to the QCD production of  $V + \text{jets}$  samples. NLO  $W + \text{jets}$  samples are utilised when deriving the  $k$ -factors for the LO counterpart. Drell-Yan  $Z(\rightarrow \ell\bar{\ell}) + \text{jets}$  samples are utilised for both the LO Drell-Yan and  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$  processes. The  $k$ -factors are calculated independently for the two processes as the LO sample production and cross sections are different. One or two additional jets in the matrix element calculations are permitted in the NLO samples. Events from the above datasets are processed through the following generator-level event selection to ensure they are in a similar phase space to the analysis:

- $p_T^{j_1} > 80 \text{ GeV}$
- $p_T^V > 200 \text{ GeV}$
- $H_T^{\text{Gen}} > 200 \text{ GeV}$
- $H_T^{\text{Gen}}/p_T^V < 1.2$

where  $p_T^V$  is the generator-level boson  $p_T$ , and  $H_T^{\text{Gen}}$  is the scalar sum of generator-level jet  $p_T$ . The  $k$ -factors are binned in two dimensions,  $p_T^V$  and  $p_T^{j_1}$ .

Uncertainties for the QCD renormalisation scale, factorisation scale, and in the parton distribution function are treated as individual, uncorrelated sources. The same values for the nominal  $k$ -factors and uncertainties are used for each data taking year in Run-2. The impact of the renormalisation scale uncertainty was approximately 0–5 % on the yields, while the factorisation and PDF scales had effects of roughly 1–10 % and 1–5 %, respectively. Distributions of the NLO QCD  $k$ -factors are presented in Fig. 6.14.

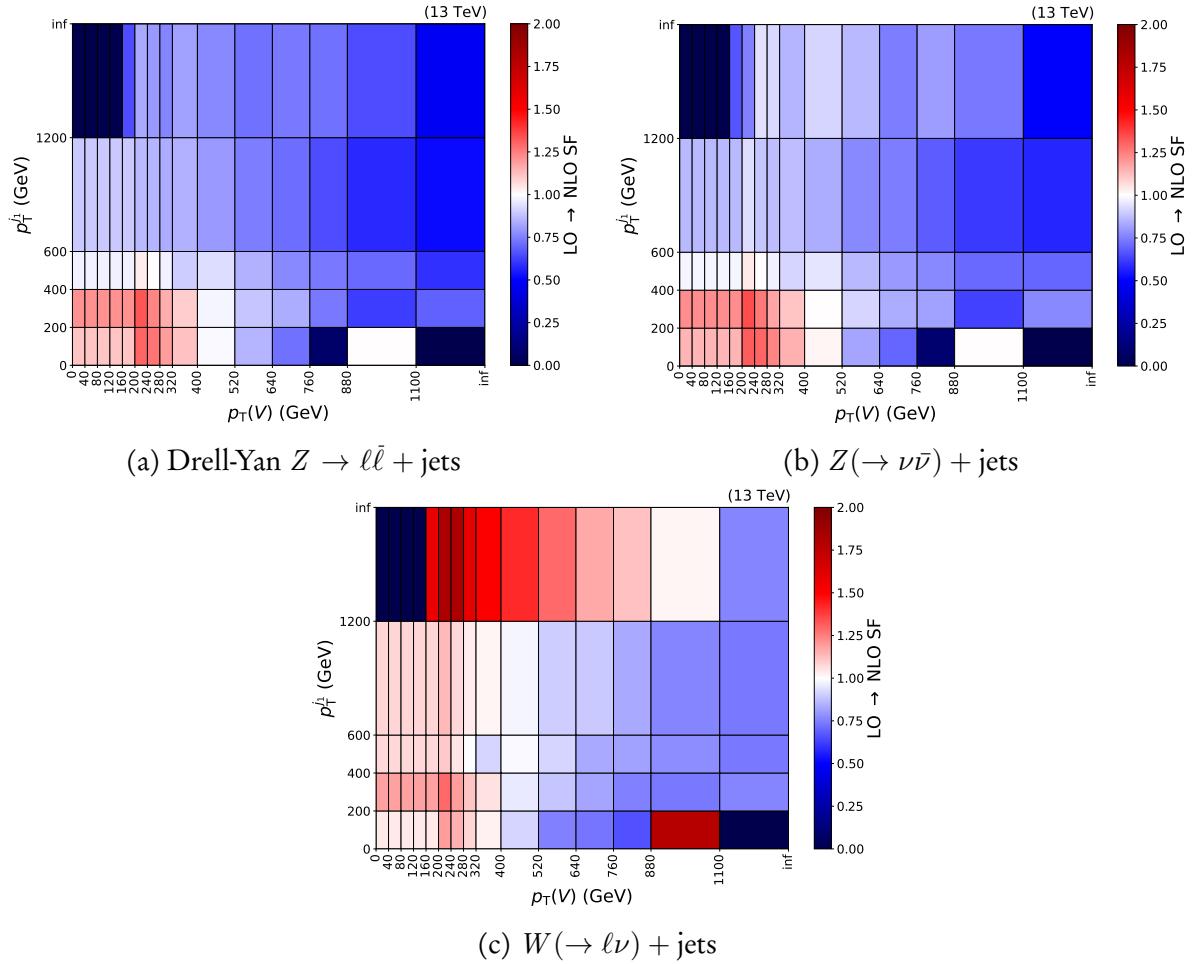


Figure 6.14: NLO  $k$ -factors—as a function of the generator-level boson  $p_T$  and leading jet  $p_T$ —for reweighting reweight events in LO QCD  $W$  and  $Z + \text{jets}$  simulation to NLO accuracy in QCD.

Since all of the  $V + \text{jets}$  samples were generated at LO electroweak accuracy, an additional  $k$ -factor is needed to reweight events to NLO. Ref. 112 provides NLO electroweak corrections for  $W$ ,  $Z$ , and  $\gamma + \text{jets}$  processes as a function of  $p_T^V$ , depicted in Fig. 6.15. The  $Z$   $k$ -factor is adopted by both the Drell-Yan and  $Z \rightarrow \nu\bar{\nu}$  processes.

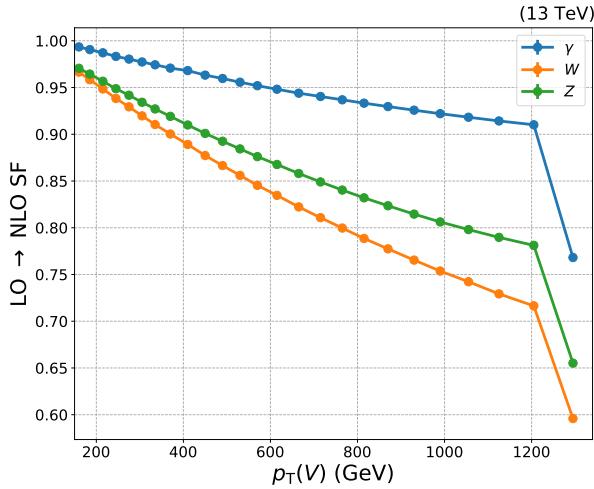


Figure 6.15: NLO  $k$ -factors as a function of generator-level boson  $p_T$  for reweighting QCD  $V + \text{jets}$  events generated at NLO electroweak accuracy. Separate  $k$ -factors correct the  $W$ ,  $Z$ , and  $\gamma + \text{jets}$  processes. Values obtained from Ref. 112.

### 6.8.4 Top quark processes

$t\bar{t} + \text{jets}$  is the dominant background for the  $t\bar{t}H$  channel. The  $t\bar{t}X$  backgrounds also have a non-negligible presence. As such, much effort must be made to ensure they are modelled correctly. Variations of the QCD renormalisation and factorisation scales provide an uncertainty only to the shape of the  $p_T^{\text{miss}}$  spectrum of the  $t\bar{t}$  and  $t\bar{t}X$  backgrounds, and  $t\bar{t}H$  signal. Reweighting of the top quark  $p_T$  distribution to NNLO accuracy is also performed for  $t\bar{t}$ , yielding a shape correction and associated uncertainty.

#### 6.8.4.1 QCD scale systematic uncertainty

The renormalisation  $\mu_R$  and factorisation scale  $\mu_F$  chosen for MC samples at generation may be somewhat arbitrary, and therefore not representative of nature. Relative to the nominal value, weights were derived centrally for each combination of these scales being varied independently (and together) up or down by a factor of two. Eight variations then exist for each event:

- $\mu_{R\downarrow} \& \mu_{F\downarrow}$
- $\mu_{R\downarrow} \& \mu_{F\text{ nom.}}$
- $\mu_{R\downarrow} \& \mu_{F\uparrow}$
- $\mu_{R\text{ nom.}} \& \mu_{F\downarrow}$

- $\mu_{R,\text{nom.}} \& \mu_{F\uparrow}$
- $\mu_{R\uparrow} \& \mu_{F\downarrow}$
- $\mu_{R\uparrow} \& \mu_{F,\text{nom.}}$
- $\mu_{R\uparrow} \& \mu_{F\uparrow}$

To remove the dependence on normalisation (since the skim and event selection may bias it) and ensure the systematic is only shape-based, each variation was also multiplied by an additional factor  $f_{\text{var}}$ :

$$(6.22) \quad f_{\text{var}} = \frac{\sum_{\text{unskimmed events}} w_{\mu_{R,\text{nom.}} \& \mu_{F,\text{nom.}}}}{\sum_{\text{unskimmed events}} w_{\text{var}}}$$

These are computed individually for each decay and data taking year. The envelope is characterised by the correlated variations  $\mu_{R\uparrow} \& \mu_{F\uparrow}$ , and  $\mu_{R\downarrow} \& \mu_{F\downarrow}$ . Confirmed by Fig. 6.16, an example is given for the  $t\bar{t}$  background in the signal region. While the sizes of the variations appear large, they are reduced when matching the fit binning, and very similar between the signal and control regions, cancelling to a large degree to minimally impact the results.

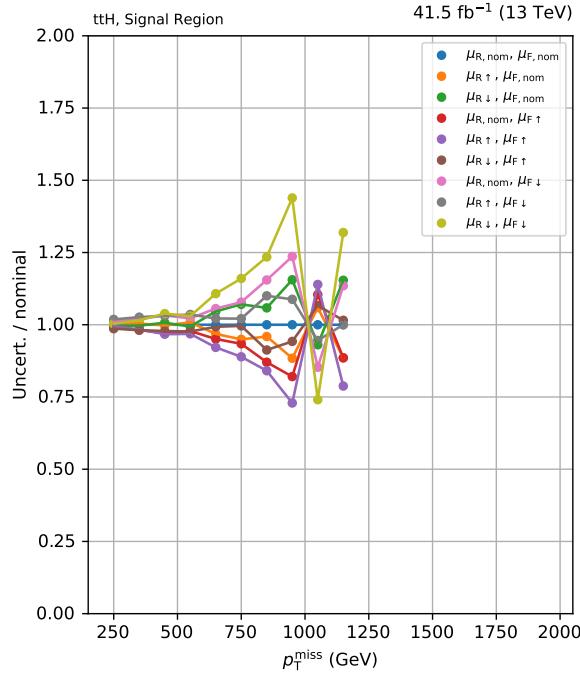


Figure 6.16: The deviations relative to the nominal weight of the combinations of QCD renormalisation and factorisation scale variations. They are presented as a function of  $p_T^{\text{miss}}$  inclusive over the  $t\bar{t}H$  categories for the  $t\bar{t}$  samples. Events are from the signal region in the 2017 dataset after the analysis-level selection.

For each process (i.e.,  $t\bar{t}$ , and the accompanying boson  $X$  in  $t\bar{t}X$ ), a separate systematic uncertainty is derived.<sup>5</sup> Since different generators and settings may be used for each dataset, it is appropriate to assume they are uncorrelated.

#### 6.8.4.2 Top quark $p_{\mathrm{T}}$ reweighting

The top quark  $p_{\mathrm{T}}$  distribution has been shown to be harder in simulation at NLO than in data [136]. NLO  $t\bar{t} + \text{jets}$  MC generated in POWHEG is reweighted to NNLO QCD + NLO electroweak accuracy as function of top quark  $p_{\mathrm{T}}$ . An analytic function fits the distribution that determines the weight. The same nominal function applies to all  $t\bar{t} + \text{jets}$  samples for all data taking years:

$$(6.23) \quad f(p_{\mathrm{T}}) = \exp(a + b \cdot p_{\mathrm{T}} + c \cdot p_{\mathrm{T}}^2)$$

where the coefficients are  $a = 1.614 \times 10^{-2}$ ,  $b = -1.966 \times 10^{-4}$ , and  $c = -1.454 \times 10^{-8}$ . For a top or antitop quark, the scale factor is given by the value of  $f$  for the particle's generator-level  $p_{\mathrm{T}}$ . The event weight is then the square root of the product of the top and antitop scale factors.

The uncertainties are acquired from the fit itself over the theoretical uncertainties inherent in MC calculations. This was to avoid double counting with respect to the renormalisation and factorisation scale uncertainty in Chpt. 6.8.4.1, as overlap would be present and otherwise difficult to disentangle. For each parameter in the fit, the upward and downward variations were estimated. The fit function was then recalculated, fixing the parameter of interest to the variation. Providing upward and downward functions, the systematic uncertainty attributed to the parameter is simply the event weight calculated with these functions over the nominal. Separate, uncorrelated uncertainties were estimated for each parameter in the fit, and were each found to have a  $< 2\%$  impact on the event yields. The dependence on normalisation was removed in the same fashion as for the QCD scale to consider the correction only to the shape of the distributions and not also the normalisation. The nominal function and associated variations are shown in Fig. 6.17.

---

<sup>5</sup>  $t\bar{t}H(H \rightarrow \text{invisible})$  signal and  $t\bar{t}H(H \rightarrow \text{visible})$  background are grouped as one process since the kinematics and generation procedure are consistent.

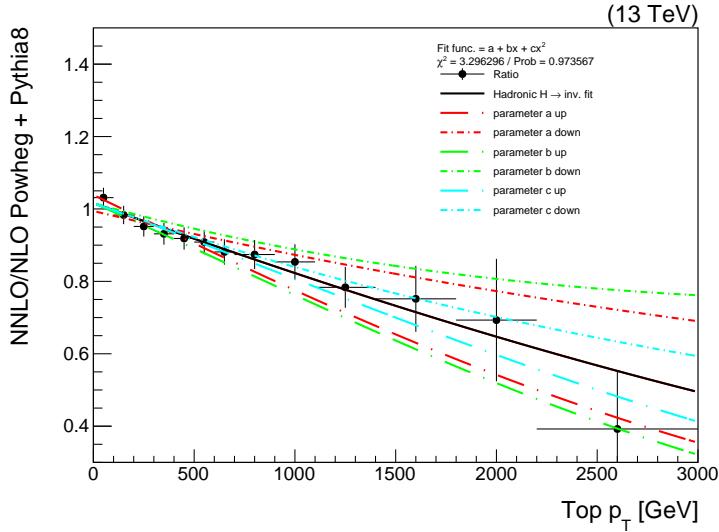


Figure 6.17: The ratio of the NNLO to NLO  $t\bar{t}$  cross section as a function of top quark  $p_T$ . The fit to the distribution—along with the uncertainties in the parameters—are shown.

### 6.8.5 Object-level scale factors and uncertainties

There are multiple corrections applied to simulation based on the physics objects with an event as opposed to the event topology as a whole. One example of which is Chpt. 6.8.4.2. A weight or efficiency  $\epsilon$  attributed to an object itself is typically referred to as a *scale factor*, while the event-level selection weight  $w_{\text{sel.}}$  is the efficiency for successfully tagging or selecting an event with those object contents, and is the product of the scale factors:

$$(6.24) \quad w_{\text{sel.}} = \prod_i^{N_{\text{objects}}} \epsilon_i$$

These weights are calculated individually for each class of object in an event. For electrons, muons, and photons, these scale factors are usually from the reconstruction efficiency, identification efficiency, and  $p_T$ - or  $\eta$ -dependent energy corrections. In the case of  $b$ -tagged jets, it is the data–simulation scale factor at the given working point from the algorithm used to identify them. Systematic uncertainties associated with the scale factors are aggregated in the same manner. Unless stated otherwise, the following scale factors and systematics are derived separately for each year in Run-2.

#### 6.8.5.1 Lepton and photon identification, isolation, and reconstruction

The scale factors and systematic uncertainties for electrons, photons, and muons were derived centrally for CMS by the Physics Object Groups. Electrons and photons are identified using

the cut-based identification described in Chpts. 4.2.4 and 4.2.5, respectively. Scale factors are applied separately to each of these objects in simulation that account for the efficiency of identification, isolation, and reconstruction in data. They are derived as a function of  $\eta$  and ECAL supercluster  $p_T$ . Uncertainties are given as the error in the  $(\eta, p_T)$  bin of the histogram that describes the efficiencies. The ID and isolation are grouped as a single uncertainty, while reconstruction is another. Event weights simply are the products of the individual object scale factors. The uncertainty on the event weight is the product of the scale factor plus/minus its uncertainty. For each source, they are on the order of 1–5 %.

Muons follow a similar prescription, where the ID and isolation efficiencies are calculated in bins of  $\eta$  and  $p_T$ . The systematic uncertainties for each of the two sources are separated—contrary to electrons and photons—and are usually 1 % or smaller in size. An equivalent reconstruction efficiency is not given as the dedicated muon chambers and reconstruction algorithms are more than sufficient.

#### 6.8.5.2 $b$ -jets tagged by the DEEPCSV algorithm

Data–simulation scale factors are calculated for each  $b$ -jet as tagged by the DEEPCSV algorithm. They are defined as the tagging efficiency in data divided by the efficiency for simulation for a given working point, and as a function of jet  $p_T$ . Various methods are employed to derive separate scale factors for separate topologies, such as QCD multijet or  $t\bar{t}$  events, and are described in Sec. 8.4 of Ref. 132. A weighted average of the scale factors computed by each method are then taken to yield the final scale factor for a  $b$ -tagged jet. The resultant systematic uncertainty is typically less than 5 %.

#### 6.8.5.3 Boosted jets tagged by the DEEPAK8 algorithm

In a similar fashion to the section above, objects classified by the DEEPAK8 algorithm are given scale factors according to the data/MC efficiencies as a function of AK8 jet  $p_T$ . Scale factors are derived separately for jets tagged as originating from a top quark and  $V$  boson. Systematic variations associated with the scale factors on the order of 1–10 % are also derived separately, according to Sec. 8.1 of Ref. 60.

### 6.8.6 Minor contributions

In the following subsection, minor corrections and uncertainties which had little impact on the yields are documented.

### 6.8.6.1 Pre-firing in the ECAL

A weight is applied to events in simulation that emulates the effect of pre-firing in 2016 and 2017 data, given in Ref. 64:

$$(6.25) \quad w_{\text{pre-firing}} = \prod_{i=\gamma, \text{jet}} 1 - \epsilon_i^{\text{pre-firing}}(\eta, p_T^{\text{EM}})$$

where  $\epsilon$  is the probability for an object to pre-fire as a function of its  $\eta$  and electromagnetic  $p_T$  (simply  $p_T$  for photons, and  $p_T$  multiplied by the sum of the charged and neutral electromagnetic energy fractions for jets). Pre-firing probabilities are assumed to be uncorrelated between photons and jets. The systematic uncertainty in the event weight is found by taking the uncertainty on  $\epsilon$ , defined the maximum value from either  $0.2\epsilon$  or the relative statistical uncertainty in the  $(\eta, p_T^{\text{EM}})$  bin of the pre-firing map from which  $\epsilon$  is extracted. It was found to impact the yields in the analysis by  $< 1\%$ .

### 6.8.6.2 Jet energy scale and resolution

To estimate the magnitude of the jet energy scale (JES) and resolution (JER) uncertainties, each event is copied once per variation (i.e., upward and downward for the JES and JER). In the copied event, the  $p_T$  of the AK4 jets are varied according to scale factors from resolution/scale maps as a function of  $p_T$  and  $\eta$ . Consequently, the  $\vec{p}_T^{\text{miss}}$  also changes for the event. The copied events are propagated through the analysis with the same treatment as the standard events. After categorisation and binning in  $p_T^{\text{miss}}$ , the weighted event yields from the copy in a given subcategory and  $p_T^{\text{miss}}$  bin serves as the uncertainty on the nominal yields.

As the event selection for the signal region is mirrored in the control regions, variations from the JER and JES variations should affect both in a similar way and their effects should largely cancel in the fit to data. This cancellation is however not complete, resulting in residual differences of 2 % and 3 % on average for JER and JES, respectively. Similar variations for the signal processes were observed. As such, these values are applied as systematic uncertainties to all signal and background processes in all regions of the analysis.

### 6.8.6.3 Luminosity

Uncertainties are associated with the various methods designed to measure the luminosity received by CMS. For 2016, 2017, and 2018, the total systematic uncertainties on certified data are 2.5 %, 2.3 %, and 2.5 %, respectively. They may be decomposed into several individual sources, some of which are correlated between years. These are documented in Tab. 6.6 with further information available in Refs. 50, 51, and 52.

Uncertainty source	Impact in 2016	Impact in 2017	Impact in 2018
Uncorrelated sources	2.2	2.0	1.5
$x\text{-}y$ factorisation	0.9	0.8	2.0
Length scale	—	0.3	0.2
Beam-beam deflection	0.4	0.4	—
Dynamic $\beta$	0.5	0.5	—
Beam current calibration	—	0.3	0.2
Ghosts and satellites	0.4	0.1	—

Table 6.6: Sources of uncertainty in the luminosity measurements for each year of Run-2, and their impacts in percent.

#### 6.8.6.4 Cross section for signal processes

There are several sources of uncertainty in the computation of high order cross sections, from both theoretical and experimental sides. As with the nominal cross sections for signal processes, the uncertainties are retrieved from Ref. 43. Each source for each signal process is treated as a separate uncertainty on the normalisation. They are correlated across years as the cross sections are constant. Nominal values of the cross sections at  $\sqrt{s} = 13$  TeV are tabulated in Tab. 2.3. A summary of the uncertainties are presented in Tab. 6.7.

Signal	$\delta(\text{scale}) (\%)$	$\delta(\text{PDF} + \alpha_S) (\%)$
$t\bar{t}H$	+6.0, -9.2	$\pm 3.5$
$W^+H$	+0.74, -0.73	$\pm 1.79$
$W^-H$	+0.59, -0.63	$\pm 2.03$
$pp \rightarrow ZH$	+3.50, -2.68	$\pm 1.65$
$gg \rightarrow ZH$	+2.49, -1.88	$\pm 4.37$
VBF	+0.43, -0.33	$\pm 2.1$

Table 6.7: Systematic uncertainties on the cross section of the simulated  $H \rightarrow$  invisible signal samples. Values obtained from Ref. 43.

#### 6.8.7 Summary of systematic uncertainties

Tab. 6.8 contains a summary of the systematic uncertainties incorporated in the fits to data. The type of uncertainty, whether it affects primarily the normalisation of the event yields and/or the shapes of distributions, the applicable data taking years and correlations among them, the samples affected (where “all” corresponds to all signal and background simulation), and an estimation of the impact on event yields in the  $p_T^{\text{miss}}$  distribution are presented.

Source	Type	Years	Samples	Size (%)
Pileup	Normalisation, shape	All, correlated	LO V + jets	1-2
NLO renormalisation scale	Normalisation, shape	All, correlated	LO V + jets	0-5
NLO factorisation scale	Normalisation, shape	All, correlated	LO V + jets	1-10
NLO PDF	Normalisation, shape	All, correlated	LO V + jets	2-20
Trigger efficiency	Normalisation, shape	All, correlated	All	1-2
QCD $\mu_R$ & $\mu_F$	Shape	All, correlated	$t\bar{t}H, t\bar{t}, t\bar{t}X$	0-10
Top $p_T$	Shape	All, correlated	$t\bar{t}$	< 2 %
Muon ID	Normalisation	All, correlated	All	< 1
Muon isolation	Normalisation	All, correlated	All	1
Electron ID & isolation	Normalisation	All, correlated	All	1-5
Electron reconstruction	Normalisation	All, correlated	All	1-5
Photon ID & isolation	Normalisation	All, correlated	All	1-5
$b$ -tagging with DEEPCSV	Normalisation	All, correlated	All	< 5
DEEPAK8 tagging	Normalisation	All, correlated	All	1-10
ECAL pre-firing	Normalisation	2016-17, uncorrelated	All	< 1
jet energy resolution	Normalisation	All, correlated	All	2
jet energy scale	Normalisation	All, correlated	All	3
Luminosity	Normalisation	All	All	2-3
Photon purity	All, uncorrelated	QCD in $\gamma + \text{jets}$ CR		25
QCD prediction	All, uncorrelated	QCD in signal region		50
Signal cross section	All, correlated	Signal		0.5-9

Table 6.8: The experimental uncertainties present in the analysis with the applicable samples and impact on the corresponding event counts.

## 6.9 Results

The fits of signal and background to the observed data are performed as outlined in Chpt. 6.4. Tabulated event counts in the signal region after the control region-only fit are shown in Chpt. 6.9.1 to compare the background prediction from those regions to the observed data in the signal region. To check whether the SM background by itself can accurately model the data, event yields in the signal region from the fit in the background-only hypothesis are exhibited in Chpt. 6.9.2. No significant excesses in the data are identified that can be plugged by signal, demonstrating that indeed the background-only fit can bring the data and SM background into agreement within uncertainties. Limits can therefore be set.

In Chpts. 6.9.3 and 6.9.4, the results of the fits for the  $t\bar{t}H$  and  $VH$  channels individually are shown in several forms: pre-fit vs. post-fit distributions in each category and  $p_T^{\text{miss}}$  bin; the observed upper limit on the signal strength parameter  $\sigma \times \mathcal{B}(H \rightarrow \text{invisible})/\sigma_{\text{SM}}$  at 95 % confidence level from the signal plus background hypothesis, accompanied by the median expected limit with 68 % and 95 % confidence level intervals from the background-only hypothesis; and profile likelihood scans as a function of the signal strength parameter. The outcome from combining both channels, yielding the final result, is given in Chpt. 6.9.5.

### 6.9.1 Results from the control region-only fit

From the fit invoking only the control regions to scale the standard model background events in the signal region, a sufficient description of the data can be shown. The total SM event counts in Tabs. 6.9, 6.10, and 6.11, which correspond to the 2016, 2017, and 2018 datasets, respectively, are very similar to data within uncertainties. The largest discrepancies occur, as expected, from the bins with small event counts such as in the  $t\bar{t}H$  boosted categories. However, with additional constraints provided by data and simulation in the background-only or signal plus background fits, dissimilarities can be mitigated to some extent.

Category	$p_T^{\text{miss}}$	Lost lepton	$Z \rightarrow \nu\bar{\nu}$	QCD	Total SM	Data
$t\bar{t}H$ 2Boosted	[200, 300)	$7.37 \pm 1.07$	$0 \pm 0$	$1.03 \pm 1.01$	$8.4 \pm 1.47$	12
	[300, 400)	$2.97 \pm 0.5$	$1.26 \pm 0.89$	$0.38 \pm 0.62$	$4.61 \pm 1.2$	1
	$[400, \infty)$	$0.61 \pm 0.14$	$1.46 \pm 0.75$	$0.18 \pm 0.43$	$2.24 \pm 0.87$	2
$t\bar{t}H$ 1t1b	[200, 300)	$33.7 \pm 2.91$	$6.34 \pm 2.01$	$7.11 \pm 2.67$	$47.1 \pm 4.43$	36
	[300, 400)	$25.8 \pm 2.35$	$6.61 \pm 1.37$	$2.62 \pm 1.62$	$35.0 \pm 3.16$	45
	$[400, \infty)$	$18.9 \pm 1.77$	$7.36 \pm 1.82$	$1.27 \pm 1.13$	$27.5 \pm 2.78$	32
$t\bar{t}H$ 1t2b	[200, 300)	$44.7 \pm 3.34$	$3.12 \pm 1.31$	$9.42 \pm 3.07$	$57.3 \pm 4.72$	51
	[300, 400)	$41.4 \pm 3.13$	$2.68 \pm 1.32$	$3.47 \pm 1.86$	$47.5 \pm 3.87$	40
	[400, 600)	$17.4 \pm 1.63$	$4.62 \pm 1.84$	$1.51 \pm 1.23$	$23.5 \pm 2.75$	21
	$[600, \infty)$	$2.13 \pm 0.56$	$1.1 \pm 1.0$	$0.17 \pm 0.41$	$3.39 \pm 1.22$	4
$t\bar{t}H$ 1W1b	[200, 300)	$375 \pm 15.7$	$46.2 \pm 10.8$	$21.6 \pm 4.64$	$443 \pm 19.6$	396
	[300, 400)	$83.2 \pm 5.33$	$34.9 \pm 9.96$	$7.94 \pm 2.82$	$126 \pm 11.6$	112
	$[400, \infty)$	$15.1 \pm 1.67$	$8.23 \pm 2.77$	$3.84 \pm 1.96$	$27.2 \pm 3.78$	32
$t\bar{t}H$ 1W2b	[200, 300)	$266 \pm 11.0$	$4.0 \pm 2.2$	$14.4 \pm 3.8$	$284 \pm 11.8$	264
	[300, 400)	$50.7 \pm 4.27$	$3.66 \pm 2.12$	$5.31 \pm 2.31$	$59.7 \pm 5.29$	65
	$[400, \infty)$	$7.73 \pm 1.04$	$1.07 \pm 0.7$	$2.57 \pm 1.6$	$11.4 \pm 2.03$	11
$t\bar{t}H$ 5j1b	[200, 300)	$1522 \pm 36.4$	$473 \pm 68.6$	$54.5 \pm 7.38$	$2049 \pm 78.1$	2192
	[300, 400)	$285 \pm 11.8$	$157 \pm 25.5$	$20.1 \pm 4.48$	$461 \pm 28.4$	516
	$[400, \infty)$	$55.1 \pm 4.11$	$92.1 \pm 23.0$	$9.69 \pm 3.11$	$157 \pm 23.5$	189
$t\bar{t}H$ 6j1b	[200, 300)	$900 \pm 29.3$	$232 \pm 38.0$	$41.6 \pm 6.45$	$1173 \pm 48.4$	1193
	[300, 400)	$183 \pm 9.5$	$91.8 \pm 19.3$	$15.3 \pm 3.91$	$291 \pm 21.9$	315
	$[400, \infty)$	$51.5 \pm 4.15$	$34.6 \pm 12.5$	$7.4 \pm 2.72$	$93.5 \pm 13.5$	114
$t\bar{t}H$ 5j2b	[200, 300)	$406 \pm 15.3$	$111 \pm 25.8$	$10.4 \pm 3.23$	$528 \pm 30.1$	555
	[300, 400)	$47.3 \pm 3.73$	$29.0 \pm 11.5$	$3.83 \pm 1.96$	$80.2 \pm 12.3$	87
	$[400, \infty)$	$9.38 \pm 1.53$	$16.1 \pm 6.95$	$1.85 \pm 1.36$	$27.3 \pm 7.25$	36
$t\bar{t}H$ 6j2b	[200, 300)	$291 \pm 13.5$	$73.7 \pm 20.1$	$10.9 \pm 3.3$	$375 \pm 24.5$	357
	[300, 400)	$45.3 \pm 4.11$	$49.4 \pm 13.4$	$4.01 \pm 2.0$	$98.7 \pm 14.2$	76
	$[400, \infty)$	$12.4 \pm 1.89$	$4.36 \pm 4.16$	$1.94 \pm 1.39$	$18.7 \pm 4.77$	29
$VH$ 2j0b	[200, 300)	$3914 \pm 82.2$	$6497 \pm 247$	$259 \pm 16.1$	$10669 \pm 260$	9744
	[300, 400)	$513 \pm 21.7$	$1125 \pm 40.9$	$0 \pm 0$	$1638 \pm 46.3$	1663
	$[400, \infty)$	$48.8 \pm 6.0$	$99.5 \pm 8.52$	$0 \pm 0$	$148 \pm 10.4$	172
$VH$ 2j1b	[200, 300)	$545 \pm 20.7$	$571 \pm 34.8$	$1.41 \pm 1.19$	$1117 \pm 40.5$	1060
	[300, 400)	$56.1 \pm 4.9$	$115 \pm 10.7$	$0 \pm 0$	$171 \pm 11.7$	139
	$[400, \infty)$	$3.24 \pm 0.95$	$9.91 \pm 2.73$	$0 \pm 0$	$13.1 \pm 2.89$	14
$VH$ 2j2b	$[200, \infty)$	$38.4 \pm 4.91$	$160 \pm 18.1$	$0 \pm 0$	$198 \pm 18.7$	143
$VH$ 1V	[200, 300)	$1525 \pm 39.8$	$2836 \pm 120$	$23.5 \pm 4.85$	$4384 \pm 127$	3899
	[300, 400)	$163 \pm 8.95$	$497 \pm 26.6$	$0 \pm 0$	$661 \pm 28.0$	654
	$[400, \infty)$	$6.92 \pm 1.42$	$22.1 \pm 4.73$	$0 \pm 0$	$29.1 \pm 4.94$	42

Table 6.9: Background event yields in the signal region as predicted by the control regions, and observed data event counts from the 2016 dataset.

Category	$p_T^{\text{miss}}$	Lost lepton	$Z \rightarrow \nu\bar{\nu}$	QCD	Total SM	Data
$t\bar{t}H$ 2Boosted	[200, 300)	$7.11 \pm 1.07$	$2.28 \pm 1.19$	$1.6 \pm 1.27$	$11.0 \pm 2.04$	9
	[300, 400)	$2.48 \pm 0.36$	$2.23 \pm 0.93$	$0.61 \pm 0.78$	$5.32 \pm 1.27$	6
	[400, $\infty$ )	$1.17 \pm 0.25$	$1.4 \pm 0.5$	$0.32 \pm 0.56$	$2.88 \pm 0.79$	3
$t\bar{t}H$ 1t1b	[200, 300)	$45.8 \pm 3.21$	$5.53 \pm 1.58$	$10.2 \pm 3.2$	$61.6 \pm 4.8$	45
	[300, 400)	$38.0 \pm 3.09$	$5.96 \pm 1.07$	$3.88 \pm 1.97$	$47.8 \pm 3.82$	52
	[400, $\infty$ )	$21.2 \pm 1.64$	$8.79 \pm 1.87$	$2.02 \pm 1.42$	$32.1 \pm 2.86$	28
$t\bar{t}H$ 1t2b	[200, 300)	$52.9 \pm 3.35$	$2.28 \pm 1.13$	$12.8 \pm 3.58$	$68.0 \pm 5.03$	57
	[300, 400)	$45.2 \pm 3.23$	$4.12 \pm 1.28$	$4.86 \pm 2.2$	$54.1 \pm 4.12$	38
	[400, 600)	$20.3 \pm 1.99$	$4.75 \pm 1.59$	$2.27 \pm 1.51$	$27.4 \pm 2.96$	33
	[600, $\infty$ )	$1.7 \pm 0.42$	$4.34 \pm 1.74$	$0.25 \pm 0.5$	$6.29 \pm 1.86$	1
$t\bar{t}H$ 1W1b	[200, 300)	$385 \pm 15.3$	$32.5 \pm 7.73$	$26.8 \pm 5.18$	$445 \pm 17.9$	410
	[300, 400)	$94.9 \pm 5.03$	$13.7 \pm 4.53$	$10.2 \pm 3.19$	$119 \pm 7.48$	110
	[400, $\infty$ )	$15.8 \pm 1.64$	$5.54 \pm 1.57$	$5.3 \pm 2.3$	$26.6 \pm 3.23$	35
$t\bar{t}H$ 1W2b	[200, 300)	$288 \pm 10.4$	$8.06 \pm 3.4$	$18.1 \pm 4.26$	$315 \pm 11.8$	302
	[300, 400)	$53.0 \pm 3.35$	$4.2 \pm 2.1$	$6.88 \pm 2.62$	$64.1 \pm 4.74$	74
	[400, $\infty$ )	$6.73 \pm 0.9$	$1.74 \pm 0.68$	$3.58 \pm 1.89$	$12.0 \pm 2.2$	14
$t\bar{t}H$ 5j1b	[200, 300)	$1808 \pm 42.9$	$502 \pm 59.8$	$73.0 \pm 8.54$	$2383 \pm 74.1$	2514
	[300, 400)	$289 \pm 11.9$	$244 \pm 36.4$	$27.7 \pm 5.26$	$561 \pm 38.6$	540
	[400, $\infty$ )	$55.2 \pm 3.81$	$82.3 \pm 18.3$	$14.4 \pm 3.8$	$152 \pm 19.1$	152
$t\bar{t}H$ 6j1b	[200, 300)	$997 \pm 25.2$	$250 \pm 34.4$	$54.0 \pm 7.35$	$1301 \pm 43.2$	1295
	[300, 400)	$202 \pm 9.39$	$110 \pm 20.8$	$20.5 \pm 4.53$	$332 \pm 23.3$	343
	[400, $\infty$ )	$54.0 \pm 4.07$	$40.1 \pm 11.9$	$10.7 \pm 3.26$	$105 \pm 13.0$	121
$t\bar{t}H$ 5j2b	[200, 300)	$511 \pm 18.2$	$122 \pm 25.6$	$13.5 \pm 3.68$	$647 \pm 31.6$	651
	[300, 400)	$45.6 \pm 3.51$	$24.9 \pm 10.4$	$5.14 \pm 2.27$	$75.6 \pm 11.3$	106
	[400, $\infty$ )	$8.78 \pm 1.4$	$22.7 \pm 9.85$	$2.67 \pm 1.63$	$34.1 \pm 10.1$	27
$t\bar{t}H$ 6j2b	[200, 300)	$335 \pm 12.9$	$57.4 \pm 16.6$	$12.9 \pm 3.59$	$406 \pm 21.3$	422
	[300, 400)	$51.2 \pm 4.4$	$27.7 \pm 11.0$	$4.9 \pm 2.21$	$83.8 \pm 12.1$	87
	[400, $\infty$ )	$6.86 \pm 1.1$	$4.62 \pm 4.0$	$2.55 \pm 1.6$	$14.0 \pm 4.45$	36
$VH$ 2j0b	[200, 300)	$4152 \pm 90.8$	$6090 \pm 208$	$379 \pm 19.5$	$10622 \pm 228$	11263
	[300, 400)	$574 \pm 23.9$	$964 \pm 36.6$	$0 \pm 0$	$1538 \pm 43.7$	1696
	[400, $\infty$ )	$39.0 \pm 4.37$	$103 \pm 9.76$	$0 \pm 0$	$142 \pm 10.7$	149
$VH$ 2j1b	[200, 300)	$569 \pm 20.9$	$619 \pm 36.1$	$2.07 \pm 1.44$	$1190 \pm 41.8$	1235
	[300, 400)	$56.6 \pm 4.26$	$126 \pm 12.1$	$0 \pm 0$	$183 \pm 12.8$	190
	[400, $\infty$ )	$4.27 \pm 1.06$	$9.3 \pm 2.42$	$0 \pm 0$	$13.6 \pm 2.64$	14
$VH$ 2j2b	[200, $\infty$ )	$32.7 \pm 3.21$	$132 \pm 15.8$	$0 \pm 0$	$165 \pm 16.1$	164
$VH$ 1V	[200, 300)	$1349 \pm 42.9$	$2394 \pm 106$	$34.5 \pm 5.88$	$3778 \pm 114$	4005
	[300, 400)	$169 \pm 8.88$	$511 \pm 30.4$	$0 \pm 0$	$679 \pm 31.7$	612
	[400, $\infty$ )	$6.34 \pm 1.14$	$34.8 \pm 5.87$	$0 \pm 0$	$41.2 \pm 5.98$	50

Table 6.10: Background event yields in the signal region as predicted by the control regions, and observed data event counts from the 2017 dataset.

Category	$p_T^{\text{miss}}$	Lost lepton	$Z \rightarrow \nu\bar{\nu}$	QCD	Total SM	Data
$t\bar{t}H$ 2Boosted	[200, 300)	$11.2 \pm 1.3$	$1.97 \pm 1.11$	$2.2 \pm 1.48$	$15.4 \pm 2.26$	9
	[300, 400)	$3.48 \pm 0.51$	$1.25 \pm 0.67$	$0.86 \pm 0.93$	$5.58 \pm 1.26$	7
	$[400, \infty)$	$1.59 \pm 0.4$	$1.58 \pm 0.54$	$0.45 \pm 0.67$	$3.62 \pm 0.95$	5
$t\bar{t}H$ 1t1b	[200, 300)	$52.4 \pm 3.22$	$7.5 \pm 1.55$	$12.4 \pm 3.52$	$72.3 \pm 5.02$	74
	[300, 400)	$42.6 \pm 3.14$	$8.43 \pm 1.29$	$4.83 \pm 2.2$	$55.8 \pm 4.04$	49
	$[400, \infty)$	$26.1 \pm 1.99$	$11.2 \pm 2.45$	$2.52 \pm 1.59$	$39.8 \pm 3.53$	38
$t\bar{t}H$ 1t2b	[200, 300)	$72.2 \pm 4.04$	$7.31 \pm 2.21$	$17.9 \pm 4.23$	$97.4 \pm 6.25$	82
	[300, 400)	$58.8 \pm 3.58$	$4.8 \pm 1.32$	$6.95 \pm 2.64$	$70.5 \pm 4.64$	72
	[400, 600)	$27.5 \pm 1.96$	$3.67 \pm 1.34$	$3.26 \pm 1.81$	$34.4 \pm 2.98$	42
	$[600, \infty)$	$2.27 \pm 0.45$	$3.03 \pm 1.25$	$0.36 \pm 0.6$	$5.66 \pm 1.45$	5
$t\bar{t}H$ 1W1b	[200, 300)	$523 \pm 15.7$	$47.8 \pm 9.34$	$36.7 \pm 6.06$	$607 \pm 19.3$	612
	[300, 400)	$114 \pm 4.9$	$15.5 \pm 4.11$	$14.3 \pm 3.78$	$144 \pm 7.43$	149
	$[400, \infty)$	$17.7 \pm 1.71$	$6.14 \pm 2.15$	$7.45 \pm 2.73$	$31.3 \pm 3.87$	48
$t\bar{t}H$ 1W2b	[200, 300)	$430 \pm 14.9$	$21.1 \pm 5.75$	$30.5 \pm 5.52$	$481 \pm 16.9$	467
	[300, 400)	$84.2 \pm 4.9$	$5.88 \pm 2.82$	$11.9 \pm 3.44$	$102 \pm 6.62$	92
	$[400, \infty)$	$14.1 \pm 1.33$	$3.62 \pm 1.28$	$6.18 \pm 2.49$	$23.9 \pm 3.1$	27
$t\bar{t}H$ 5j1b	[200, 300)	$2043 \pm 37.9$	$672 \pm 66.6$	$93.0 \pm 9.65$	$2808 \pm 77.2$	2989
	[300, 400)	$390 \pm 14.6$	$283 \pm 35.4$	$36.2 \pm 6.01$	$709 \pm 38.7$	698
	$[400, \infty)$	$82.0 \pm 4.81$	$109 \pm 20.6$	$18.9 \pm 4.34$	$210 \pm 21.6$	238
$t\bar{t}H$ 6j1b	[200, 300)	$1206 \pm 27.5$	$291 \pm 41.5$	$68.2 \pm 8.26$	$1565 \pm 50.4$	1574
	[300, 400)	$261 \pm 11.1$	$184 \pm 31.1$	$26.5 \pm 5.15$	$472 \pm 33.4$	445
	$[400, \infty)$	$64.4 \pm 4.09$	$109 \pm 22.7$	$13.8 \pm 3.72$	$187 \pm 23.3$	188
$t\bar{t}H$ 5j2b	[200, 300)	$610 \pm 18.1$	$151 \pm 27.4$	$20.5 \pm 4.53$	$782 \pm 33.2$	788
	[300, 400)	$82.4 \pm 4.97$	$58.9 \pm 14.4$	$7.99 \pm 2.83$	$149 \pm 15.5$	159
	$[400, \infty)$	$16.7 \pm 1.94$	$26.4 \pm 10.8$	$4.17 \pm 2.04$	$47.2 \pm 11.1$	45
$t\bar{t}H$ 6j2b	[200, 300)	$449 \pm 15.5$	$119 \pm 23.1$	$20.6 \pm 4.54$	$589 \pm 28.2$	593
	[300, 400)	$86.6 \pm 6.25$	$41.3 \pm 12.6$	$8.02 \pm 2.83$	$136 \pm 14.3$	144
	$[400, \infty)$	$19.1 \pm 2.19$	$23.8 \pm 10.1$	$4.19 \pm 2.05$	$47.2 \pm 10.5$	47
$VH$ 2j0b	[200, 300)	$4543 \pm 81.3$	$7000 \pm 217$	$400 \pm 20$	$11943 \pm 233$	12086
	[300, 400)	$542 \pm 22.1$	$1149 \pm 37.6$	$0 \pm 0$	$1691 \pm 43.6$	1827
	$[400, \infty)$	$65.0 \pm 6.19$	$124 \pm 10.9$	$0 \pm 0$	$189 \pm 12.6$	208
$VH$ 2j1b	[200, 300)	$707 \pm 23.1$	$685 \pm 36.6$	$2.19 \pm 1.48$	$1394 \pm 43.4$	1519
	[300, 400)	$93.0 \pm 6.35$	$132 \pm 10.7$	$0 \pm 0$	$225 \pm 12.5$	236
	$[400, \infty)$	$7.71 \pm 1.79$	$17.6 \pm 3.28$	$0 \pm 0$	$25.3 \pm 3.74$	27
$VH$ 2j2b	$[200, \infty)$	$58.1 \pm 5.06$	$216 \pm 21.9$	$0 \pm 0$	$274 \pm 22.5$	208
$VH$ 1V	[200, 300)	$1481 \pm 37$	$2710 \pm 93.8$	$36.4 \pm 6.04$	$4227 \pm 101$	4815
	[300, 400)	$173 \pm 7.24$	$572 \pm 26.8$	$0 \pm 0$	$746 \pm 27.7$	816
	$[400, \infty)$	$6.08 \pm 0.95$	$70.7 \pm 9.81$	$0 \pm 0$	$76.8 \pm 9.85$	51

Table 6.11: Background event yields in the signal region as predicted by the control regions, and observed data event counts from the 2018 dataset.

### 6.9.2 Results from the background-only fit

Category	$p_T^{\text{miss}}$	Lost lepton	$Z \rightarrow \nu\bar{\nu}$	QCD	Total SM	Data
$t\bar{t}H$ 2Boosted	[200, 300)	$7.88 \pm 1.04$	$0 \pm 0$	$0.88 \pm 0.27$	$8.76 \pm 1.08$	12
	[300, 400)	$3.03 \pm 0.52$	$0.84 \pm 0.54$	$0.32 \pm 0.1$	$4.19 \pm 0.75$	1
	[400, $\infty$ )	$0.61 \pm 0.13$	$1.5 \pm 0.74$	$0.16 \pm 0.05$	$2.26 \pm 0.75$	2
$t\bar{t}H$ 1t1b	[200, 300)	$32.6 \pm 2.52$	$5.79 \pm 1.64$	$5.97 \pm 1.82$	$44.3 \pm 3.51$	36
	[300, 400)	$27 \pm 2.39$	$7.15 \pm 1.56$	$2.24 \pm 0.67$	$36.4 \pm 2.93$	45
	[400, $\infty$ )	$19.3 \pm 1.66$	$8.34 \pm 2.0$	$1.08 \pm 0.33$	$28.7 \pm 2.62$	32
$t\bar{t}H$ 1t2b	[200, 300)	$44.2 \pm 2.72$	$3.02 \pm 1.28$	$7.95 \pm 2.44$	$55.2 \pm 3.88$	51
	[300, 400)	$40.3 \pm 2.58$	$2.38 \pm 1.1$	$2.94 \pm 0.87$	$45.6 \pm 2.94$	40
	[400, 600)	$17.1 \pm 1.61$	$4.49 \pm 1.62$	$1.28 \pm 0.39$	$22.8 \pm 2.32$	21
	[600, $\infty$ )	$2.22 \pm 0.52$	$1.36 \pm 0.99$	$0.14 \pm 0.04$	$3.72 \pm 1.12$	4
$t\bar{t}H$ 1W1b	[200, 300)	$362 \pm 11.7$	$41.3 \pm 8.72$	$18.1 \pm 5.54$	$422 \pm 15.6$	396
	[300, 400)	$81.5 \pm 5.1$	$32.4 \pm 7.12$	$6.71 \pm 2.04$	$121 \pm 8.99$	112
	[400, $\infty$ )	$15.3 \pm 1.45$	$10.6 \pm 2.88$	$3.27 \pm 1.01$	$29.1 \pm 3.38$	32
$t\bar{t}H$ 1W2b	[200, 300)	$260 \pm 9.2$	$3.9 \pm 2.19$	$12.2 \pm 3.77$	$276 \pm 10.2$	264
	[300, 400)	$51.3 \pm 3.26$	$4.41 \pm 2.52$	$4.53 \pm 1.38$	$60.3 \pm 4.35$	65
	[400, $\infty$ )	$7.8 \pm 1.01$	$1.2 \pm 0.73$	$2.18 \pm 0.68$	$11.2 \pm 1.41$	11
$t\bar{t}H$ 5j1b	[200, 300)	$1557 \pm 32.9$	$558 \pm 53.4$	$46.5 \pm 14.0$	$2161 \pm 64.3$	2192
	[300, 400)	$292 \pm 11.0$	$193 \pm 18.6$	$17.1 \pm 5.51$	$502 \pm 22.3$	516
	[400, $\infty$ )	$54.5 \pm 4.13$	$119 \pm 12.6$	$8.25 \pm 2.51$	$182 \pm 13.5$	189
$t\bar{t}H$ 6j1b	[200, 300)	$907 \pm 23.1$	$250 \pm 30.6$	$35.3 \pm 10.9$	$1192 \pm 39.9$	1193
	[300, 400)	$186 \pm 9.7$	$109 \pm 15.1$	$13 \pm 3.97$	$309 \pm 18.4$	315
	[400, $\infty$ )	$51.5 \pm 3.68$	$50.1 \pm 10.1$	$6.31 \pm 1.92$	$108 \pm 10.9$	114
$t\bar{t}H$ 5j2b	[200, 300)	$413 \pm 14.3$	$130 \pm 23.3$	$8.84 \pm 2.69$	$552 \pm 27.5$	555
	[300, 400)	$48.1 \pm 3.85$	$34.2 \pm 7.73$	$3.26 \pm 1.01$	$85.5 \pm 8.7$	87
	[400, $\infty$ )	$9.32 \pm 1.42$	$22.6 \pm 4.15$	$1.57 \pm 0.48$	$33.4 \pm 4.41$	36
$t\bar{t}H$ 6j2b	[200, 300)	$288 \pm 12.3$	$70.9 \pm 15.2$	$9.21 \pm 2.8$	$368 \pm 19.7$	357
	[300, 400)	$43.7 \pm 3.43$	$37.3 \pm 7.39$	$3.39 \pm 1.08$	$84.4 \pm 8.22$	76
	[400, $\infty$ )	$12.8 \pm 1.94$	$11.3 \pm 4.77$	$1.65 \pm 0.51$	$25.7 \pm 5.17$	29
$VH$ 2j0b	[200, 300)	$3804 \pm 72.6$	$5792 \pm 108$	$200 \pm 73.3$	$9796 \pm 150$	9744
	[300, 400)	$520 \pm 17.9$	$1090 \pm 24.8$	$0 \pm 0$	$1610 \pm 30.6$	1663
	[400, $\infty$ )	$51.5 \pm 5.63$	$109 \pm 7.55$	$0 \pm 0$	$161 \pm 9.42$	172
$VH$ 2j1b	[200, 300)	$535 \pm 18.2$	$520 \pm 21.4$	$1.1 \pm 0.41$	$1056 \pm 28.1$	1060
	[300, 400)	$52.5 \pm 4.32$	$101 \pm 7.62$	$0 \pm 0$	$153 \pm 8.76$	139
	[400, $\infty$ )	$3.35 \pm 1.0$	$10.4 \pm 2.15$	$0 \pm 0$	$13.7 \pm 2.37$	14
$VH$ 2j2b	[200, $\infty$ )	$35.4 \pm 3.84$	$125 \pm 10.4$	$0 \pm 0$	$160 \pm 11.1$	143
$VH$ 1V	[200, 300)	$1459 \pm 37.1$	$2521 \pm 55.8$	$18.3 \pm 6.71$	$3998 \pm 67.4$	3899
	[300, 400)	$162 \pm 9.69$	$484 \pm 17.8$	$0 \pm 0$	$646 \pm 20.2$	654
	[400, $\infty$ )	$7.51 \pm 1.41$	$28 \pm 4.17$	$0 \pm 0$	$35.5 \pm 4.4$	42

Table 6.12: Results of the combined fit to signal and control regions with the 2016 dataset. The total predicted standard model background is compared to events observed in data. A breakdown of the individual contributions from the lost lepton,  $Z \rightarrow \nu\bar{\nu}$ , and QCD processes as determined by the fit is also presented.

Category	$p_T^{\text{miss}}$	Lost lepton	$Z \rightarrow \nu\bar{\nu}$	QCD	Total SM	Data
$t\bar{t}H$ 2Boosted	[200, 300)	$7.03 \pm 0.95$	$2.04 \pm 1.04$	$1.28 \pm 0.35$	$10.4 \pm 1.45$	9
	[300, 400)	$2.49 \pm 0.32$	$2.39 \pm 1.01$	$0.49 \pm 0.13$	$5.37 \pm 1.07$	6
	$[400, \infty)$	$1.13 \pm 0.24$	$1.56 \pm 0.54$	$0.25 \pm 0.07$	$2.95 \pm 0.59$	3
$t\bar{t}H$ 1t1b	[200, 300)	$44.4 \pm 2.64$	$4.95 \pm 1.29$	$8.05 \pm 2.26$	$57.4 \pm 3.71$	45
	[300, 400)	$38.8 \pm 2.79$	$6.07 \pm 1.1$	$3.12 \pm 0.88$	$48.0 \pm 3.12$	52
	$[400, \infty)$	$20.8 \pm 1.63$	$9.34 \pm 1.72$	$1.62 \pm 0.45$	$31.8 \pm 2.41$	28
$t\bar{t}H$ 1t2b	[200, 300)	$51.9 \pm 3.12$	$2.1 \pm 1.11$	$10.1 \pm 2.74$	$64.2 \pm 4.29$	57
	[300, 400)	$43.2 \pm 2.72$	$3.72 \pm 1.24$	$3.86 \pm 1.05$	$50.7 \pm 3.17$	38
	[400, 600)	$20.9 \pm 2.0$	$5.53 \pm 1.79$	$1.83 \pm 0.51$	$28.3 \pm 2.73$	33
$t\bar{t}H$ 1W1b	[600, $\infty$ )	$1.57 \pm 0.37$	$3.01 \pm 1.06$	$0.2 \pm 0.06$	$4.79 \pm 1.12$	1
	[200, 300)	$378 \pm 12.6$	$30.1 \pm 5.82$	$21.3 \pm 5.83$	$430 \pm 15.0$	410
	[300, 400)	$93.8 \pm 4.48$	$13.8 \pm 4.38$	$8.14 \pm 2.25$	$116 \pm 6.66$	110
$t\bar{t}H$ 1W2b	$[400, \infty)$	$16.4 \pm 1.35$	$6.95 \pm 1.78$	$4.3 \pm 1.19$	$27.6 \pm 2.53$	35
	[200, 300)	$286 \pm 11.5$	$7.8 \pm 3.38$	$14.5 \pm 3.93$	$309 \pm 12.6$	302
	[300, 400)	$54.9 \pm 3.56$	$5.11 \pm 2.46$	$5.56 \pm 1.57$	$65.6 \pm 4.6$	74
$t\bar{t}H$ 5j1b	$[400, \infty)$	$6.92 \pm 0.88$	$2.04 \pm 0.71$	$2.88 \pm 0.79$	$11.8 \pm 1.39$	14
	[200, 300)	$1842 \pm 33.1$	$570 \pm 48.7$	$59.1 \pm 16.5$	$2471 \pm 61.2$	2514
	[300, 400)	$290 \pm 10.4$	$239 \pm 21.9$	$22.1 \pm 6.49$	$551 \pm 25.1$	540
$t\bar{t}H$ 6j1b	$[400, \infty)$	$54.6 \pm 3.61$	$85.6 \pm 11.9$	$11.6 \pm 3.24$	$152 \pm 12.9$	152
	[200, 300)	$1001 \pm 22.2$	$259 \pm 32.5$	$43.2 \pm 12.2$	$1303 \pm 41.2$	1295
	[300, 400)	$204 \pm 8.13$	$120 \pm 15.5$	$16.5 \pm 4.66$	$341 \pm 18.1$	343
$t\bar{t}H$ 5j2b	$[400, \infty)$	$54.3 \pm 3.89$	$52.2 \pm 9.05$	$8.59 \pm 2.48$	$115 \pm 10.2$	121
	[200, 300)	$515 \pm 15.4$	$129 \pm 19.0$	$10.9 \pm 2.93$	$654 \pm 24.6$	651
	[300, 400)	$47.5 \pm 3.67$	$44.0 \pm 8.57$	$4.14 \pm 1.16$	$95.7 \pm 9.39$	106
$t\bar{t}H$ 6j2b	$[400, \infty)$	$8.59 \pm 1.2$	$17.9 \pm 4.75$	$2.14 \pm 0.58$	$28.7 \pm 4.93$	27
	[200, 300)	$340 \pm 11.5$	$64.9 \pm 14.6$	$10.4 \pm 2.87$	$415 \pm 18.8$	422
	[300, 400)	$51.8 \pm 4.39$	$30.0 \pm 7.21$	$3.94 \pm 1.12$	$85.8 \pm 8.51$	87
$VH$ 2j0b	$[400, \infty)$	$7.05 \pm 1.14$	$21.8 \pm 4.79$	$2.05 \pm 0.56$	$30.9 \pm 4.96$	36
	[200, 300)	$4225 \pm 76.3$	$6459 \pm 162$	$546 \pm 194$	$11230 \pm 264$	11263
	[300, 400)	$605 \pm 23.5$	$1033 \pm 26.3$	$0 \pm 0$	$1638 \pm 35.3$	1696
$VH$ 2j1b	$[400, \infty)$	$40 \pm 3.7$	$106 \pm 7.53$	$0 \pm 0$	$146 \pm 8.39$	149
	[200, 300)	$577 \pm 18.2$	$652 \pm 26.8$	$2.94 \pm 1.1$	$1232 \pm 32.4$	1235
	[300, 400)	$57.4 \pm 4.8$	$132 \pm 9.66$	$0 \pm 0$	$189 \pm 10.8$	190
$VH$ 2j2b	$[400, \infty)$	$4.27 \pm 1.06$	$9.54 \pm 1.93$	$0 \pm 0$	$13.8 \pm 2.2$	14
	[200, $\infty$ )	$32.9 \pm 3.48$	$133 \pm 10.2$	$0 \pm 0$	$166 \pm 10.8$	164
	$VH$ 1V		$1381 \pm 39.3$	$2536 \pm 65.7$	$49.2 \pm 18.7$	$3965 \pm 78.8$
	[200, 300)	$167 \pm 8.61$	$489 \pm 19.7$	$0 \pm 0$	$656 \pm 21.5$	612
	[300, 400)	$6.53 \pm 0.95$	$38.9 \pm 5.0$	$0 \pm 0$	$45.4 \pm 5.09$	50

Table 6.13: Results of the combined fit to signal and control regions with the 2017 dataset. The total predicted standard model background is compared to events observed in data. A breakdown of the individual contributions from the lost lepton,  $Z \rightarrow \nu\bar{\nu}$ , and QCD processes as determined by the fit is also presented.

Category	$p_T^{\text{miss}}$	Lost lepton	$Z \rightarrow \nu\bar{\nu}$	QCD	Total SM	Data
$t\bar{t}H$ 2Boosted	[200, 300)	$11.0 \pm 1.05$	$1.52 \pm 0.89$	$2.25 \pm 0.51$	$14.8 \pm 1.47$	9
	[300, 400)	$3.54 \pm 0.45$	$1.35 \pm 0.73$	$0.88 \pm 0.2$	$5.78 \pm 0.88$	7
	$[400, \infty)$	$1.64 \pm 0.34$	$1.9 \pm 0.61$	$0.46 \pm 0.1$	$4.0 \pm 0.7$	5
$t\bar{t}H$ 1t1b	[200, 300)	$52.7 \pm 2.93$	$7.68 \pm 1.29$	$12.8 \pm 2.89$	$73.2 \pm 4.31$	74
	[300, 400)	$41.5 \pm 2.76$	$8.29 \pm 1.11$	$4.95 \pm 1.06$	$54.7 \pm 3.16$	49
	$[400, \infty)$	$25.6 \pm 1.81$	$12.5 \pm 2.22$	$2.59 \pm 0.58$	$40.7 \pm 2.93$	38
$t\bar{t}H$ 1t2b	[200, 300)	$70.2 \pm 3.63$	$6.67 \pm 1.76$	$18.0 \pm 3.85$	$94.8 \pm 5.57$	82
	[300, 400)	$58.9 \pm 3.39$	$4.69 \pm 1.31$	$7.17 \pm 1.61$	$70.8 \pm 3.98$	72
	[400, 600)	$28.0 \pm 1.98$	$4.26 \pm 1.47$	$3.38 \pm 0.77$	$35.7 \pm 2.58$	42
	$[600, \infty)$	$2.24 \pm 0.44$	$3.15 \pm 1.08$	$0.37 \pm 0.08$	$5.76 \pm 1.16$	5
$t\bar{t}H$ 1W1b	[200, 300)	$522 \pm 12.8$	$48.5 \pm 8.15$	$37.9 \pm 8.5$	$608 \pm 17.4$	612
	[300, 400)	$115 \pm 5.28$	$17.0 \pm 4.18$	$14.7 \pm 3.38$	$146 \pm 7.53$	149
	$[400, \infty)$	$18.4 \pm 1.46$	$8.65 \pm 2.47$	$7.89 \pm 1.87$	$35.0 \pm 3.43$	48
$t\bar{t}H$ 1W2b	[200, 300)	$426 \pm 12.3$	$20.3 \pm 5.64$	$31.2 \pm 6.93$	$478 \pm 15.2$	467
	[300, 400)	$82.6 \pm 3.82$	$5.71 \pm 2.46$	$12.1 \pm 2.84$	$100 \pm 5.36$	92
	$[400, \infty)$	$14.2 \pm 1.25$	$4.5 \pm 1.47$	$6.4 \pm 1.45$	$25.1 \pm 2.42$	27
$t\bar{t}H$ 5j1b	[200, 300)	$2074 \pm 33.6$	$761 \pm 56.3$	$97.6 \pm 22.1$	$2933 \pm 69.2$	2989
	[300, 400)	$387 \pm 13.5$	$284 \pm 23.7$	$37.0 \pm 8.75$	$708 \pm 28.7$	698
	$[400, \infty)$	$80.8 \pm 4.84$	$129 \pm 14.2$	$19.6 \pm 4.65$	$230 \pm 15.7$	238
$t\bar{t}H$ 6j1b	[200, 300)	$1207 \pm 25.2$	$302 \pm 29.1$	$70.1 \pm 16.2$	$1579 \pm 41.7$	1574
	[300, 400)	$258 \pm 9.96$	$176 \pm 17.6$	$27.1 \pm 6.29$	$461 \pm 21.2$	445
	$[400, \infty)$	$63.2 \pm 3.97$	$111 \pm 12.9$	$14.2 \pm 3.32$	$189 \pm 13.9$	188
$t\bar{t}H$ 5j2b	[200, 300)	$612 \pm 15.7$	$161 \pm 21.9$	$21.1 \pm 4.97$	$794 \pm 27.4$	788
	[300, 400)	$82.8 \pm 5.01$	$65.7 \pm 10.4$	$8.24 \pm 1.82$	$157 \pm 11.7$	159
	$[400, \infty)$	$16.5 \pm 1.62$	$25 \pm 5.5$	$4.29 \pm 0.98$	$45.8 \pm 5.81$	45
$t\bar{t}H$ 6j2b	[200, 300)	$450 \pm 13.6$	$123 \pm 18.0$	$21.2 \pm 4.79$	$594 \pm 23.1$	593
	[300, 400)	$87.1 \pm 5.78$	$46.3 \pm 9.9$	$8.27 \pm 1.85$	$142 \pm 11.6$	144
	$[400, \infty)$	$18.8 \pm 2.28$	$24.1 \pm 6.38$	$4.31 \pm 1.0$	$47.1 \pm 6.85$	47
$VH$ 2j0b	[200, 300)	$4560 \pm 65.9$	$7298 \pm 132$	$321 \pm 112$	$12180 \pm 185$	12086
	[300, 400)	$561 \pm 19.5$	$1231 \pm 25.9$	$0 \pm 0$	$1792 \pm 32.4$	1827
	$[400, \infty)$	$67.7 \pm 7.05$	$135 \pm 8.7$	$0 \pm 0$	$202 \pm 11.2$	208
$VH$ 2j1b	[200, 300)	$729 \pm 19.9$	$747 \pm 31.1$	$1.8 \pm 0.64$	$1477 \pm 36.9$	1519
	[300, 400)	$95.7 \pm 5.76$	$140 \pm 9.24$	$0 \pm 0$	$235 \pm 10.9$	236
	$[400, \infty)$	$7.81 \pm 1.69$	$18.7 \pm 2.95$	$0 \pm 0$	$26.5 \pm 3.4$	27
$VH$ 2j2b	$[200, \infty)$	$56.3 \pm 4.52$	$182 \pm 13.1$	$0 \pm 0$	$238 \pm 13.8$	208
$VH$ 1V	[200, 300)	$1545 \pm 34.7$	$3049 \pm 63.1$	$30.4 \pm 11.0$	$4625 \pm 72.9$	4815
	[300, 400)	$179 \pm 7.8$	$617 \pm 20.1$	$0 \pm 0$	$796 \pm 21.6$	816
	$[400, \infty)$	$5.95 \pm 0.85$	$58.9 \pm 5.64$	$0 \pm 0$	$64.9 \pm 5.7$	51

Table 6.14: Results of the combined fit to signal and control regions with the 2018 dataset. The total predicted standard model background is compared to events observed in data. A breakdown of the individual contributions from the lost lepton,  $Z \rightarrow \nu\bar{\nu}$ , and QCD processes as determined by the fit is also presented.

The results of the background-only fit to data model it well, in large part. As noted above, no significant excess is observed in the data. With the SM background describing the data sufficiently, only upper limits can be set on  $\mathcal{B}(H \rightarrow \text{invisible})$ . Post-fit distributions in the signal region—visualising the event counts in the Tabs. 6.12, 6.13, and 6.14—are depicted in Apps. A.5 and A.6 for the  $t\bar{t}H$  and  $VH$  categories, respectively. The “pre-fit” yields are from the results of the control region-only fit documented in Chpt. 6.9.1.

As expected, knowledge of the background and data in the signal region permits the background-only fit to model the data better than the control region-only fit. Noticeable improvements are in the  $t\bar{t}H$  and  $VH$  resolved categories. While small in some cases, the addition of the QCD multijet background with 50 % uncertainty creates an additional avenue for constraining the backgrounds. Strong sensitivity translates to a small value of the expected upper limit on  $\mathcal{B}(H \rightarrow \text{invisible})$ .

### 6.9.3 Results from the $t\bar{t}H$ channel

Distributions of the yields in the signal region for 2016, 2017, and 2018 from the signal plus background fit are displayed in Fig. 6.18. The SM prediction from the control region-only fit is also overlaid, denoted as “pre-fit.” Event counts from said fit are tabulated in the previous section. Corresponding figures for the control region distributions from the control region-only fit are given in App. A.1.

It is evident from the distributions that the fit succeeds in most cases to match the signal and background to the data within uncertainties. Relatively few bins contain a post-fit data/prediction ratio largely different from unity. Sizeable deviations are seen only in the  $t\bar{t}H$  boosted categories which are statistically limited. In 2016, a general over-prediction of background is seen, corrected for the most part by the fit. The pre-fit ratio is overall better in 2016 than in the other years; the effect of pre-firing was less severe than in 2017. Both of the later years suffered from additional problems such as noise in the ECAL end caps in 2017—affecting  $t\bar{t}H$  more given its high jet multiplicity—and the HEM issue in 2018 that largely affects jets, and therefore  $\vec{p}_T^{\text{miss}}$ . Techniques were employed to mitigate the problems, but residual mis-modelling may still be present.

QCD multijet is a somewhat large background in several of the categories despite the  $\Delta\phi_{\min}$ ,  $\tilde{\omega}_{\min}$ , and  $b$ -jet/boosted object requirements. Even though they are designed to reject a significant fraction of the multijet background, the high cross section of the process leads to a noticeable contribution. QCD comprises roughly 2–3 % of the total background in the resolved categories. However, at low  $p_T^{\text{miss}}$  in the boosted categories, its contribution is as high as approximately 10 %.

## CHAPTER 6. SEARCH FOR INVISIBLY DECAYING HIGGS BOSONS

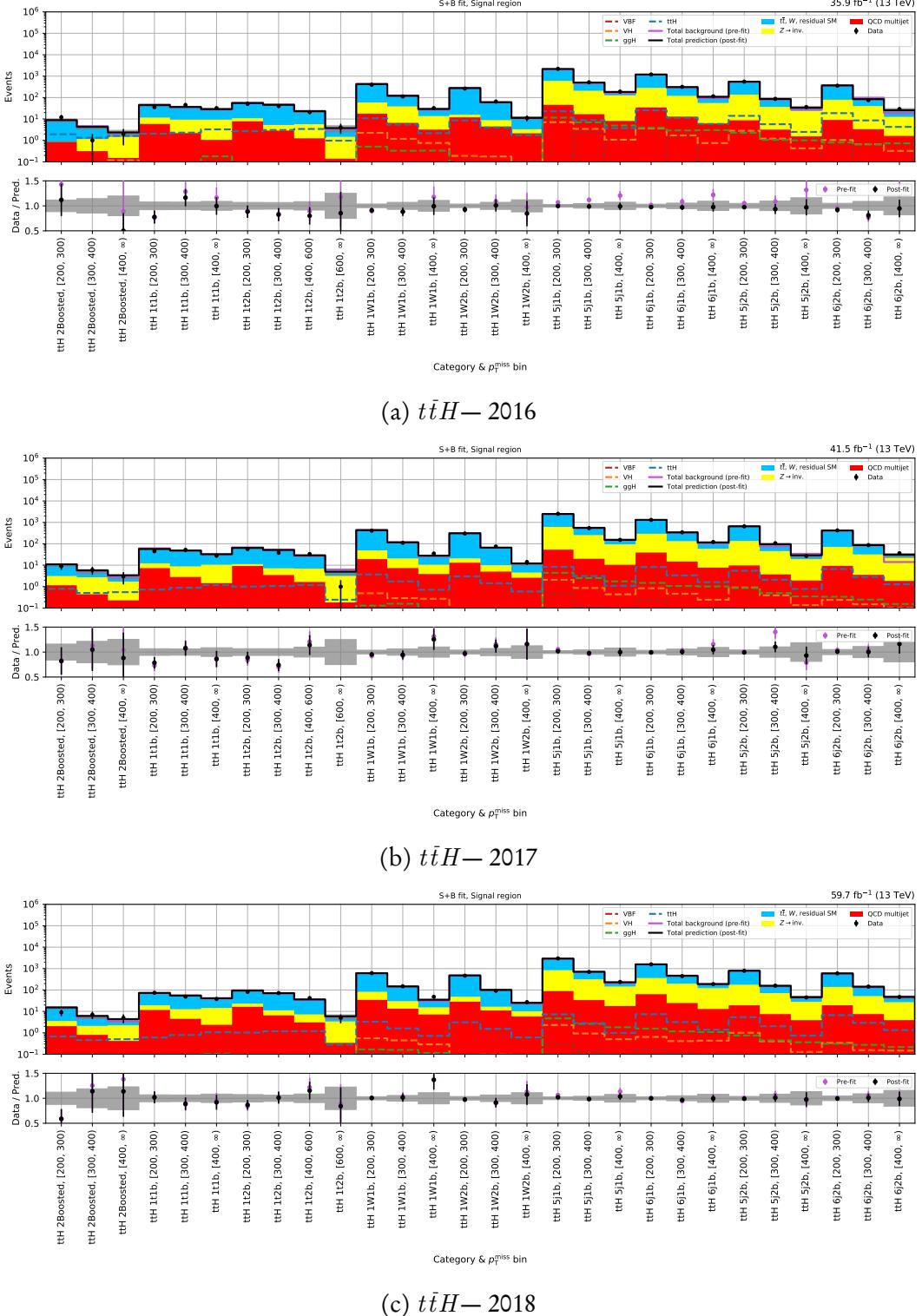


Figure 6.18: Pre-fit and post-fit yields in the signal region for every  $t\bar{t}H$  category and  $p_T^{\text{miss}}$  bin in each year of Run-2. “Pre-fit” refers to outcome of the control region-only prediction.

These post-fit distributions translate directly into the upper limit on  $\sigma \times \mathcal{B}(H \rightarrow \text{invisible})/\sigma_{\text{SM}}$ . Fig. 6.19 showcases the limit and profile likelihood scan for the  $t\bar{t}H$  channel in each data taking year individually and the combination over the full Run-2 dataset. Limits broken down by category in each year are presented in Fig. A.11.

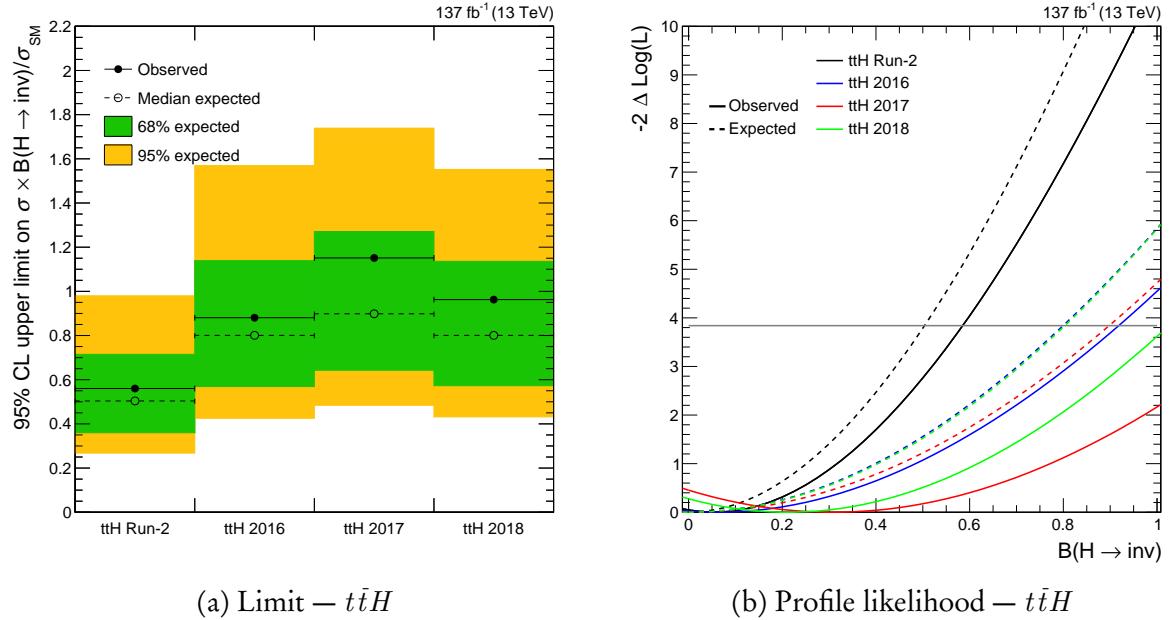


Figure 6.19: Observed and expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction  $\mathcal{B}(H \rightarrow \text{invisible})$  (left) and the corresponding profile likelihood scan (right) in the  $t\bar{t}H$  channel. The result from each data taking period is presented along with their combination.

The sensitivity landscape is mostly consistent between years, more so for 2017 and 2018 given the similarities in simulation and detector configuration. All of the observations are within the  $2\sigma$  boundaries of the expected limit, with many within  $1\sigma$ , demonstrating a consistent picture between the SM expectation and data. The combined Run-2 limit of 50 % expected and 56 % observed is more sensitive than the analogous combination from ATLAS with 94 % observed and 64 % expected [24]. Comparing to the preliminary 2016 result from CMS exhibits a very similar sensitivity, with 85 % observed and 73 % expected in the 0-lepton channel [83]. The better sensitivity in the preliminary result could be explained by a number of reasons. Additional systematic uncertainties were included in this analysis such as the QCD scale for top quark processes, jet energy resolution, and those associated with the NLO corrections to  $V + \text{jets}$  backgrounds.

By category, those targeting the boosted topologies are the most sensitive to the invisible decays of the Higgs boson (see Fig. A.13). Given the purity of these categories, it is

understandable why. Despite greater statistical power in the resolved regime, systematic uncertainties would have a larger absolute effect and possibly conceal signal due to their size. These result in generally weaker limits than the boosted categories.

The profile likelihood scans illustrate how the likelihood function for the fit over a set of categories evolves with the upper limit on branching ratio. The horizontal line  $-2\Delta \ln(\mathcal{L}) = 3.84$  represents the 95 % confidence level. It can therefore be seen that the intersection of a dashed curve with the line is equal to the median expected limit, and the intersection of a solid curve with it equals the measured observed limit. Indicative of the stability and health of the fit, the likelihood scans in Fig. 6.19 are satisfactory with minima near  $\mathcal{B} = 0$ , rising smoothly either side of it.

#### 6.9.4 Results from the $VH$ channel

Distributions of the prediction from the signal plus background fit to data, for 2016, 2017, and 2018, are displayed in Fig. 6.20. The SM prediction from the control region-only fit (with event yields in Chpt. 6.9.1) is overlaid for comparison. Corresponding figures of the control regions resulting from the control region-only fit are given in App. A.2.

The  $VH$  topology is dominated by the irreducible  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$  background, followed by lost lepton, and finally a small amount of QCD multijet. Especially with the  $b$ -jet/boosted object requirements and dijet signature coupled to a small mass window around the electroweak bosons, it is a percent-level background in the bins where it appears. The post-fit prediction agrees with the data well in most cases, demonstrating an adequate description and leaving little room for signal to be inflated. An exception is the 1V category in 2018 whose observation is much larger than in the preceding years. The observed limits for 2j0b and 2j1b also fall close to, if not slightly outside, the  $2\sigma$  bound of the expected limit in 2017 and 2018.

Fig. 6.21 showcases the expected and observed limits for the  $VH$  channel in each year and for the Run-2 combination. Limits for the individual categories can be found in Fig. A.12.

## 6.9. RESULTS

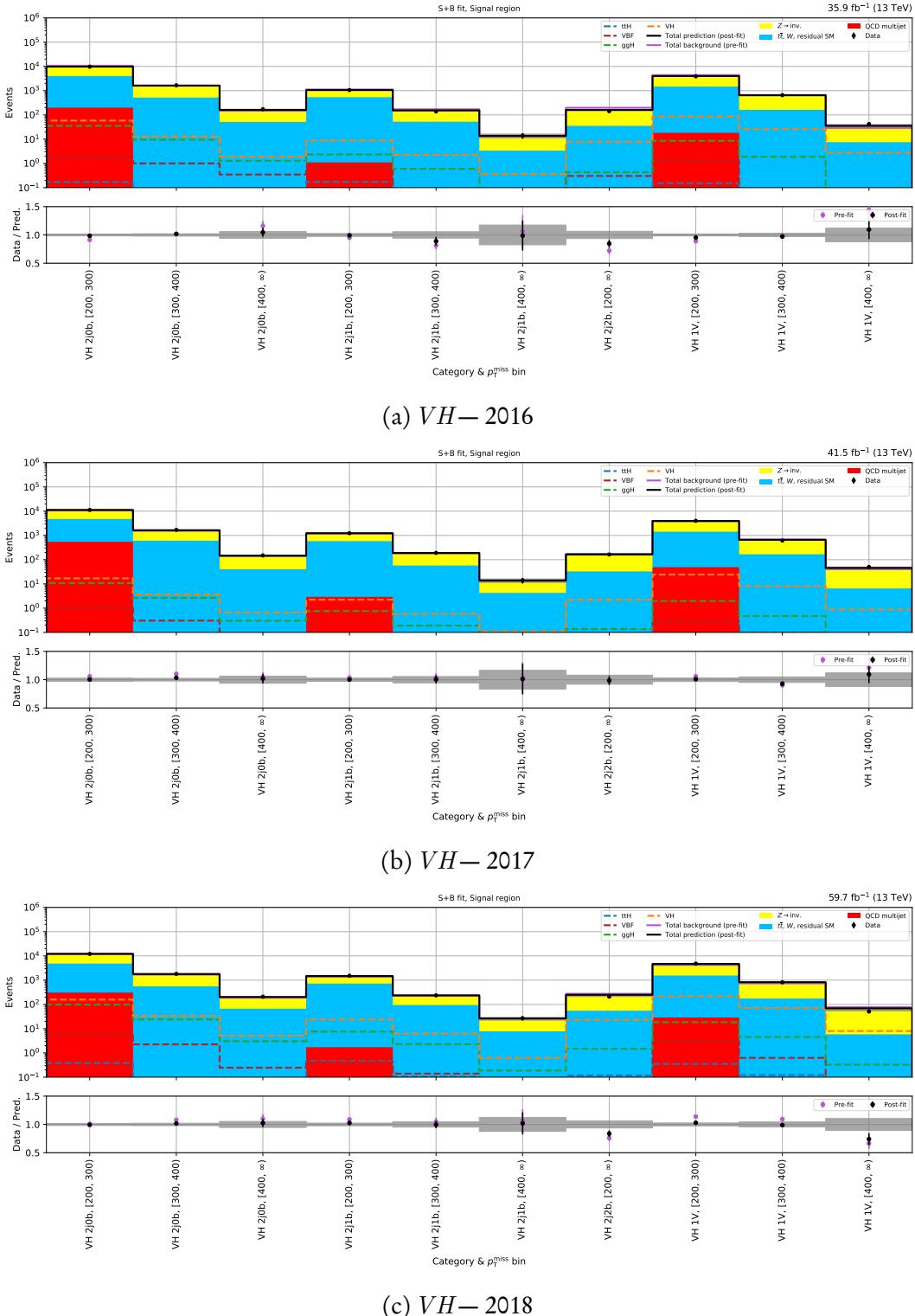


Figure 6.20: Pre-fit and post-fit yields in the signal region for every  $VH$  category and  $p_T^{\text{miss}}$  bin in each year of Run-2. “Pre-fit” refers to outcome of the control region-only prediction.

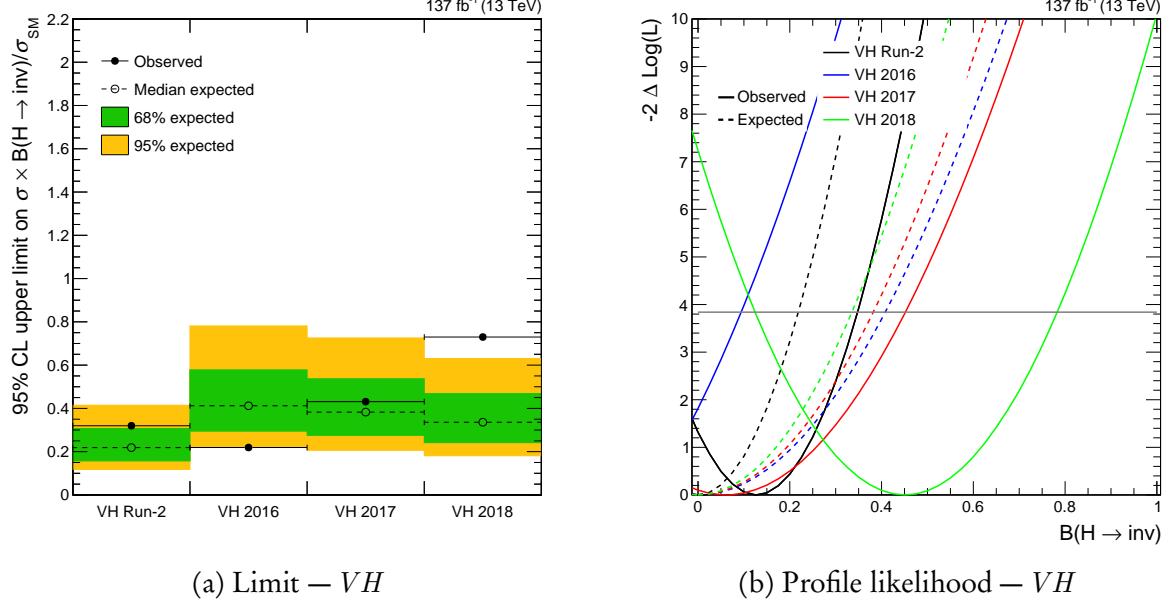


Figure 6.21: Observed and expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction  $\mathcal{B}(H \rightarrow \text{invisible})$  (left) and the corresponding profile likelihood scan (right) in the  $VH$  channel. The result from each data taking period is presented along with their combination.

The obtained limits demonstrate that the sensitivity of the  $VH$  topology to  $H \rightarrow \text{invisible}$  is second only to VBF. Fig. A.12 makes apparent that the sensitivity of  $VH$  within a given year, and overall, is driven by the 1V category. It is closely followed by 2j0b (rich in  $V \rightarrow q\bar{q}$ ) and 2j2b (targeting  $Z \rightarrow b\bar{b}$ ). 2j1b lags behind as it relies on a  $b$ -jet either being falsely tagged by the DEEPCSV algorithm or missed by it. Fluctuations between years average to yield a full Run-2 observed limit of 32 % and expected of 22 %—a very competitive result in the invisible Higgs landscape for a single production mode. Contributions to the overall limit from the boosted (1V) and resolved categories can be seen in Fig. A.13. While 1V contributes most to the expected limit, the resolved categories collectively drive the observed limit.

For comparisons to public results, there are no full Run-2 hadronic  $VH(H \rightarrow \text{invisible})$  searches or interpretations. Results from both ATLAS and CMS are, however, available using 2016 data. The latter is a mono- $V$  search comparable to the 1V category from this thesis. The ATLAS  $VH$  result found an observed limit of 83 % and expected of 58 % [2], while from CMS they were 50 % and 48 %, respectively [130]. From this thesis, an observed limit of 67 % and expected of 55 % is obtained for the 1V category. Sensitivity is better than that of ATLAS but worse than the CMS result. Differences in signal modelling are likely a factor. Older simulated signal have been found to use fixed factorisation scales at the  $W$  or  $H$  mass as opposed to the running scales in newer samples such as those used in this thesis; consequently,

a harder Higgs boson  $p_T$  (i.e.,  $p_T^{\text{miss}}$ ) spectrum is exhibited in the older samples. Dissimilarities such as these make direct comparisons to previous results more difficult.

The limits per year in Fig. 6.21 are inconsistent to some degree, with an observation slightly below the  $2\sigma$  boundary of the expected limit in 2016, correspondence in 2017, and an excess in 2018. In 2018, the observed limit is noticeably worse than in the preceding two. Inference from Figs. A.7 and A.8 suggests it is due the over-prediction of simulation in all of the leptonic control regions of the 1V category, consequently scaling down the background in the signal region. The 2j0b observation is also just outside the  $2\sigma$  interval, further affecting the fit to the whole channel. Reassuringly, the expected limits per year are stable and as assumed (sensitivity improving each year, following the trend in integrated luminosity). No issues are found with the likelihood scans, either.

### 6.9.5 Combined results

The combined upper limit on  $\sigma \times \mathcal{B}(H \rightarrow \text{invisible})/\sigma_{\text{SM}}$  from combining the  $t\bar{t}H$  and  $VH$  channels for the full Run-2 dataset is illustrated in Fig. 6.22. Profile likelihood scans are presented opposite the limits.

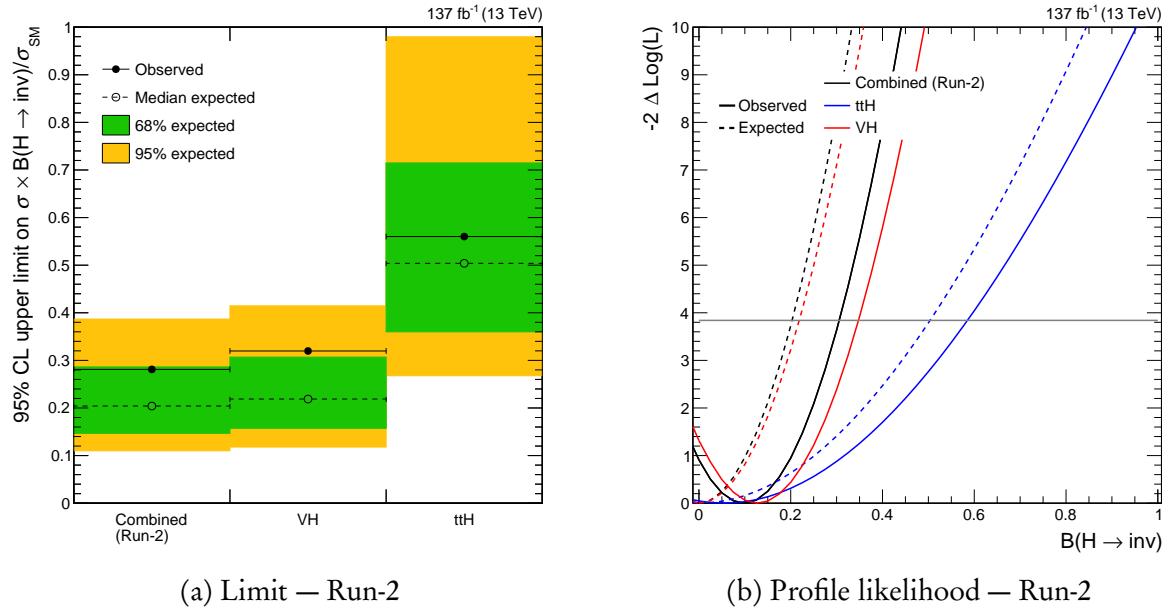


Figure 6.22: Observed and expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction  $\mathcal{B}(H \rightarrow \text{invisible})$  (left) and the corresponding profile likelihood scan (right), for both the individual categories, as well as the combination of them, for the full Run-2 dataset. The standard model Higgs boson with its associated mass and production cross section are assumed.

The  $VH$  channel predominantly drives the combined observed and expected sensitivity, with upper limits of 28 % and 20 %, respectively. Nevertheless, the  $t\bar{t}H$  category does constrain the combination and reduces the uncertainty bands around the expected limit. Both results are competitive with existing searches, with comparable results with the 2016 dataset to those published by both CMS and ATLAS. No full Run-2 combination exists for the  $VH$  channel, so the results in this thesis set the benchmark for future endeavours. The  $t\bar{t}H$  result with the full Run-2 dataset is stronger than that of ATLAS, also setting the standard for later searches to improve upon. Expected limits evolve naturally over the years, consistent with the integrated luminosity. A summary of the observed and median expected limits for both the  $t\bar{t}H$  and  $VH$  channels, together with their combination, is given in Tab. 6.15.

Dataset	$t\bar{t}H$	$VH$	Combined
2016	88 % (obs.)	22 % (obs.)	22 % (obs.)
	80 % (exp.)	41 % (exp.)	36 % (exp.)
2017	115 % (obs.)	43 % (obs.)	42 % (obs.)
	90 % (exp.)	38 % (exp.)	36 % (exp.)
2018	96 % (obs.)	73 % (obs.)	68 % (obs.)
	80 % (exp.)	34 % (exp.)	31 % (exp.)
Run-2	56 % (obs.)	32 % (obs.)	28 % (obs.)
	50 % (exp.)	22 % (exp.)	20 % (exp.)

Table 6.15: Observed and median expected upper limits on  $\mathcal{B}(H \rightarrow \text{invisible})$  at 95 % confidence level for each combination of channel and dataset analysed.

## 6.10 Interpretations in simplified dark matter scenarios

The results of the analysis fail to observe the Higgs boson to invisible state process, with an upper limit on the branching ratio much larger than the standard model prediction. However, limits on certain properties of dark matter may be set. One BSM interpretation is presented on a Higgs portal model where the SM Higgs acts as the mediator. The same procedure is followed as described in Ref. 139.

### 6.10.1 Higgs portal model with the standard model Higgs boson

An interpretation of the results in this thesis may be in terms of a simplified Higgs portal model—coupling the dark sector to the visible where the SM Higgs boson acts as the mediator

bridging them [77]. An effective field theory approach assumes that the invisible decays of the Higgs boson result in the pair production of dark matter particles  $\chi$ , with a rate constrained by the upper limit on  $\mathcal{B}(H \rightarrow \text{invisible})$ . If the dark matter mass  $m_\chi < m_H/2$ , on-shell production allows the translation of the  $H \rightarrow \text{invisible}$  width  $\Gamma_{\text{inv.}}$  into the spin-independent  $\chi$ -nucleon scattering cross section  $\sigma_{\text{SI}}$ . The interaction between dark and baryonic matter may be mediated by a Higgs boson, making direct detection experiments particularly sensitive to the recoil. The sensitivity of these experiments is typically parametrised by  $\sigma_{\text{SI}}$  as a function of  $m_\chi$ . Comparisons can therefore be made between direct detection and collider experiments.

The LHC has the ability to probe smaller dark matter masses  $\mathcal{O}(1\text{--}10\,\text{GeV})$  better than direct detection experiments. Light dark matter still results in  $\vec{p}_{\text{T}}^{\text{miss}}$  from a collision, but does not transfer much momentum when scattering against a heavy nucleus such as xenon, resulting in a weaker signature.

In this Higgs portal model, two cases are considered: the dark matter candidate is a scalar, or a Majorana fermion. The branching ratio is  $\mathcal{B}(H \rightarrow \text{invisible}) = \Gamma_{\text{inv.}}/(\Gamma_{\text{SM}} + \Gamma_{\text{inv.}})$ , where it is assumed assume  $\Gamma_{\text{SM}} = 4.07\,\text{MeV}$  [92]. From Ref. 77,  $\Gamma_{\text{inv.}}$  can be calculated for scalar  $\chi_S$  and fermion dark matter  $\chi_f$  as

$$(6.26) \quad \begin{aligned} \Gamma_{H \rightarrow \chi_S \chi_S}^{\text{inv.}} &= \frac{\lambda_{HSS}^2 v^2 \beta_S}{64\pi m_H}, \\ \Gamma_{H \rightarrow \chi_f \chi_f}^{\text{inv.}} &= \frac{\lambda_{Hff}^2 v^2 m_H \beta_f^3}{32\pi \Lambda^2} \end{aligned}$$

where  $\lambda$  is the coupling (scaled by  $\Lambda$  in the case of fermions),  $v$  is the vacuum expectation value of the SM Higgs field, and  $\beta_i = \sqrt{1 - 4m_{\chi_i}^2/m_H^2}$ . The masses of the dark matter particles  $m_\chi$  are the physical masses after electroweak symmetry breaking with this new Higgs field. The spin-independent  $\chi$ -nucleon scattering cross sections are

$$(6.27) \quad \begin{aligned} \sigma_{\chi_S - N}^{\text{SI}} &= \frac{\lambda_{HSS}^2}{16\pi m_H^4} \frac{m_N^4 f_N^2}{(m_{\chi_S} + m_N)^2}, \\ \sigma_{\chi_f - N}^{\text{SI}} &= \frac{\lambda_{Hff}^2}{4\pi \Lambda^2 m_H^4} \frac{m_N^4 m_{\chi_f}^2 f_N^2}{(m_{\chi_f} + m_N)^2} \end{aligned}$$

where  $f_N$  parametrises the Higgs-nucleon coupling. The value  $f_N = 0.308 \pm 0.018$  is taken from Ref. 96, recommended over that proposed in Ref. 77. Limits on these cross sections as a function of dark matter mass are displayed in Fig. 6.23, computed at 90 % confidence level to compare with direct detection experiments, whose latest results are also presented.

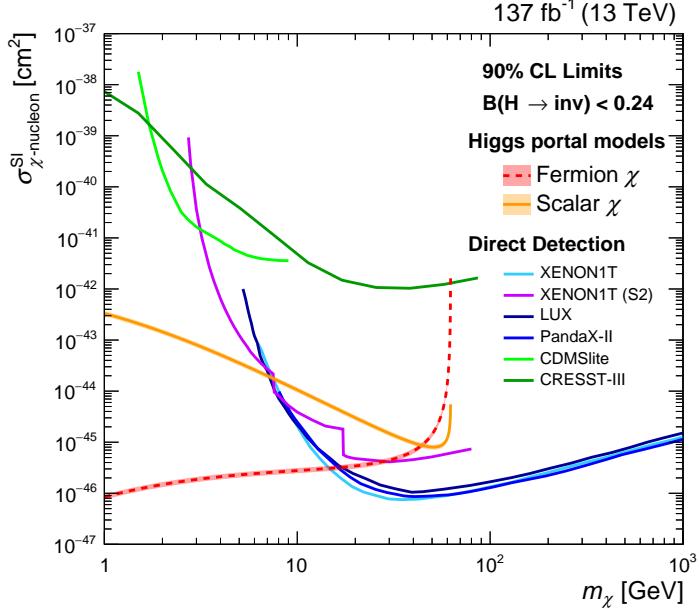


Figure 6.23: 90 % confidence level upper limits on the spin-independent dark matter-nucleon scattering cross section in Higgs portal models, where the standard model Higgs boson decays into a pair of scalar (solid orange) or fermion (dashed red) dark matter particles. Comparisons to direct detection experiments are also provided: XENON1T [18] (additionally with the S2-only analysis [19]), LUX [12], PandaX-II [73], CDMSlite [9], and CRESST-III [6].

For  $\mathcal{B}(H \rightarrow \text{invisible}) < 0.28$  observed at a 95 % confidence level (corresponding to 0.24 at a 90 % confidence level), the Higgs portal models assuming scalar and fermion dark matter candidates set the lowest limits on  $\sigma_{\text{SI}}$  for  $m_\chi$  below 7 and 15 GeV, respectively.

## CONCLUSIONS

Satisfaction lies in the effort, not in the attainment.

— Mahatma Gandhi

Scientific research can often be imagined as searching for bright city lights at the end of a dimly-lit, winding road. One hopes to reach it quickly with minimal setbacks. Despite somewhat stumbling in the dark and following the faintest strands, and unexpected twists and turns, progress is made through effort and determination even if the destination seems still out of sight. Reflecting on the work carried out over the course of this PhD and documented in this thesis, the seemingly unscalable wall that is the nature of dark matter has been modestly chipped away.

### 7.1 $H \rightarrow \text{invisible}$

Constraining the Higgs boson to invisible state branching ratio is important in understanding whether it couples to dark matter at all, and if so how strongly. The measurement can also refute or restrict the viability of several dark matter models. From the analysis in Chpt. 6 utilising the full Run-2 dataset from CMS, a combined upper limit on  $\sigma \times \mathcal{B}(H \rightarrow \text{invisible})/\sigma_{\text{SM}}$  of 28 % was observed at 95 % confidence level, and 20 % expected in the background-only hypothesis. While still far above the standard model prediction of 0.1 %, significant improvements have been made with respect to those obtained from Run-1 or 2016-only data.

The results with the full Run-2 dataset for both the  $t\bar{t}H$  and  $VH$  channels set new, world-leading benchmarks for measuring the branching ratio of invisibly decaying Higgs bosons from these production modes. For  $t\bar{t}H$ , modest improvements are seen compared to the equivalent measurement from ATLAS [24], while large gains are achieved in comparison to the previous result by CMS [83]. An observed upper limit of 56 % and 50 % expected was achieved. Despite it being a more novel and challenging signature than some of the other production modes, this analysis has demonstrated that the sensitivity of the  $t\bar{t}H$  channel in hadronic final states is comparable to that of the cleaner leptonic channels and the other production modes of the Higgs boson.

For the  $VH$  mechanism, there is no other public result (at the time of writing) encompassing both resolved and boosted topologies with a comparable dataset. Leveraging the full Run-2 dataset from CMS culminates in notably stronger sensitivity over previous results that analysed 2016 or Run-1 data: 22 % and 32 % upper limits were expected and observed, respectively. In addition to a superior limit in the  $VH$  channel overall, the Run-2 limit from the 1V category—equivalent to a mono- $V$  phase space—sets a new precedent for sensitivity to the  $H \rightarrow$  invisible decay.

Further improvements from a full combination over all Higgs production modes are inevitable. Given VBF is the most sensitive channel—and leptonic final states such as Ref. 83 have been discussed little—a grand combination with the other full Run-2 analyses will provide a much stronger limit. If there is a significant enhancement to the SM value of  $\mathcal{B}(H \rightarrow \text{invisible})$  due to a coupling to dark matter, it may be observable at the HL-LHC. The volume of available data is expected to increase by more than ten times. Coupled with improvements to the CMS detector for improved particle identification, noise mitigation, higher bandwidth links to read out more detector information, and more precise measurements of particle properties, a promising result is within the realm of possibility. New and improved machine learning algorithms, for example to tag resolved decays of the top quark, and more generally to discriminate signal from background, will also contribute. In summary, advancements in several areas can each incrementally boost the sensitivity to this process.

Failing any enhancement from a dark matter coupling, observing the SM rate of  $H \rightarrow 4\nu$  may require a precision-focused experiment. FCC- $ee$  and the ILC are strong candidates. Such “Higgs factories” would produce large quantities of Higgs bosons in  $e^+e^-$  collisions predominantly through the  $VH$  and VBF mechanisms.

## 7.2 Semi-visible jets

The search for semi-visible jets pushes into unexplored territory. This novel array of theoretical models—with final states overlooked by most dark matter searches—provides a rich and diverse playground with many avenues to consider. Only a small fraction of signal models for the  $s$ - and  $t$ -channel production mechanisms were analysed in Chpt. 5. However, their characteristics suggest unique signatures compared to the expected standard model background and can be exploited with variables as simple as the transverse mass of the dijet system.

The analysis focused on the  $s$ -channel mode, performed by collaborators in CMS, is nearing completion at the time of writing. A dedicated search for the  $t$ -channel mechanism has recently begun, as well as for semi-visible jets originating from the decays of boosted  $Z'$  bosons. Any of these could yield important results with the full Run-2 dataset. With the aid of Run-3 and HL-LHC data, and the other experimental improvements noted above, the boundaries will be further pushed. More advanced and versatile tagging algorithms could categorise a wide range of signal-like topologies while rejecting background to a larger degree. Higher energy future colliders, such as FCC-hh would widen the search window. Mediator masses up to several tens of TeV could be probed, with a commensurate increase to the dark quark mass range. Additional topologies arising from variations of the  $s$ - and  $t$ -channel modes, such as searches for displaced vertices/long-lived particles, open an even greater assortment of avenues to explore.





## ADDITIONAL MATERIAL FOR THE HIGGS TO INVISIBLE ANALYSIS

To avoid detracting too much from the predominant results in Chpt. 6, some figures and tables were omitted from the main text. Apps. A.1 and A.2 illustrate the distributions from the control region-only fit for  $t\bar{t}H$  and  $VH$  channels, respectively. Each category and  $p_T^{\text{miss}}$  bin is shown for the 2016 and 2018 datasets, as those for 2017 are displayed in Chpt. 6.7 of the main text. While the yields for individual backgrounds are post-fit, the total pre-fit MC is also displayed for comparisons to data as well as the post-fit prediction. Event counts in the signal region from the control region-only fit are tabulated in the main text (Chpt. 6.9.1).

The rate parameters from the control region-only fits to predict the electroweak backgrounds in the signal region are tabulated in App. A.3. Inputs to the QCD multijet prediction in the signal region are documented in App. A.4.

Apps. A.5 and A.6 comprise the  $p_T^{\text{miss}}$  distribution in the signal region from the fits to only the background. Corresponding event counts are presented in the main text under Chpt. 6.9.2.

App. A.7 includes limit plots broken down by category and data taking period for each of the analysis categories. The full Run-2 limits and likelihood scans for the boosted and resolved category groupings are also presented.

## A.1 Control region-only fits to the $t\bar{t}H$ categories

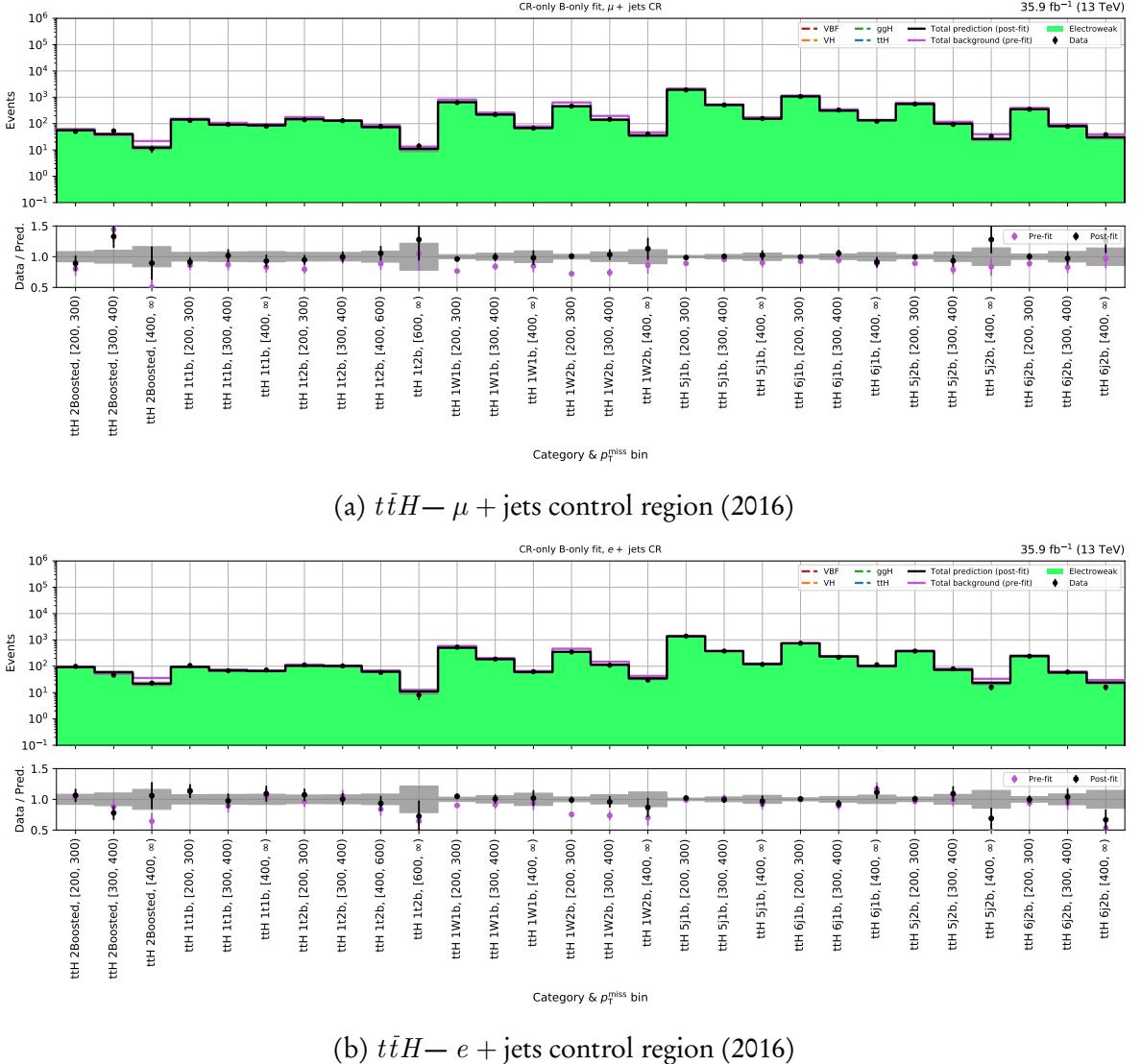


Figure A.1: Post-fit yields for each  $t\bar{t}H$  category and  $p_T^{\text{miss}}$  bin in the single lepton control regions for the 2016 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subplot.

## A.1. CONTROL REGION-ONLY FITS TO THE $t\bar{t}H$ CATEGORIES

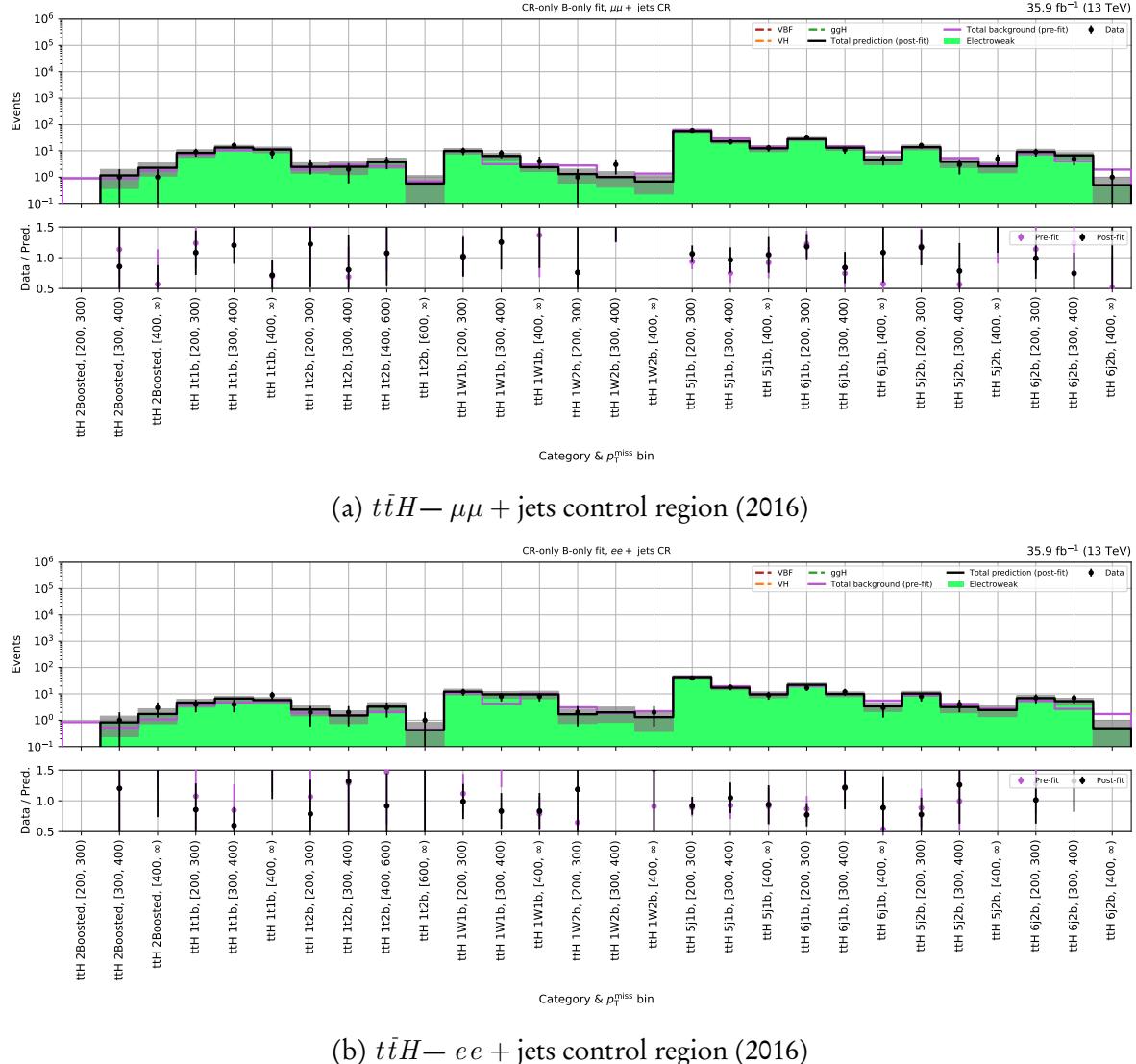


Figure A.2: Post-fit yields for each  $t\bar{t}H$  category and  $p_T^{\text{miss}}$  bin in the dilepton control regions for the 2016 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subfigure.

## APPENDIX A. ADDITIONAL MATERIAL FOR THE HIGGS TO INVISIBLE ANALYSIS

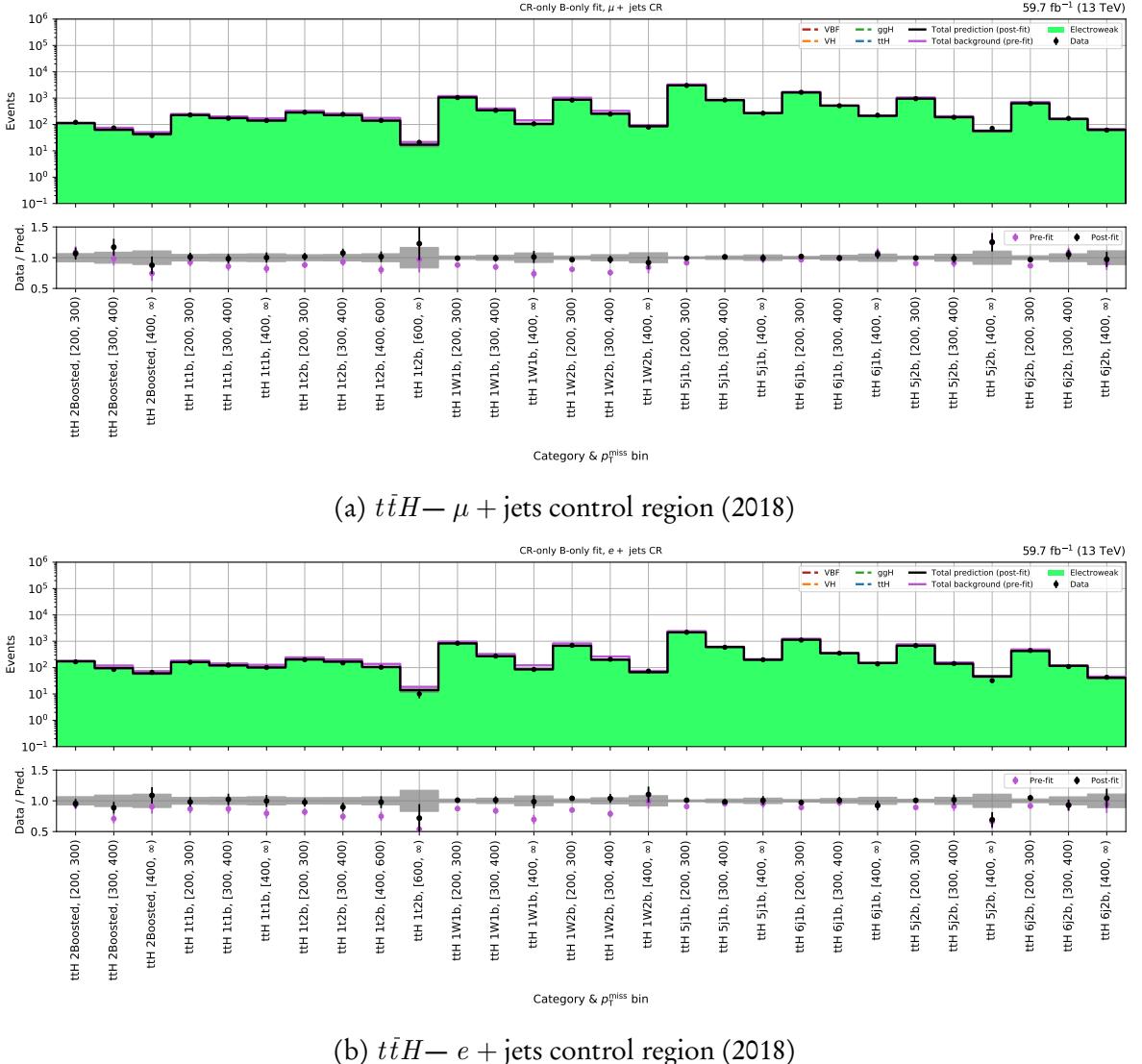


Figure A.3: Post-fit yields for each  $t\bar{t}H$  category and  $p_T^{\text{miss}}$  bin in the single lepton control regions for the 2018 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subfigure.

## A.1. CONTROL REGION-ONLY FITS TO THE $t\bar{t}H$ CATEGORIES

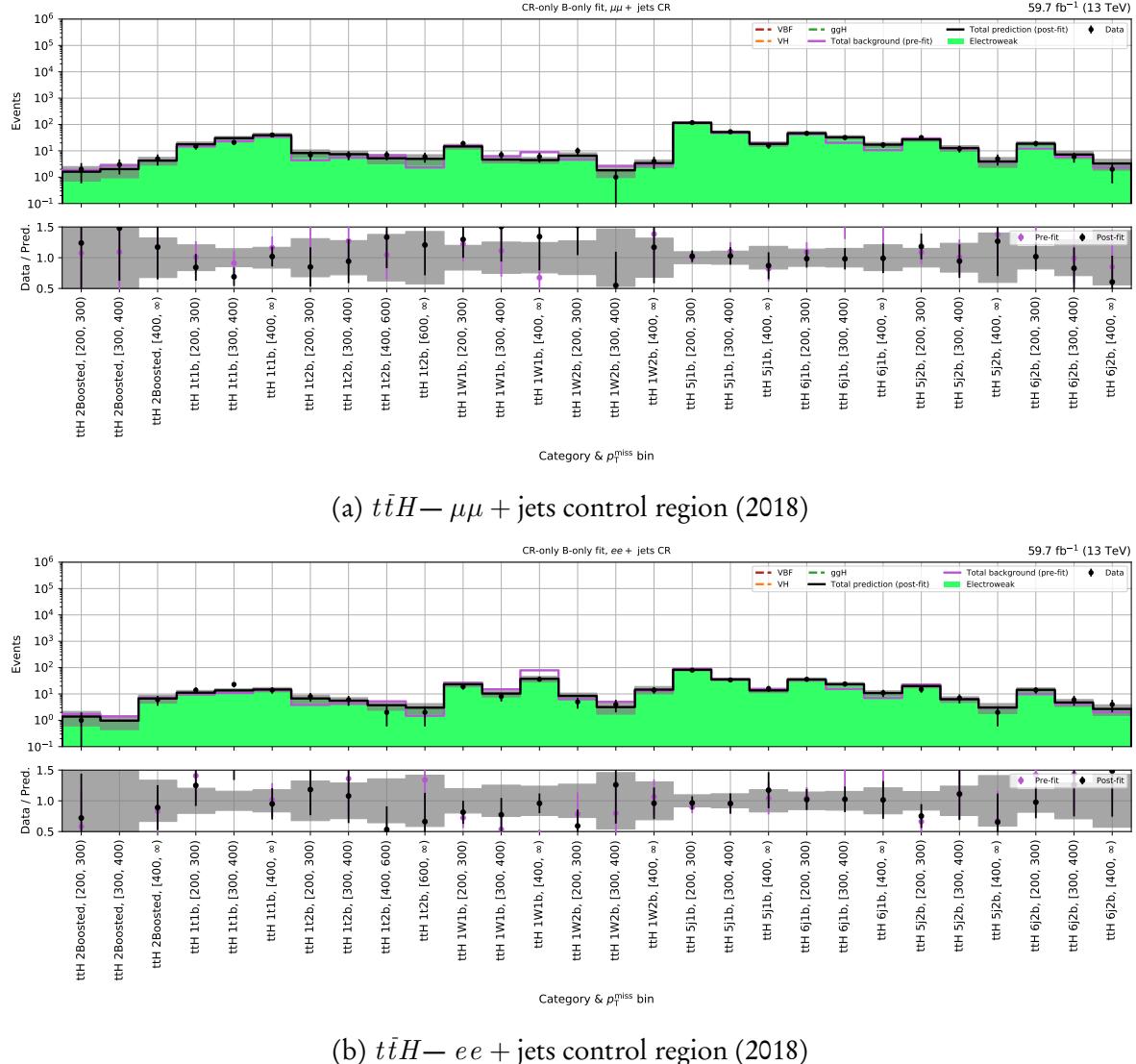


Figure A.4: Post-fit yields for each  $t\bar{t}H$  category and  $p_T^{\text{miss}}$  bin in the dilepton control regions for the 2018 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subplot.

## A.2 Control region-only fits to the $VH$ categories

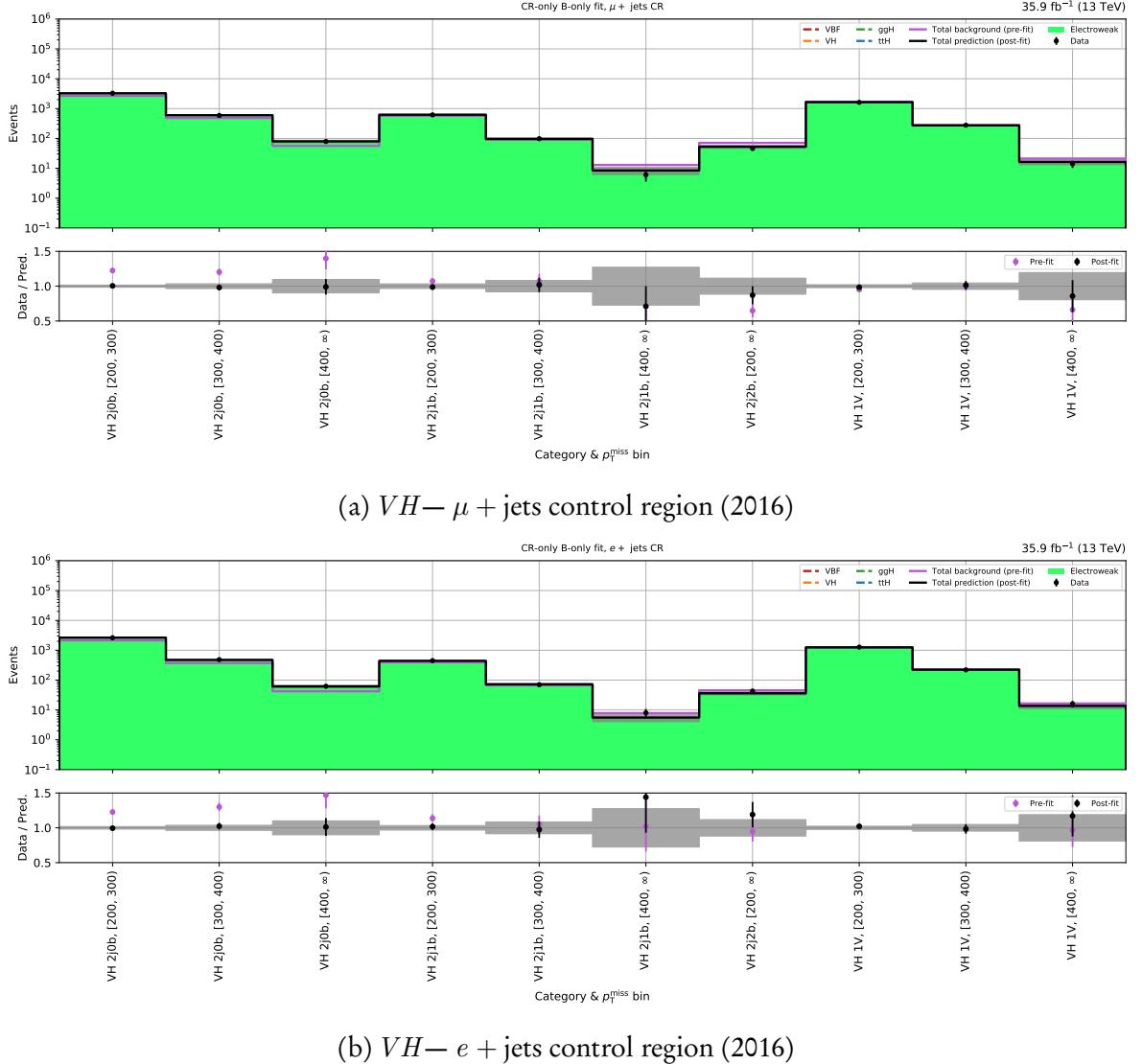


Figure A.5: Post-fit yields for each  $VH$  category and  $p_T^{\text{miss}}$  bin in the single lepton control regions for the 2016 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subfigure.

## A.2. CONTROL REGION-ONLY FITS TO THE $VH$ CATEGORIES

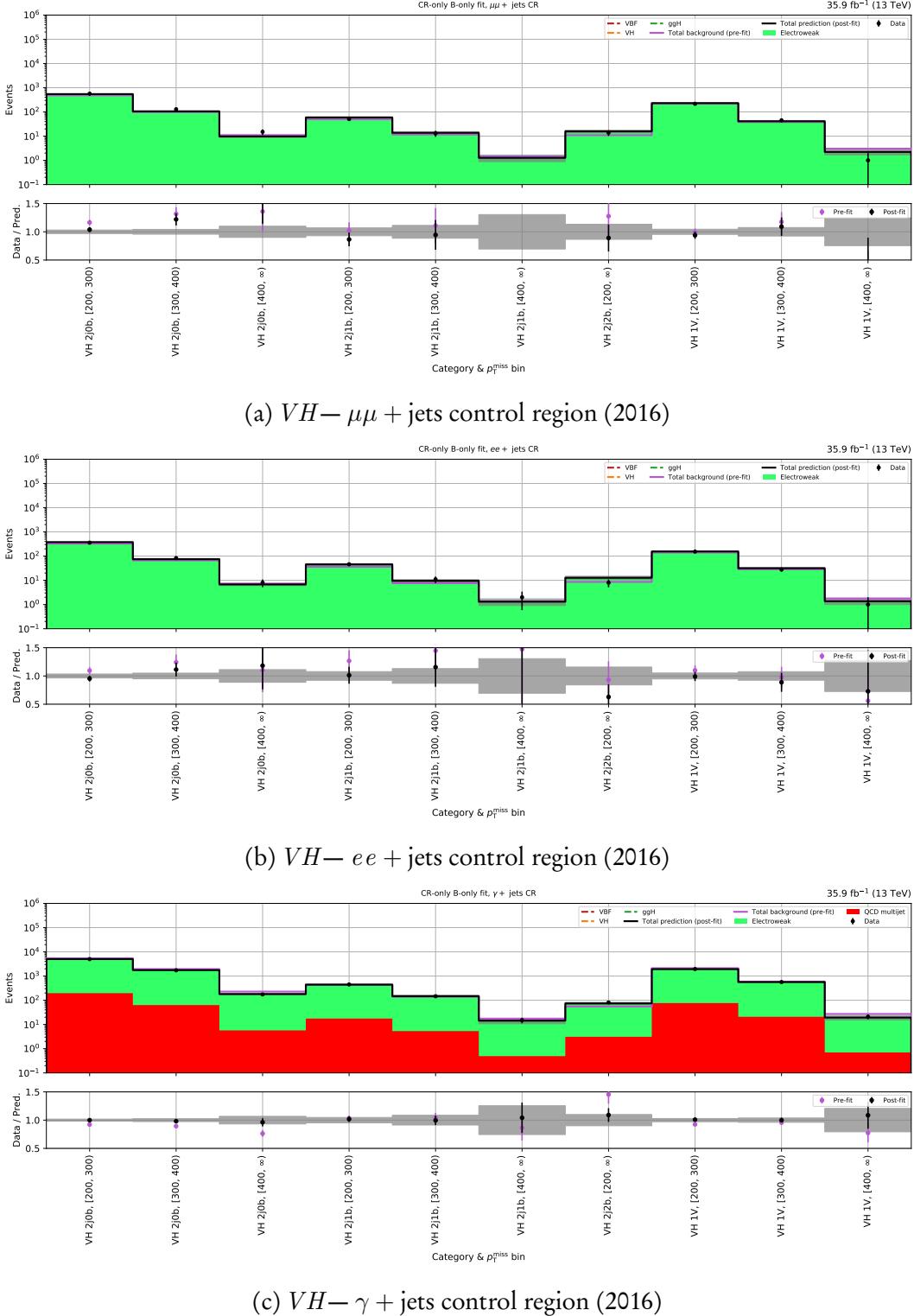


Figure A.6: Post-fit yields for each  $VH$  category and  $p_T^{\text{miss}}$  bin in the dilepton and photon control regions for the 2016 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subfigure.

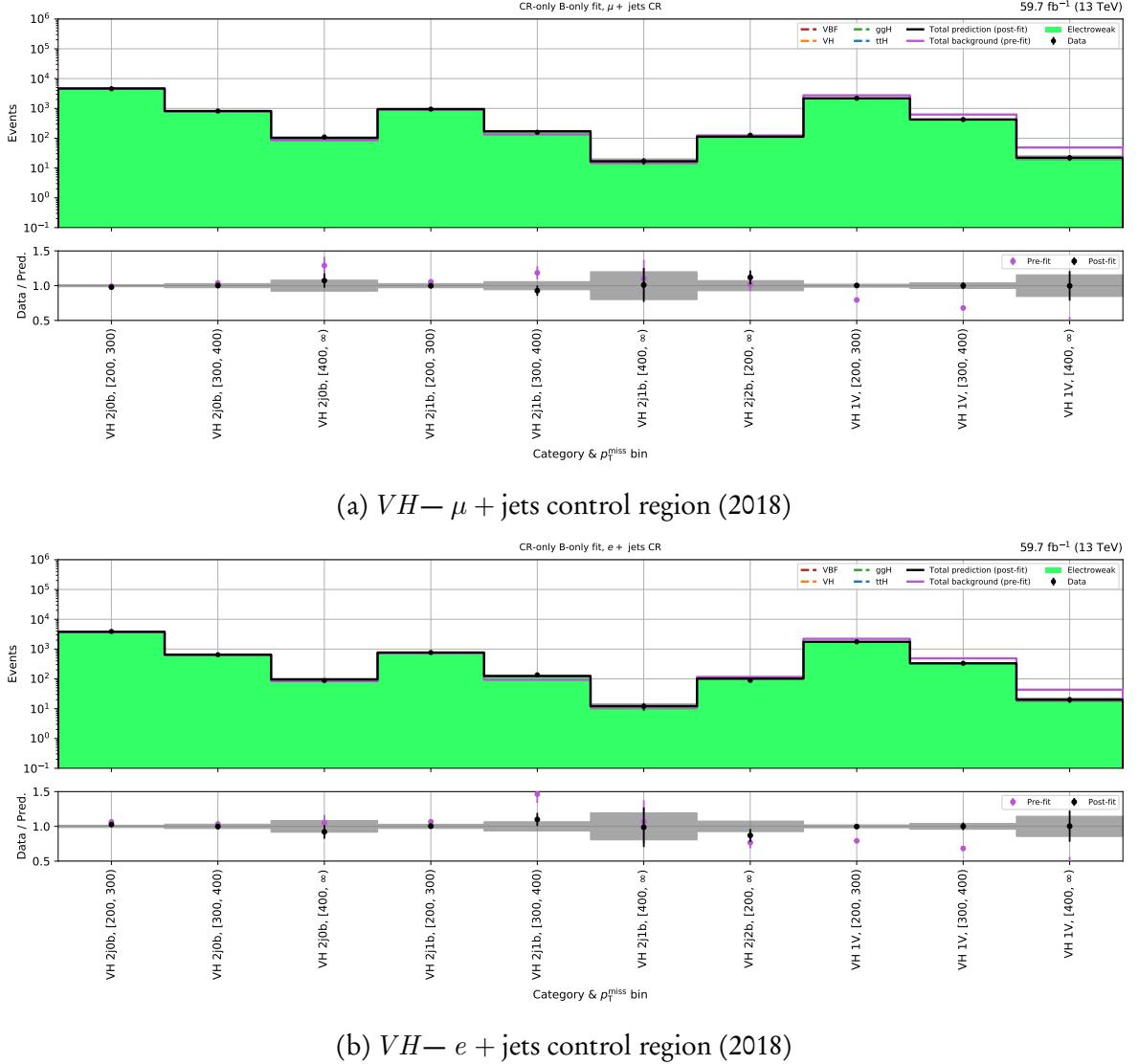


Figure A.7: Post-fit yields for each  $VH$  category and  $p_T^{\text{miss}}$  bin in the single lepton control regions for the 2018 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subfigure.

## A.2. CONTROL REGION-ONLY FITS TO THE $VH$ CATEGORIES

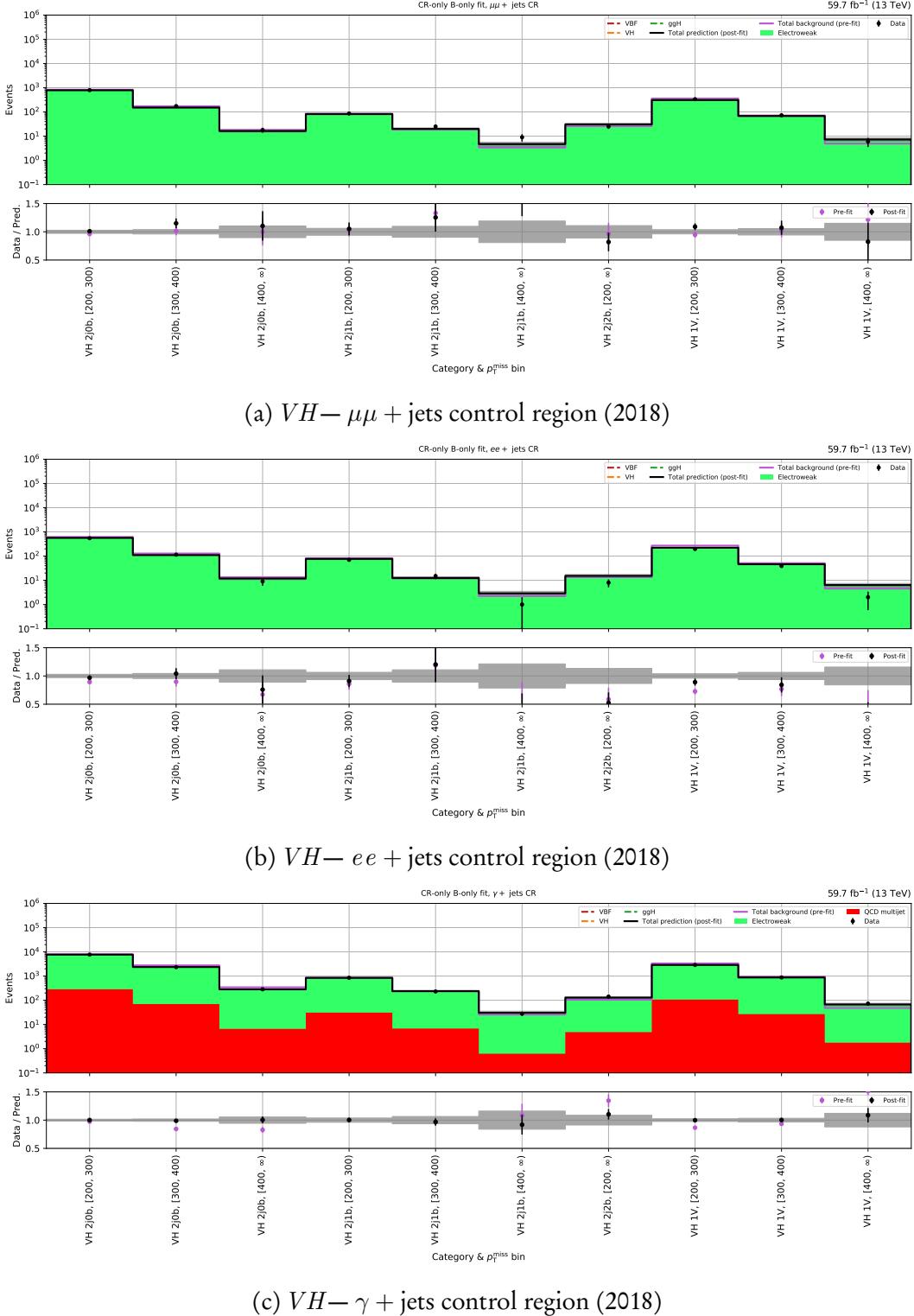


Figure A.8: Post-fit yields for each  $VH$  category and  $p_T^{\text{miss}}$  bin in the dilepton and photon control regions for the 2018 dataset. The total background pre-fit and post-fit is compared to data in the lower panel of each subplot.

### A.3 Rate parameters from the control region-only fit

Category	$p_T^{\text{miss}}$	$\mu + \text{jets}$		$e + \text{jets}$		$a_{\ell_{\text{lost}}}$	$N_{\ell_{\text{lost}}}^{\text{pred.}}$
		MC	Data	MC	Data		
$t\bar{t}H$ 2Boosted	[200, 300)	62.2	50	95.1	99	0.89	7.37
	[300, 400)	36.7	53	52.3	46	1.05	2.97
	[400, $\infty$ )	21.7	11	35.7	23	0.56	0.61
$t\bar{t}H$ 1t1b	[200, 300)	153	131	93.3	107	0.93	33.7
	[300, 400)	107	93	74.9	67	0.85	25.8
	[400, $\infty$ )	96.3	80	68.3	72	0.88	18.9
$t\bar{t}H$ 1t2b	[200, 300)	176	140	116	112	0.83	44.7
	[300, 400)	133	129	98.1	103	0.96	41.4
	[400, 600)	88.1	78	69.1	58	0.83	17.4
	[600, $\infty$ )	13.3	14	12.4	8	0.80	2.13
$t\bar{t}H$ 1W1b	[200, 300)	815	625	588	529	0.80	375
	[300, 400)	260	219	204	187	0.84	83.2
	[400, $\infty$ )	78.9	67	65.6	62	0.84	15.1
$t\bar{t}H$ 1W2b	[200, 300)	636	460	461	348	0.71	266
	[300, 400)	196	145	148	109	0.71	50.7
	[400, $\infty$ )	46.5	40	42.8	30	0.73	7.73
$t\bar{t}H$ 5j1b	[200, 300)	2137	1903	1416	1398	0.91	1522
	[300, 400)	536	512	369	377	0.95	285
	[400, $\infty$ )	173	156	127	117	0.85	55.1
$t\bar{t}H$ 6j1b	[200, 300)	1159	1073	750	748	0.94	900
	[300, 400)	349	328	245	218	0.89	183
	[400, $\infty$ )	137	122	97.6	114	0.94	51.5
$t\bar{t}H$ 5j2b	[200, 300)	620	553	388	376	0.89	406
	[300, 400)	118	93	80.3	80	0.84	47.3
	[400, $\infty$ )	39.5	33	33.3	16	0.63	9.38
$t\bar{t}H$ 6j2b	[200, 300)	394	350	256	241	0.88	291
	[300, 400)	94.2	78	63.0	60	0.84	45.3
	[400, $\infty$ )	39.2	38	29.9	16	0.73	12.4
$VH$ 2j0b	[200, 300)	2670	3271	2140	2625	1.07	3914
	[300, 400)	490	589	373	485	1.11	513
	[400, $\infty$ )	56.5	79	42.2	62	1.32	48.8
$VH$ 2j1b	[200, 300)	573	615	395	449	1.04	545
	[300, 400)	91.7	98	66.7	70	1.00	56.1
	[400, $\infty$ )	12.9	6	7.85	8	0.64	3.24
$VH$ 2j2b	[200, $\infty$ )	70.9	46	45.3	43	0.72	38.4
	[200, 300)	1698	1616	1248	1273	0.89	1525
	[300, 400)	278	277	223	220	0.93	163
	[400, $\infty$ )	21.2	14	16.50	16	0.76	6.92

Table A.1: Monte Carlo and data yields in the  $\mu + \text{jets}$  and  $e + \text{jets}$  control regions of the 2016 dataset, and corresponding rate parameter  $a_{\ell_{\text{lost}}}$  connecting them to the signal region, for each category and  $p_T^{\text{miss}}$  bin. The predicted lost lepton background in the signal region  $N_{\ell_{\text{lost}}}^{\text{pred.}}$  is also given.

## A.3. RATE PARAMETERS FROM THE CONTROL REGION-ONLY FIT

Category	$p_T^{\text{miss}}$	$\mu\mu + \text{jets}$		$ee + \text{jets}$		$\gamma + \text{jets}$		$a_{Z \rightarrow \nu\bar{\nu}}$	$N_{Z \rightarrow \nu\bar{\nu}}^{\text{pred.}}$
		MC	Data	MC	Data	MC	Data		
$t\bar{t}H$ 2Boosted	[200, 300)	0.90	0	0.85	0	—	—	0.00	0.00
	[300, 400)	0.88	1	0.53	1	—	—	1.31	1.26
	[400, $\infty$ )	1.76	1	1.07	3	—	—	1.24	1.46
$t\bar{t}H$ 1t1b	[200, 300)	7.28	9	3.72	4	—	—	1.19	6.34
	[300, 400)	10.3	16	4.73	4	—	—	1.28	6.61
	[400, $\infty$ )	11.4	8	5.23	9	—	—	0.88	7.36
$t\bar{t}H$ 1t2b	[200, 300)	1.97	3	1.87	2	—	—	1.27	3.12
	[300, 400)	2.91	2	1.55	2	—	—	0.85	2.68
	[400, 600)	2.53	4	2.04	3	—	—	1.35	4.62
	[600, $\infty$ )	0.69	0	0.45	1	—	—	0.70	1.10
$t\bar{t}H$ 1W1b	[200, 300)	9.72	10	10.7	12	—	—	1.06	46.2
	[300, 400)	3.13	8	4.24	8	—	—	2.03	34.9
	[400, $\infty$ )	2.93	4	10.1	8	—	—	0.74	8.23
$t\bar{t}H$ 1W2b	[200, 300)	2.77	1	3.09	2	—	—	0.50	4.00
	[300, 400)	1.02	3	1.87	0	—	—	0.96	3.66
	[400, $\infty$ )	1.36	0	2.20	2	—	—	0.46	1.07
$t\bar{t}H$ 5j1b	[200, 300)	64.1	60	45.1	40	—	—	0.96	473
	[300, 400)	29.3	22	19.5	18	—	—	0.80	157
	[400, $\infty$ )	14.1	13	9.86	9	—	—	0.82	92.1
$t\bar{t}H$ 6j1b	[200, 300)	26.9	33	19.6	17	—	—	1.12	232
	[300, 400)	14.7	11	9.79	12	—	—	0.91	91.8
	[400, $\infty$ )	8.72	5	5.58	3	—	—	0.49	34.6
$t\bar{t}H$ 5j2b	[200, 300)	13.5	16	9.05	8	—	—	1.08	111
	[300, 400)	5.31	3	4.02	4	—	—	0.73	29.0
	[400, $\infty$ )	3.05	5	2.75	0	—	—	0.77	16.1
$t\bar{t}H$ 6j2b	[200, 300)	7.90	9	5.48	7	—	—	1.21	73.7
	[300, 400)	4.03	5	2.69	7	—	—	1.69	49.4
	[400, $\infty$ )	1.95	1	1.72	0	—	—	0.24	4.36
$VH$ 2j0b	[200, 300)	486	566	328	359	5236	5013	0.93	6497
	[300, 400)	97.9	129	66.8	83	1871	1729	0.92	1125
	[400, $\infty$ )	11	15	7.25	8	222	174	0.79	99.5
$VH$ 2j1b	[200, 300)	49.9	51	36.3	46	418	450	1.02	571
	[300, 400)	11.7	13	7.61	11	136	147	1.06	115
	[400, $\infty$ )	1.43	0	1.36	2	16.9	15	0.83	9.91
$VH$ 2j2b	[200, $\infty$ )	11	14	8.59	8	52.6	81	1.35	160
$VH$ 1V	[200, 300)	219	217	140	154	2030	1950	0.92	2836
	[300, 400)	38.3	45	28.7	28	571	566	0.96	497
	[400, $\infty$ )	3.04	1	1.79	1	26.4	21	0.71	22.1

Table A.2: Monte Carlo and data yields in the  $\mu\mu + \text{jets}$ ,  $ee + \text{jets}$ , and  $\gamma + \text{jets}$  control regions of the 2016 dataset, and corresponding rate parameter  $a_{Z \rightarrow \nu\bar{\nu}}$  connecting them to the signal region, for each category and  $p_T^{\text{miss}}$  bin. The predicted invisible  $Z$  background in the signal region  $N_{Z \rightarrow \nu\bar{\nu}}^{\text{pred.}}$  is also given.

Category	$p_T^{\text{miss}}$	$\mu + \text{jets}$		$e + \text{jets}$		$a_{\ell_{\text{lost}}}$	$N_{\ell_{\text{lost}}}^{\text{pred.}}$
		MC	Data	MC	Data		
$t\bar{t}H$ 2Boosted	[200, 300)	59.1	72	93.8	99	1.13	7.11
	[300, 400)	38.8	46	61.5	74	1.20	2.48
	[400, $\infty$ )	26.0	36	38.3	41	1.17	1.17
$t\bar{t}H$ 1t1b	[200, 300)	155	174	111	117	1.12	45.8
	[300, 400)	119	128	90	95	1.08	38.0
	[400, $\infty$ )	99.9	105	82.1	74	0.98	21.2
$t\bar{t}H$ 1t2b	[200, 300)	178	168	124	129	0.99	52.90
	[300, 400)	138	138	105	125	1.08	45.20
	[400, 600)	88.1	94	69.4	72	1.05	20.3
	[600, $\infty$ )	11.7	13	8.94	7	0.94	1.70
$t\bar{t}H$ 1W1b	[200, 300)	731	683	562	520	0.97	385
	[300, 400)	242	248	188	186	1.04	94.9
	[400, $\infty$ )	80.9	83	64.7	64	1.04	15.8
$t\bar{t}H$ 1W2b	[200, 300)	592	534	431	360	0.90	288
	[300, 400)	176	141	137	112	0.82	53.0
	[400, $\infty$ )	50.7	28	40.7	36	0.72	6.73
$t\bar{t}H$ 5j1b	[200, 300)	2143	2317	1410	1611	1.12	1808
	[300, 400)	536	550	377	404	1.05	289
	[400, $\infty$ )	166	184	125	112	1.02	55.2
$t\bar{t}H$ 6j1b	[200, 300)	1088	1205	738	804	1.13	997
	[300, 400)	320	357	232	262	1.14	202
	[400, $\infty$ )	121	153	85.2	104	1.26	54.0
$t\bar{t}H$ 5j2b	[200, 300)	592	690	411	423	1.11	511
	[300, 400)	111	99	88.2	73	0.85	45.6
	[400, $\infty$ )	34.4	22	27.7	34	0.89	8.78
$t\bar{t}H$ 6j2b	[200, 300)	381	386	264	290	1.05	335
	[300, 400)	86.5	99	61.6	59	1.06	51.2
	[400, $\infty$ )	32.1	16	24.2	21	0.65	6.86
$VH$ 2j0b	[200, 300)	3269	3687	2352	2894	1.21	4152
	[300, 400)	530	653	411	529	1.28	574
	[400, $\infty$ )	55.9	70	53.8	60	1.19	39.0
$VH$ 2j1b	[200, 300)	593	706	452	517	1.18	569
	[300, 400)	96.7	96	70.0	77	1.03	56.6
	[400, $\infty$ )	8.75	12	8.05	6	1.03	4.27
$VH$ 2j2b	[200, $\infty$ )	71.7	57	57.2	46	0.79	32.7
$VH$ 1V	[200, 300)	1989	1680	1392	1157	0.88	1349
	[300, 400)	420	322	327	252	0.78	169
	[400, $\infty$ )	35.2	24	28.1	20	0.70	6.34

Table A.3: Monte Carlo and data yields in the  $\mu + \text{jets}$  and  $e + \text{jets}$  control regions of the 2017 dataset, and corresponding rate parameter  $a_{\ell_{\text{lost}}}$  connecting them to the signal region, for each category and  $p_T^{\text{miss}}$  bin. The predicted lost lepton background in the signal region  $N_{\ell_{\text{lost}}}^{\text{pred.}}$  is also given.

## A.3. RATE PARAMETERS FROM THE CONTROL REGION-ONLY FIT

Category	$p_T^{\text{miss}}$	$\mu\mu + \text{jets}$		$ee + \text{jets}$		$\gamma + \text{jets}$		$a_{Z \rightarrow \nu\bar{\nu}}$	$N_{Z \rightarrow \nu\bar{\nu}}^{\text{pred.}}$
		MC	Data	MC	Data	MC	Data		
$t\bar{t}H$ 2Boosted	[200, 300)	1.09	1	1.38	2	—	—	1.34	2.28
	[300, 400)	1.26	2	1.42	3	—	—	1.83	2.23
	[400, $\infty$ )	2.16	6	3.34	3	—	—	1.61	1.40
$t\bar{t}H$ 1t1b	[200, 300)	6.77	9	4.64	5	—	—	1.21	5.53
	[300, 400)	13.0	26	5.65	7	—	—	1.73	5.96
	[400, $\infty$ )	21.2	38	8.08	6	—	—	1.42	8.79
$t\bar{t}H$ 1t2b	[200, 300)	2.64	2	2.11	2	—	—	0.87	2.28
	[300, 400)	3.24	6	1.75	4	—	—	2.01	4.12
	[400, 600)	3.42	6	2.43	4	—	—	1.65	4.75
	[600, $\infty$ )	1.03	4	0.38	3	—	—	4.63	4.34
$t\bar{t}H$ 1W1b	[200, 300)	8.43	13	11.4	12	—	—	1.31	32.5
	[300, 400)	2.83	5	7.71	5	—	—	0.99	13.7
	[400, $\infty$ )	5.39	5	37.9	26	—	—	0.70	5.54
$t\bar{t}H$ 1W2b	[200, 300)	2.35	2	4.55	4	—	—	0.89	8.06
	[300, 400)	1.05	1	2.36	3	—	—	1.22	4.20
	[400, $\infty$ )	1.64	4	6.47	6	—	—	1.23	1.74
$t\bar{t}H$ 5j1b	[200, 300)	71.0	83	44.3	46	—	—	1.15	502
	[300, 400)	25.2	34	18.8	32	—	—	1.52	244
	[400, $\infty$ )	11.1	16	8.08	5	—	—	1.07	82.3
$t\bar{t}H$ 6j1b	[200, 300)	25.0	37	17.4	29	—	—	1.60	250
	[300, 400)	12.0	21	8.57	12	—	—	1.62	110
	[400, $\infty$ )	5.94	5	4.70	6	—	—	1.05	40.1
$t\bar{t}H$ 5j2b	[200, 300)	15.5	19	12.4	15	—	—	1.27	122
	[300, 400)	5.85	4	3.93	3	—	—	0.72	24.9
	[400, $\infty$ )	1.94	1	1.40	4	—	—	1.46	22.7
$t\bar{t}H$ 6j2b	[200, 300)	7.57	11	5.97	5	—	—	1.25	57.4
	[300, 400)	3.20	4	2.07	3	—	—	1.31	27.7
	[400, $\infty$ )	1.02	1	0.98	0	—	—	0.48	4.62
$VH$ 2j0b	[200, 300)	621	680	428	435	5025	5662	1.04	6090
	[300, 400)	120	106	82.9	78	1771	1824	0.93	964
	[400, $\infty$ )	14.9	11	8.89	6	217	214	0.88	103
$VH$ 2j1b	[200, 300)	67.0	71	42.1	34	391	578	1.22	619
	[300, 400)	12.1	13	7.02	11	114	168	1.37	126
	[400, $\infty$ )	1.96	4	1.47	1	13.3	14	1.03	9.3
$VH$ 2j2b	[200, $\infty$ )	16.3	14	11.1	6	50.3	69	1.03	132
$VH$ 1V	[200, 300)	265	223	184	146	1935	2048	0.97	2394
	[300, 400)	47.3	48	34.3	36	525	618	1.08	511
	[400, $\infty$ )	3.53	8	2.15	1	32.8	39	1.17	34.8

Table A.4: Monte Carlo and data yields in the  $\mu\mu + \text{jets}$ ,  $ee + \text{jets}$ , and  $\gamma + \text{jets}$  control regions of the 2017 dataset, and corresponding rate parameter  $a_{Z \rightarrow \nu\bar{\nu}}$  connecting them to the signal region, for each category and  $p_T^{\text{miss}}$  bin. The predicted invisible  $Z$  background in the signal region  $N_{Z \rightarrow \nu\bar{\nu}}^{\text{pred.}}$  is also given.

Category	$p_T^{\text{miss}}$	$\mu + \text{jets}$		$e + \text{jets}$		$a_{\ell_{\text{lost}}}$	$N_{\ell_{\text{lost}}}^{\text{pred.}}$
		MC	Data	MC	Data		
$t\bar{t}H$ 2Boosted	[200, 300)	111	121	177	166	1.01	11.2
	[300, 400)	74.9	74	120	85	0.82	3.48
	[400, $\infty$ )	51.1	38	73.1	66	0.81	1.59
$t\bar{t}H$ 1t1b	[200, 300)	249	230	183	159	0.93	52.4
	[300, 400)	200	172	143	124	0.85	42.6
	[400, $\infty$ )	172	141	127	101	0.80	26.1
$t\bar{t}H$ 1t2b	[200, 300)	329	290	243	199	0.87	72.2
	[300, 400)	261	243	203	151	0.83	58.8
	[400, 600)	177	142	138	103	0.77	27.5
	[600, $\infty$ )	21.6	21	18.6	10	0.73	2.27
$t\bar{t}H$ 1W1b	[200, 300)	1193	1051	958	837	0.93	523
	[300, 400)	404	343	327	274	0.86	114
	[400, $\infty$ )	143	106	122	85	0.73	17.7
$t\bar{t}H$ 1W2b	[200, 300)	1036	839	823	699	0.85	430
	[300, 400)	327	248	263	207	0.77	84.2
	[400, $\infty$ )	93.9	79	74.7	74	0.89	14.1
$t\bar{t}H$ 5j1b	[200, 300)	3300	3030	2414	2191	0.95	2043
	[300, 400)	838	851	620	591	1.02	390
	[400, $\infty$ )	275	268	208	199	0.96	82
$t\bar{t}H$ 6j1b	[200, 300)	1726	1667	1242	1108	0.98	1206
	[300, 400)	521	515	365	354	1.01	261
	[400, $\infty$ )	208	225	150	138	1.03	64.4
$t\bar{t}H$ 5j2b	[200, 300)	1053	953	763	681	0.90	610
	[300, 400)	205	187	157	143	0.88	82.4
	[400, $\infty$ )	56.4	71	48.3	32	0.95	16.7
$t\bar{t}H$ 6j2b	[200, 300)	711	618	488	447	0.88	449
	[300, 400)	159	171	118	109	0.99	86.6
	[400, $\infty$ )	66.4	61	45.5	43	0.89	19.1
$VH$ 2j0b	[200, 300)	4604	4563	3673	3907	1.19	4543
	[300, 400)	788	818	626	647	1.17	542
	[400, $\infty$ )	85.2	110	83.7	88	1.23	65
$VH$ 2j1b	[200, 300)	897	947	719	767	1.10	707
	[300, 400)	133	158	93.6	137	1.31	93
	[400, $\infty$ )	15.4	17	11.2	12	1.07	7.71
$VH$ 2j2b	[200, $\infty$ )	124	126	116	89	0.94	58.1
$VH$ 1V	[200, 300)	2747	2183	2208	1748	0.89	1481
	[300, 400)	622	423	488	333	0.73	173
	[400, $\infty$ )	48.6	22	43.4	20	0.47	6.08

Table A.5: Monte Carlo and data yields in the  $\mu + \text{jets}$  and  $e + \text{jets}$  control regions of the 2018 dataset, and corresponding rate parameter  $a_{\ell_{\text{lost}}}$  connecting them to the signal region, for each category and  $p_T^{\text{miss}}$  bin. The predicted lost lepton background in the signal region  $N_{\ell_{\text{lost}}}^{\text{pred.}}$  is also given.

## A.3. RATE PARAMETERS FROM THE CONTROL REGION-ONLY FIT

Category	$p_T^{\text{miss}}$	$\mu\mu + \text{jets}$		$ee + \text{jets}$		$\gamma + \text{jets}$		$a_{Z \rightarrow \nu\bar{\nu}}$	$N_{Z \rightarrow \nu\bar{\nu}}^{\text{pred.}}$
		MC	Data	MC	Data	MC	Data		
$t\bar{t}H$ 2Boosted	[200, 300)	1.86	2	1.72	1	—	—	0.87	1.97
	[300, 400)	2.75	3	1.39	0	—	—	0.77	1.25
	[400, $\infty$ )	4.25	5	7.23	6	—	—	0.85	1.58
$t\bar{t}H$ 1t1b	[200, 300)	14.9	15	9.94	14	—	—	1.23	7.50
	[300, 400)	23.0	21	10.8	23	—	—	1.30	8.43
	[400, $\infty$ )	34.4	40	13.8	14	—	—	1.09	11.2
$t\bar{t}H$ 1t2b	[200, 300)	4.41	7	3.85	8	—	—	1.81	7.31
	[300, 400)	5.50	7	4.41	6	—	—	1.31	4.80
	[400, 600)	6.69	7	5.23	2	—	—	0.71	3.67
	[600, $\infty$ )	2.33	6	1.49	2	—	—	2.03	3.03
$t\bar{t}H$ 1W1b	[200, 300)	15.6	19	26.2	19	—	—	1.01	47.8
	[300, 400)	6.29	7	15.0	8	—	—	0.80	15.5
	[400, $\infty$ )	8.89	6	78.9	36	—	—	0.50	6.14
$t\bar{t}H$ 1W2b	[200, 300)	4.62	10	6.37	5	—	—	1.37	21.1
	[300, 400)	2.69	1	5.02	4	—	—	0.60	5.88
	[400, $\infty$ )	2.89	4	13.2	14	—	—	1.16	3.62
$t\bar{t}H$ 5j1b	[200, 300)	118	118	89.1	80	—	—	1.09	672
	[300, 400)	48.1	53	35.9	34	—	—	1.08	283
	[400, $\infty$ )	19.5	16	15.4	16	—	—	0.92	109
$t\bar{t}H$ 6j1b	[200, 300)	42.2	46	34.4	36	—	—	1.21	291
	[300, 400)	20.2	32	15.4	24	—	—	1.72	184
	[400, $\infty$ )	10.6	17	7.06	11	—	—	1.75	109
$t\bar{t}H$ 5j2b	[200, 300)	29.4	32	22.6	15	—	—	1.00	151
	[300, 400)	11.9	12	6.36	7	—	—	1.08	58.9
	[400, $\infty$ )	3.63	5	2.95	2	—	—	1.11	26.4
$t\bar{t}H$ 6j2b	[200, 300)	12.1	19	9.83	14	—	—	1.46	119
	[300, 400)	6.05	6	4.17	6	—	—	1.15	41.3
	[400, $\infty$ )	2.35	2	2.03	4	—	—	1.38	23.8
$VH$ 2j0b	[200, 300)	826	794	615	549	7636	7707	1.13	7000
	[300, 400)	171	174	128	115	2695	2332	0.92	1149
	[400, $\infty$ )	18.0	18	13.4	9	337	284	0.88	124
$VH$ 2j1b	[200, 300)	83.5	86	81.4	70	817	847	1.06	685
	[300, 400)	18.8	25	12.6	15	231	231	1.08	132
	[400, $\infty$ )	3.40	9	2.24	1	25.2	28	1.25	17.6
$VH$ 2j2b	[200, $\infty$ )	25.9	25	13.6	8	101	142	1.22	216
$VH$ 1V	[200, 300)	356	337	271	197	3212	2876	1.06	2710
	[300, 400)	71.2	73	50.8	39	911	876	1.09	572
	[400, $\infty$ )	4.94	6	4.57	2	46.5	73	1.53	70.7

Table A.6: Monte Carlo and data yields in the  $\mu\mu + \text{jets}$ ,  $ee + \text{jets}$ , and  $\gamma + \text{jets}$  control regions of the 2018 dataset, and corresponding rate parameter  $a_{Z \rightarrow \nu\bar{\nu}}$  connecting them to the signal region, for each category and  $p_T^{\text{miss}}$  bin. The predicted invisible  $Z$  background in the signal region  $N_{Z \rightarrow \nu\bar{\nu}}^{\text{pred.}}$  is also given.

## A.4 Inputs to the QCD multijet prediction

Category	$p_T^{\text{miss}}$	$f_c$	$f_p$	$N_{\text{data}}$	$N_{\text{MC}}^{\text{non-QCD}}$	$\mathcal{T}_{\text{QCD}}$	$N_{\text{pred.}}^{\text{QCD}}$
$t\bar{t}H$ 2Boosted	[200, 300)		0.647				$1.03 \pm 1.01$
	[300, 400)	0.006	0.238				$0.38 \pm 0.62$
	[400, $\infty$ )		0.115				$0.18 \pm 0.43$
$t\bar{t}H$ 1t1b	[200, 300)		0.647				$7.11 \pm 2.67$
	[300, 400)	0.042	0.238				$2.62 \pm 1.62$
	[400, $\infty$ )		0.115				$1.27 \pm 1.13$
$t\bar{t}H$ 1t2b	[200, 300)		0.647				$9.42 \pm 3.07$
	[300, 400)	0.055	0.238				$3.47 \pm 1.86$
	[400, 600)		0.104				$1.51 \pm 1.23$
	[600, $\infty$ )		0.012				$0.17 \pm 0.41$
$t\bar{t}H$ 1W1b	[200, 300)		0.647				$21.6 \pm 4.64$
	[300, 400)	0.126	0.238				$7.94 \pm 2.82$
	[400, $\infty$ )		0.115				$3.84 \pm 1.96$
$t\bar{t}H$ 1W2b	[200, 300)		0.647	7200	3426	0.0698	$14.4 \pm 3.8$
	[300, 400)	0.084	0.238				$5.31 \pm 2.31$
	[400, $\infty$ )		0.115				$2.57 \pm 1.6$
$t\bar{t}H$ 5j1b	[200, 300)		0.647				$54.5 \pm 7.38$
	[300, 400)	0.319	0.238				$20.1 \pm 4.48$
	[400, $\infty$ )		0.115				$9.69 \pm 3.11$
$t\bar{t}H$ 6j1b	[200, 300)		0.647				$41.6 \pm 6.45$
	[300, 400)	0.243	0.238				$15.3 \pm 3.91$
	[400, $\infty$ )		0.115				$7.4 \pm 2.72$
$t\bar{t}H$ 5j2b	[200, 300)		0.647				$10.4 \pm 3.23$
	[300, 400)	0.061	0.238				$3.83 \pm 1.96$
	[400, $\infty$ )		0.115				$1.85 \pm 1.36$
$t\bar{t}H$ 6j2b	[200, 300)		0.647				$10.9 \pm 3.3$
	[300, 400)	0.064	0.238				$4.01 \pm 2.0$
	[400, $\infty$ )		0.115				$1.94 \pm 1.39$
$VH$ 2j0b	[200, 300)		0.794			0.00275	$259 \pm 16.1$
	[300, 400)		0.150			0	$0 \pm 0$
	[400, $\infty$ )		0.056			0	$0 \pm 0$
$VH$ 2j1b	[200, 300)		0.794			0.000015	$1.41 \pm 1.19$
	[300, 400)	1	0.150	188260	69691	0	$0 \pm 0$
	[400, $\infty$ )		0.056			0	$0 \pm 0$
$VH$ 2j2b	[200, $\infty$ )		0.794			0	$0 \pm 0$
	[200, 300)		0.794			0.00025	$23.5 \pm 4.85$
	[300, 400)		0.150			0	$0 \pm 0$
$VH$ 1V	[400, $\infty$ )		0.056			0	$0 \pm 0$

Table A.7: Components of the QCD multijet prediction in the signal region for the 2016 dataset. These include the category ( $f_c$ ) and  $p_T^{\text{miss}}$  fractions ( $f_p$ ) from the signal region for the  $t\bar{t}H$  channel (the former being unity in  $VH$ ), and the total event counts in data ( $N_{\text{data}}$ ) and non-QCD MC ( $N_{\text{MC}}^{\text{non-QCD}}$ ) from the sideband. The transfer factor ( $\mathcal{T}_{\text{QCD}}$ ) obtained from the sideband of a smeared multijet sample.

---

A.4. INPUTS TO THE QCD MULTIJET PREDICTION

---

Category	$p_T^{\text{miss}}$	$f_c$	$f_p$	$N_{\text{data}}$	$N_{\text{MC}}^{\text{non-QCD}}$	$\mathcal{T}_{\text{QCD}}$	$N_{\text{pred.}}^{\text{QCD}}$
$t\bar{t}H$ 2Boosted	[200, 300)		0.634				$1.6 \pm 1.27$
	[300, 400)	0.007	0.241				$0.61 \pm 0.78$
	[400, $\infty$ )		0.125				$0.32 \pm 0.56$
$t\bar{t}H$ 1t1b	[200, 300)		0.634				$10.2 \pm 3.2$
	[300, 400)	0.046	0.241				$3.88 \pm 1.97$
	[400, $\infty$ )		0.125				$2.02 \pm 1.42$
$t\bar{t}H$ 1t2b	[200, 300)		0.634				$12.8 \pm 3.58$
	[300, 400)		0.241				$4.86 \pm 2.2$
	[400, 600)	0.057	0.113				$2.27 \pm 1.51$
	[600, $\infty$ )		0.012				$0.25 \pm 0.5$
$t\bar{t}H$ 1W1b	[200, 300)		0.634				$26.8 \pm 5.18$
	[300, 400)	0.120	0.241				$10.2 \pm 3.19$
	[400, $\infty$ )		0.125				$5.3 \pm 2.3$
$t\bar{t}H$ 1W2b	[200, 300)		0.634				$18.1 \pm 4.26$
	[300, 400)	0.081	0.241	9371	4347	0.0698	$6.88 \pm 2.62$
	[400, $\infty$ )		0.125				$3.58 \pm 1.89$
$t\bar{t}H$ 5j1b	[200, 300)		0.634				$73.0 \pm 8.54$
	[300, 400)	0.327	0.241				$27.7 \pm 5.26$
	[400, $\infty$ )		0.125				$14.4 \pm 3.8$
$t\bar{t}H$ 6j1b	[200, 300)		0.634				$54.0 \pm 7.35$
	[300, 400)	0.242	0.241				$20.5 \pm 4.53$
	[400, $\infty$ )		0.125				$10.7 \pm 3.26$
$t\bar{t}H$ 5j2b	[200, 300)		0.634				$13.5 \pm 3.68$
	[300, 400)	0.061	0.241				$5.14 \pm 2.27$
	[400, $\infty$ )		0.125				$2.67 \pm 1.63$
$t\bar{t}H$ 6j2b	[200, 300)		0.634				$12.9 \pm 3.59$
	[300, 400)	0.058	0.241				$4.9 \pm 2.21$
	[400, $\infty$ )		0.125				$2.55 \pm 1.6$
$VH$ 2j0b	[200, 300)		0.770			0.00275	$379 \pm 19.5$
	[300, 400)		0.166			0	$0 \pm 0$
	[400, $\infty$ )		0.064			0	$0 \pm 0$
$VH$ 2j1b	[200, 300)		0.770			0.000015	$2.07 \pm 1.44$
	[300, 400)	1	0.166	288040	108702	0	$0 \pm 0$
	[400, $\infty$ )		0.064			0	$0 \pm 0$
$VH$ 2j2b	[200, $\infty$ )		0.770			0	$0 \pm 0$
$VH$ 1V	[200, 300)		0.770			0.00025	$34.5 \pm 5.88$
	[300, 400)		0.166			0	$0 \pm 0$
	[400, $\infty$ )		0.064			0	$0 \pm 0$

Table A.8: Components of the QCD multijet prediction in the signal region for the 2017 dataset. These include the category ( $f_c$ ) and  $p_T^{\text{miss}}$  fractions ( $f_p$ ) from the signal region for the  $t\bar{t}H$  channel (the former being unity in  $VH$ ), and the total event counts in data ( $N_{\text{data}}$ ) and non-QCD MC ( $N_{\text{MC}}^{\text{non-QCD}}$ ) from the sideband. The transfer factor ( $\mathcal{T}_{\text{QCD}}$ ) obtained from the sideband of a smeared multijet sample.

Category	$p_T^{\text{miss}}$	$f_c$	$f_p$	$N_{\text{data}}$	$N_{\text{MC}}^{\text{non-QCD}}$	$\mathcal{T}_{\text{QCD}}$	$N_{\text{pred.}}^{\text{QCD}}$
$t\bar{t}H$ 2Boosted	[200, 300)		0.628				$2.2 \pm 1.48$
	[300, 400)	0.007	0.244				$0.86 \pm 0.93$
	[400, $\infty$ )		0.127				$0.45 \pm 0.67$
$t\bar{t}H$ 1t1b	[200, 300)		0.628				$12.4 \pm 3.52$
	[300, 400)	0.041	0.244				$4.83 \pm 2.2$
	[400, $\infty$ )		0.127				$2.52 \pm 1.59$
$t\bar{t}H$ 1t2b	[200, 300)		0.628				$17.9 \pm 4.23$
	[300, 400)	0.059	0.244				$6.95 \pm 2.64$
	[400, 600)		0.115				$3.26 \pm 1.81$
	[600, $\infty$ )		0.013				$0.36 \pm 0.6$
$t\bar{t}H$ 1W1b	[200, 300)		0.628				$36.7 \pm 6.06$
	[300, 400)	0.126	0.244				$14.3 \pm 3.78$
	[400, $\infty$ )		0.127				$7.45 \pm 2.73$
$t\bar{t}H$ 1W2b	[200, 300)		0.628	12395	5525	0.0698	$30.5 \pm 5.52$
	[300, 400)	0.101	0.244				$11.9 \pm 3.44$
	[400, $\infty$ )		0.127				$6.18 \pm 2.49$
$t\bar{t}H$ 5j1b	[200, 300)		0.628				$93.0 \pm 9.65$
	[300, 400)	0.308	0.244				$36.2 \pm 6.01$
	[400, $\infty$ )		0.127				$18.9 \pm 4.34$
$t\bar{t}H$ 6j1b	[200, 300)		0.628				$68.2 \pm 8.26$
	[300, 400)	0.226	0.244				$26.5 \pm 5.15$
	[400, $\infty$ )		0.127				$13.8 \pm 3.72$
$t\bar{t}H$ 5j2b	[200, 300)		0.628				$20.5 \pm 4.53$
	[300, 400)	0.068	0.244				$7.99 \pm 2.83$
	[400, $\infty$ )		0.127				$4.17 \pm 2.04$
$t\bar{t}H$ 6j2b	[200, 300)		0.628				$20.6 \pm 4.54$
	[300, 400)	0.068	0.244				$8.02 \pm 2.83$
	[400, $\infty$ )		0.127				$4.19 \pm 2.05$
$VH$ 2j0b	[200, 300)		0.775			0.00275	$400 \pm 20$
	[300, 400)		0.159			0	$0 \pm 0$
	[400, $\infty$ )		0.066			0	$0 \pm 0$
$VH$ 2j1b	[200, 300)		0.775			0.000015	$2.19 \pm 1.48$
	[300, 400)	1	0.159	313805	109959	0	$0 \pm 0$
	[400, $\infty$ )		0.066			0	$0 \pm 0$
$VH$ 2j2b	[200, $\infty$ )		0.775			0	$0 \pm 0$
	[200, 300)		0.775			0.00025	$36.4 \pm 6.04$
	[300, 400)		0.159			0	$0 \pm 0$
$VH$ 1V	[400, $\infty$ )		0.066			0	$0 \pm 0$

Table A.9: Components of the QCD multijet prediction in the signal region for the 2018 dataset. These include the category ( $f_c$ ) and  $p_T^{\text{miss}}$  fractions ( $f_p$ ) from the signal region for the  $t\bar{t}H$  channel (the former being unity in  $VH$ ), and the total event counts in data ( $N_{\text{data}}$ ) and non-QCD MC ( $N_{\text{MC}}^{\text{non-QCD}}$ ) from the sideband. The transfer factor ( $\mathcal{T}_{\text{QCD}}$ ) obtained from the sideband of a smeared multijet sample.

## A.5 Background-only fits to the $t\bar{t}H$ categories

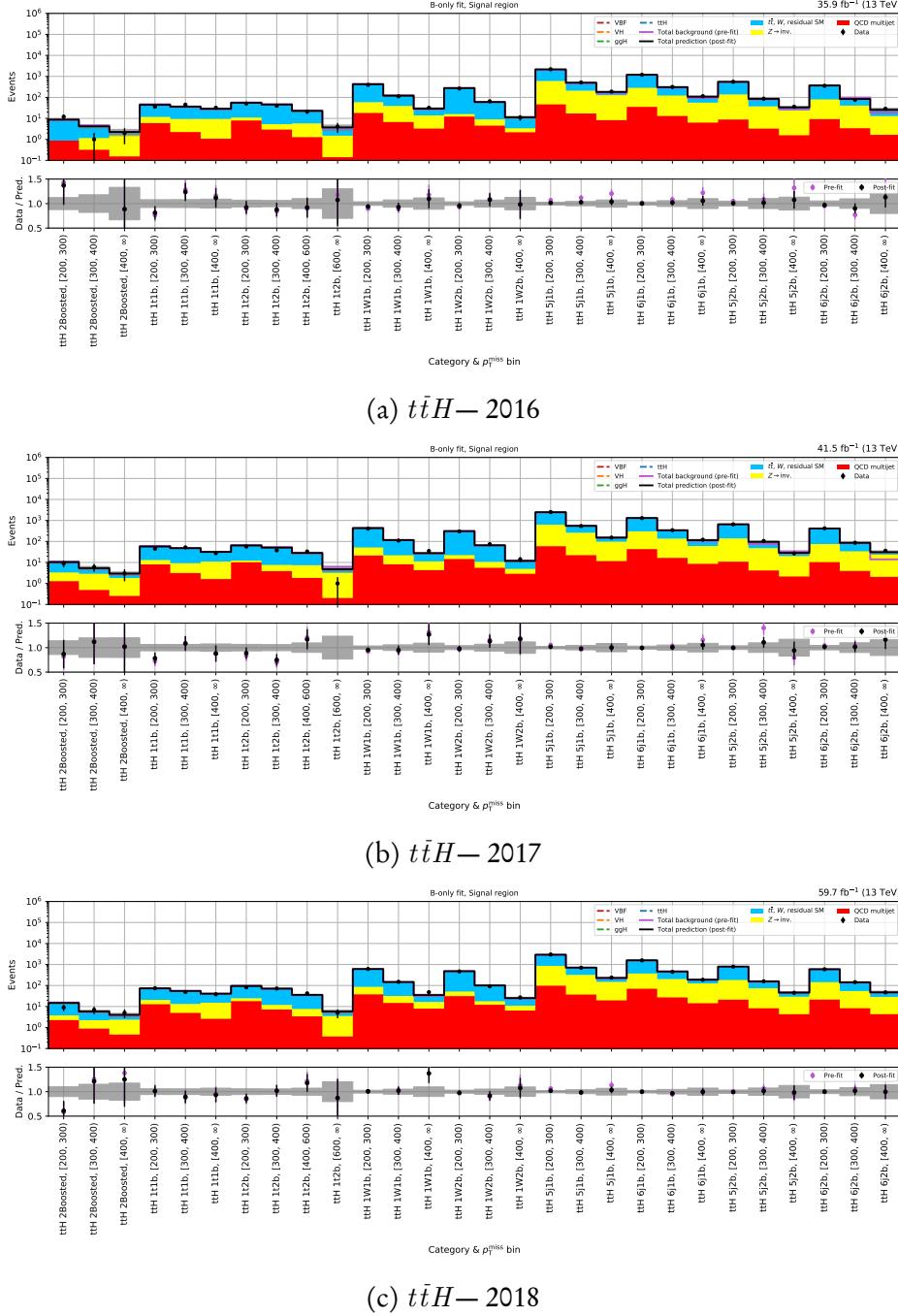


Figure A.9: Results of the background-only fit to the  $t\bar{t}H$  categories in each year of Run-2. “Pre-fit” refers to outcome of the control region-only prediction.

## A.6 Background-only fits to the $VH$ categories

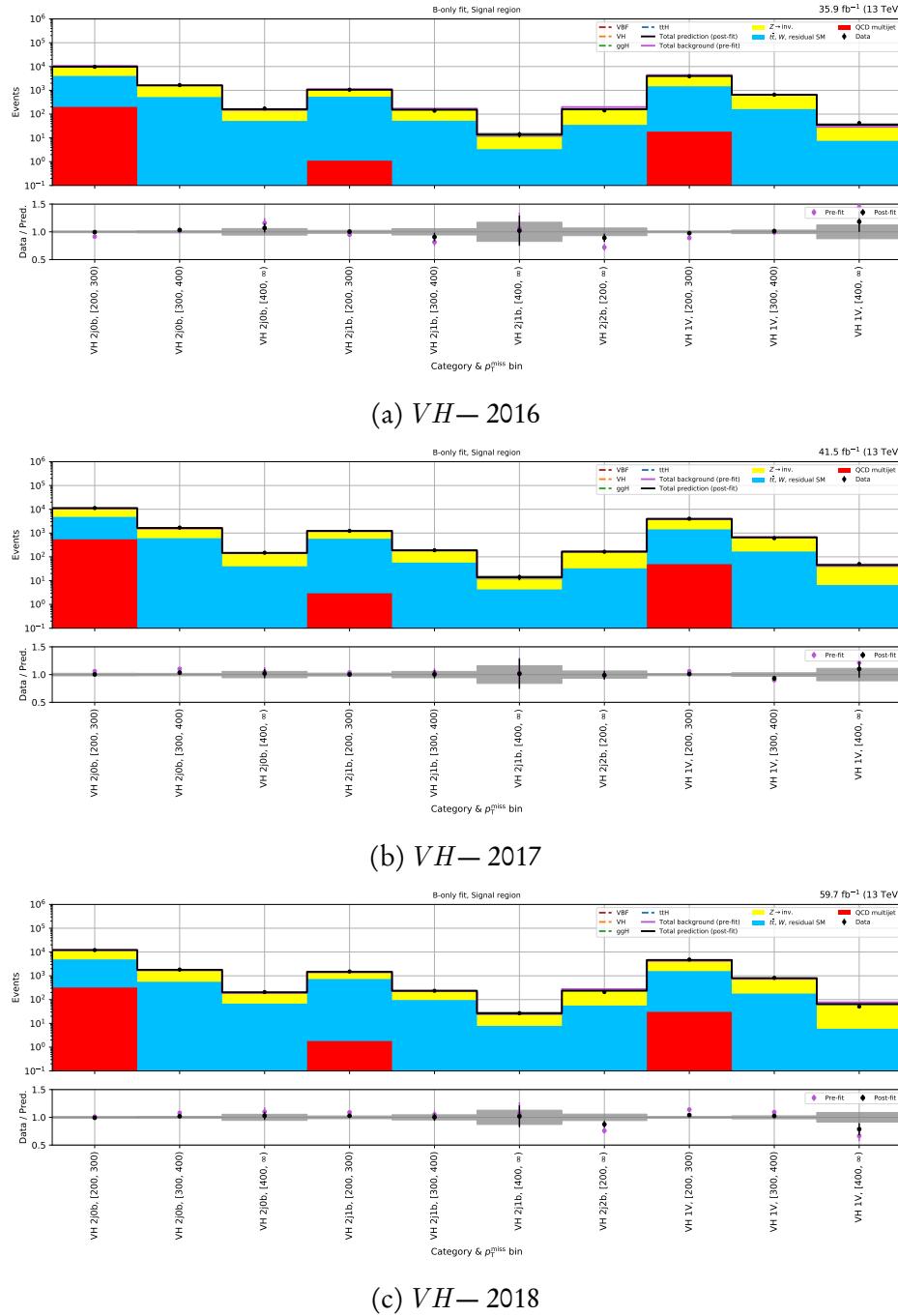


Figure A.10: Results of the background-only fit to the  $VH$  categories in each year of Run-2. “Pre-fit” refers to outcome of the control region-only prediction.

## A.7 Limits and likelihood scans for each category

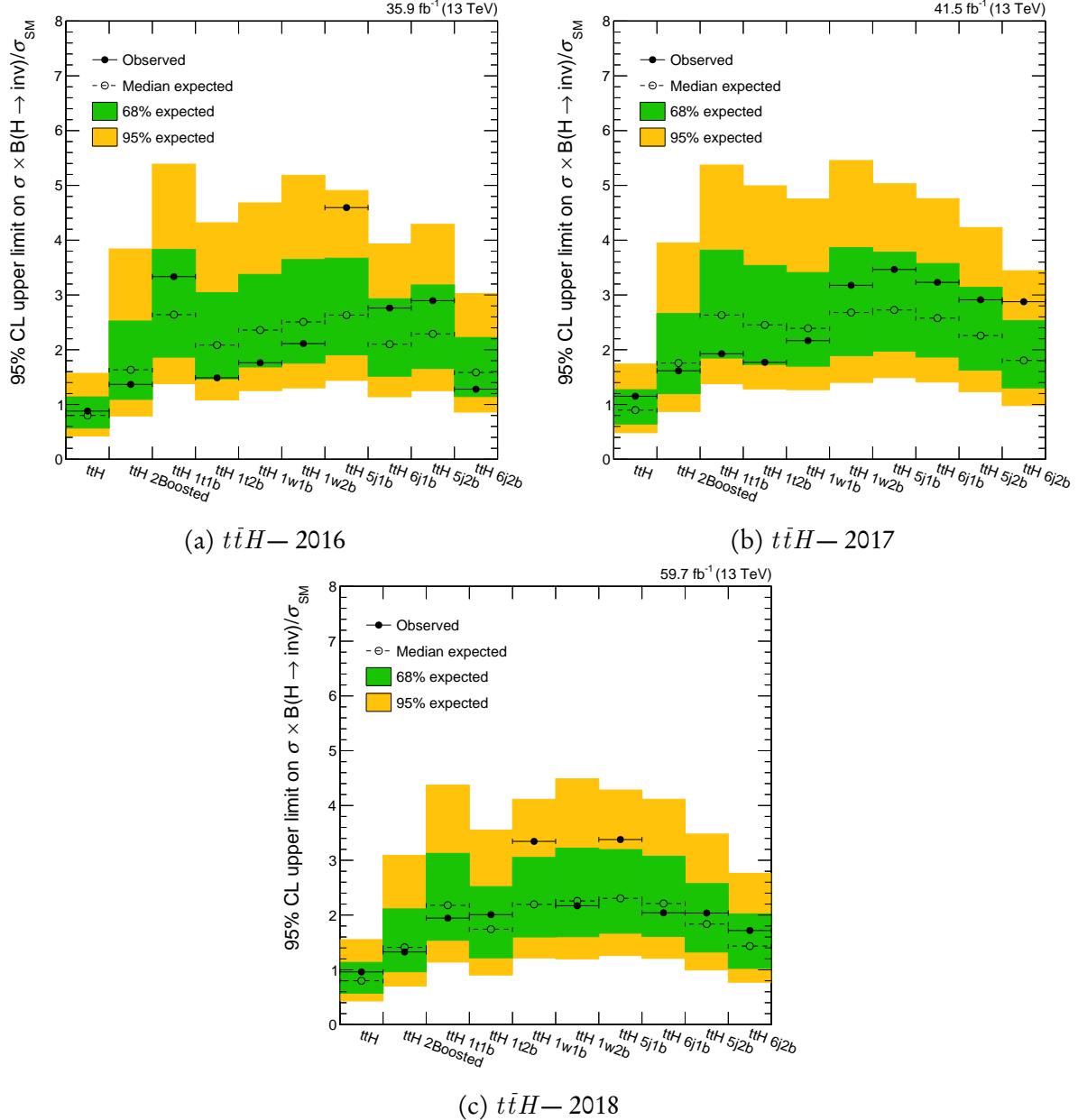


Figure A.11: Observed and expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the  $t\bar{t}H$  channel, for both the individual categories, and the combination of them, for each data-taking year in Run-2.

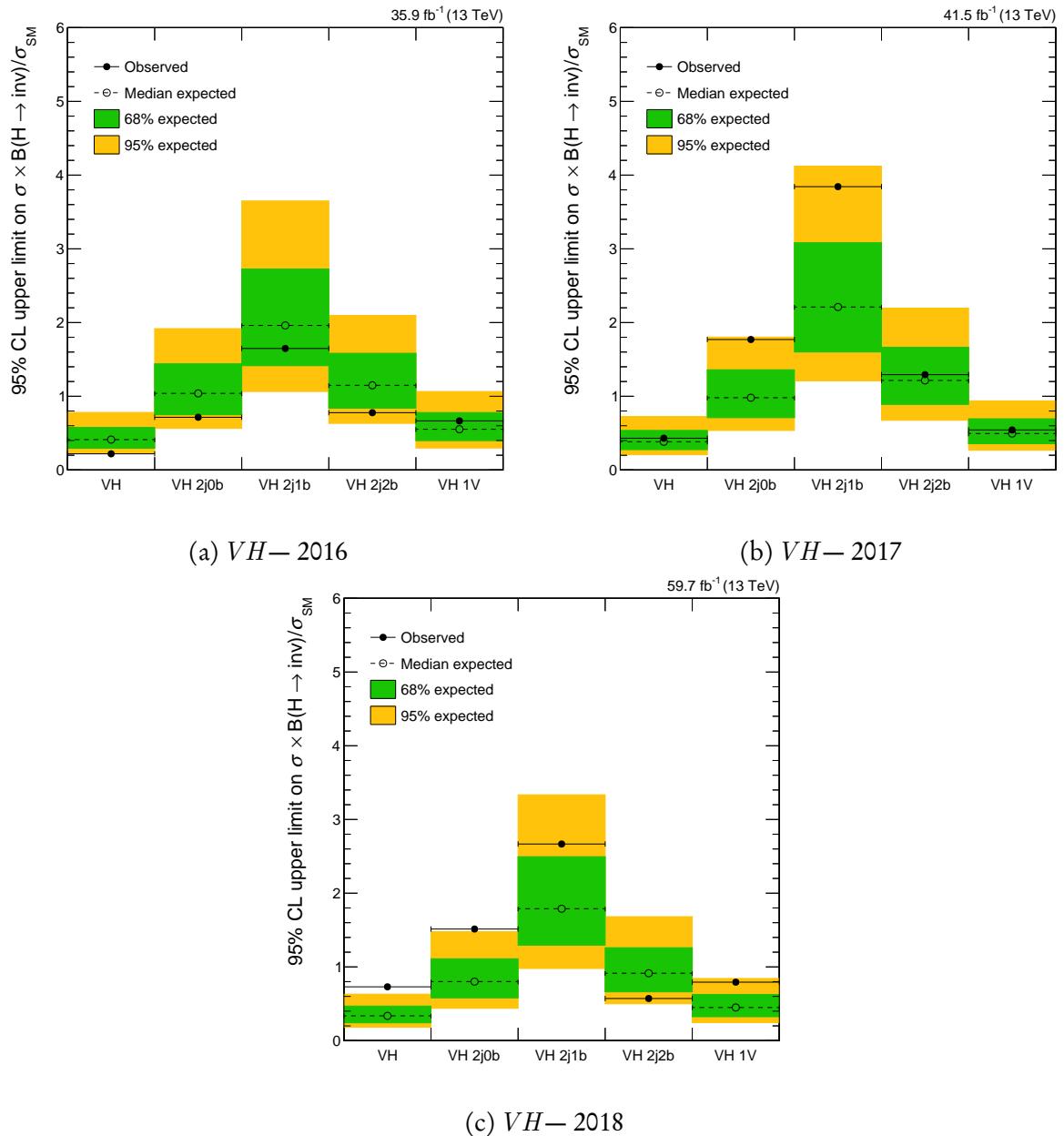
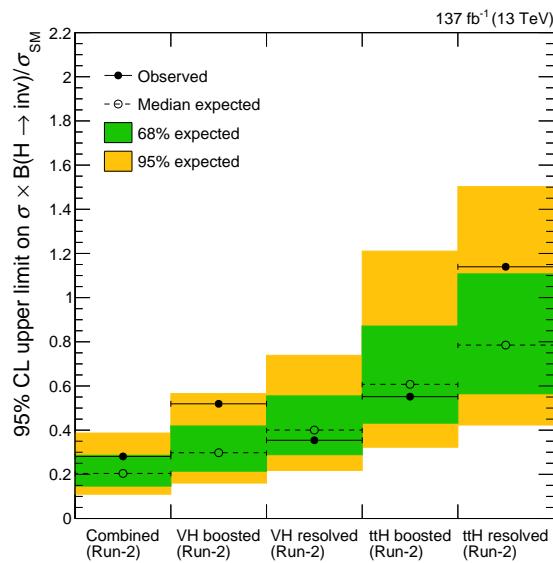
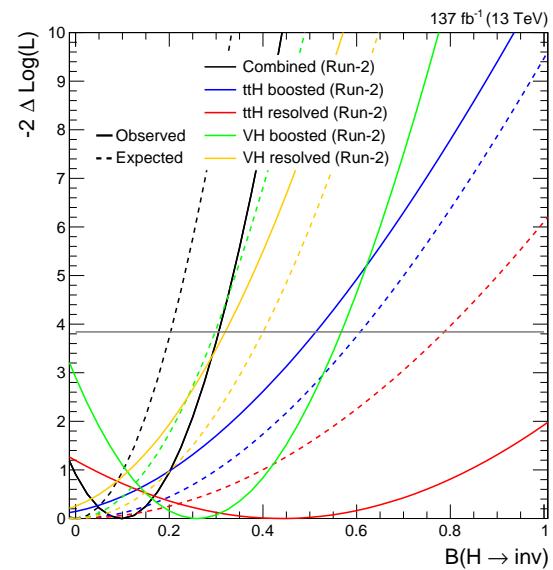


Figure A.12: Observed and expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the  $VH$  channel, for both the individual categories, and the combination of them, for each data-taking year in Run-2.



(a) Limit — Run-2



(b) Profile likelihood — Run-2

Figure A.13: Observed and expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction  $\mathcal{B}(H \rightarrow \text{invisible})$  (left) and the corresponding profile likelihood scan (right), in the  $t\bar{t}H$  and  $VH$  boosted and resolved category groupings and their combination, for the full Run-2 dataset.



## BIBLIOGRAPHY

- [1] M. AABOUD et al., *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*, Eur. Phys. J. C 77.10 (2017), p. 673, doi: [10.1140/epjc/s10052-017-5225-7](https://doi.org/10.1140/epjc/s10052-017-5225-7), arXiv: [1704.07983 \[hep-ex\]](https://arxiv.org/abs/1704.07983).
- [2] M. AABOUD et al., *Search for dark matter in events with a hadronically decaying vector boson and missing transverse momentum in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, JHEP 10 (2018), p. 180, doi: [10.1007/JHEP10\(2018\)180](https://doi.org/10.1007/JHEP10(2018)180), arXiv: [1807.11471 \[hep-ex\]](https://arxiv.org/abs/1807.11471).
- [3] M. AABOUD et al., *Observation of  $H \rightarrow b\bar{b}$  decays and  $VH$  production with the ATLAS detector*, Phys. Lett. B 786 (2018), pp. 59–86, doi: [10.1016/j.physletb.2018.09.013](https://doi.org/10.1016/j.physletb.2018.09.013), arXiv: [1808.08238 \[hep-ex\]](https://arxiv.org/abs/1808.08238).
- [4] G. AAD et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B 716 (2012), pp. 1–29, doi: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020), arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214).
- [5] R. AAIJ et al., *Measurement of the inelastic pp cross-section at a centre-of-mass energy of 13 TeV*, J. High Energy Phys. 06 (2018), p. 100, doi: [10.1007/JHEP06\(2018\)100](https://doi.org/10.1007/JHEP06(2018)100), arXiv: [1803.10974 \[hep-ex\]](https://arxiv.org/abs/1803.10974).
- [6] A. ABDELHAMEED et al., *First results from the CRESST-III low-mass dark matter program*, Phys. Rev. D 100.10 (2019), p. 102002, doi: [10.1103/PhysRevD.100.102002](https://doi.org/10.1103/PhysRevD.100.102002), arXiv: [1904.00498 \[astro-ph.CO\]](https://arxiv.org/abs/1904.00498).
- [7] D. ABERCROMBIE et al., *Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum*, Phys. Dark Univ. 27 (2020), ed. by A. BOVEIA et al., p. 100371, doi: [10.1016/j.dark.2019.100371](https://doi.org/10.1016/j.dark.2019.100371), arXiv: [1507.00966 \[hep-ex\]](https://arxiv.org/abs/1507.00966).
- [8] N. AGHANIM et al., *Planck 2018 results. VI. Cosmological parameters* (2018), arXiv: [1807.06209 \[astro-ph.CO\]](https://arxiv.org/abs/1807.06209).

## BIBLIOGRAPHY

---

- [9] R. AGNESE et al., *Search for Low-Mass Dark Matter with CDMSlite Using a Profile Likelihood Fit*, Phys. Rev. D 99.6 (2019), p. 062001, doi: [10.1103/PhysRevD.99.062001](https://doi.org/10.1103/PhysRevD.99.062001), arXiv: [1808.09098 \[astro-ph.CO\]](https://arxiv.org/abs/1808.09098).
- [10] S. AGOSTINELLI et al., *Geant4—a simulation toolkit*, Nucl. Instrum. Methods Phys. Res. A 506.3 (2003), pp. 250–303, doi: [10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [11] D. S. AKERIB et al., *The LUX-ZEPLIN (LZ) Experiment*, Nucl. Instrum. Methods Phys. Res. A 953 (2020), p. 163047, doi: [10.1016/j.nima.2019.163047](https://doi.org/10.1016/j.nima.2019.163047), arXiv: [1910.09124 \[physics.ins-det\]](https://arxiv.org/abs/1910.09124).
- [12] D. AKERIB et al., *Results from a search for dark matter in the complete LUX exposure*, Phys. Rev. Lett. 118.2 (2017), p. 021303, doi: [10.1103/PhysRevLett.118.021303](https://doi.org/10.1103/PhysRevLett.118.021303), arXiv: [1608.07648 \[astro-ph.CO\]](https://arxiv.org/abs/1608.07648).
- [13] A. ALLOUL et al., *FeynRules 2.0 - A complete toolbox for tree-level phenomenology*, Comput. Phys. Commun. 185 (2014), p. 2250, doi: [10.1016/j.cpc.2014.04.012](https://doi.org/10.1016/j.cpc.2014.04.012), arXiv: [1310.1921 \[hep-ph\]](https://arxiv.org/abs/1310.1921).
- [14] J. ALWALL et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, J. High Energy Phys. 07 (2014), p. 079, doi: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079), arXiv: [1405.0301 \[hep-ph\]](https://arxiv.org/abs/1405.0301).
- [15] J. ALWALL et al., *MadGraph 5 : Going Beyond*, J. High Energy Phys. 06 (2011), p. 128, doi: [10.1007/JHEP06\(2011\)128](https://doi.org/10.1007/JHEP06(2011)128), arXiv: [1106.0522 \[hep-ph\]](https://arxiv.org/abs/1106.0522).
- [16] A. A. ANUAR, *Electrons and photons at High Level Trigger in CMS for Run II*, Journal of Physics: Conference Series 664.8 (Dec. 2015), p. 082001, doi: [10.1088/1742-6596/664/8/082001](https://doi.org/10.1088/1742-6596/664/8/082001).
- [17] G. APOLLINARI et al., *High Luminosity Large Hadron Collider HL-LHC*, CERN Yellow Rep. 5 (2015), pp. 1–19, doi: [10.5170/CERN-2015-005.1](https://doi.org/10.5170/CERN-2015-005.1), arXiv: [1705.08830 \[physics.acc-ph\]](https://arxiv.org/abs/1705.08830).
- [18] E. APRILE et al., *Dark Matter Search Results from a One Ton-Year Exposure of XENON1T*, Phys. Rev. Lett. 121.11 (2018), p. 111302, doi: [10.1103/PhysRevLett.121.111302](https://doi.org/10.1103/PhysRevLett.121.111302), arXiv: [1805.12562 \[astro-ph.CO\]](https://arxiv.org/abs/1805.12562).
- [19] E. APRILE et al., *Light Dark Matter Search with Ionization Signals in XENON1T*, Phys. Rev. Lett. 123.25 (2019), p. 251801, doi: [10.1103/PhysRevLett.123.251801](https://doi.org/10.1103/PhysRevLett.123.251801), arXiv: [1907.11485 \[hep-ex\]](https://arxiv.org/abs/1907.11485).

- [20] G. ARCADI, A. DJOUADI and M. RAIDAL, *Dark Matter through the Higgs portal*, Phys. Rept. 842 (2020), pp. 1–180, doi: [10.1016/j.physrep.2019.11.003](https://doi.org/10.1016/j.physrep.2019.11.003), arXiv: [1903.03616 \[hep-ph\]](https://arxiv.org/abs/1903.03616).
- [21] G. ARNISON et al., *Experimental Observation of Lepton Pairs of Invariant Mass Around  $95\text{ GeV}/c^2$  at the CERN SPS Collider*, Phys. Lett. B 126.5 (1983), pp. 398–410, doi: [10.1016/0370-2693\(83\)90188-0](https://doi.org/10.1016/0370-2693(83)90188-0).
- [22] P. ARTOISENET et al., *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP 03 (2013), p. 015, doi: [10.1007/JHEP03\(2013\)015](https://doi.org/10.1007/JHEP03(2013)015), arXiv: [1212.3460 \[hep-ph\]](https://arxiv.org/abs/1212.3460).
- [23] P. ATHRON et al., *Global analyses of Higgs portal singlet dark matter models using GAMBIT*, Eur. Phys. J. C 79.1 (2019), p. 38, doi: [10.1140/epjc/s10052-018-6513-6](https://doi.org/10.1140/epjc/s10052-018-6513-6).
- [24] ATLAS COLLABORATION, *Combination of searches for invisible Higgs boson decays with the ATLAS experiment* (Oct. 2020), report number ATLAS-CONF-2020-052.
- [25] M. BACKOVIĆ et al., *Higher-order QCD predictions for dark matter production at the LHC in simplified models with s-channel mediators*, Eur. Phys. J. C 75.10 (2015), p. 482, doi: [10.1140/epjc/s10052-015-3700-6](https://doi.org/10.1140/epjc/s10052-015-3700-6), arXiv: [1508.05327 \[hep-ph\]](https://arxiv.org/abs/1508.05327).
- [26] P. BAGNAIA et al., *Evidence for  $Z^0 \rightarrow e^+e^-$  at the CERN  $\bar{p}p$  Collider*, Phys. Lett. B 129 (1983), pp. 130–140, doi: [10.1016/0370-2693\(83\)90744-X](https://doi.org/10.1016/0370-2693(83)90744-X).
- [27] I. BALDES and K. PETRAKI, *Asymmetric thermal-relic dark matter: Sommerfeld-enhanced freeze-out, annihilation signals and unitarity bounds*, J. Cosmol. Astropart. Phys. 2017.09 (2017), pp. 028–028, doi: [10.1088/1475-7516/2017/09/028](https://doi.org/10.1088/1475-7516/2017/09/028).
- [28] R. BARLOW and C. BEESTON, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. 77.2 (1993), pp. 219–228, doi: [https://doi.org/10.1016/0010-4655\(93\)90005-W](https://doi.org/10.1016/0010-4655(93)90005-W).
- [29] G. BATTAGLIA et al., *Erratum: The radial velocity dispersion profile of the Galactic halo: constraining the density profile of the dark halo of the Milky Way*, Mon. Not. R. Astron. Soc. 370.2 (2006), pp. 1055–1056, doi: [10.1111/j.1365-2966.2006.10688.x](https://doi.org/10.1111/j.1365-2966.2006.10688.x).
- [30] G. BATTAGLIA et al., *The radial velocity dispersion profile of the Galactic halo: constraining the density profile of the dark halo of the Milky Way*, Mon. Not. R. Astron. Soc. 364.2 (2005), pp. 433–442, doi: [10.1111/j.1365-2966.2005.09367.x](https://doi.org/10.1111/j.1365-2966.2005.09367.x), arXiv: [astro-ph/0506102 \[astro-ph\]](https://arxiv.org/abs/astro-ph/0506102).

## BIBLIOGRAPHY

---

- [31] G. L. BAYATYAN et al., *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*, Technical Design Report CMS.
- [32] C. M. BENDER and S. SARKAR, *Asymptotic Analysis of the Boltzmann Equation for Dark Matter Relics*, J. Math. Phys. 53 (2012), p. 103509, doi: [10.1063/1.4753990](https://doi.org/10.1063/1.4753990), arXiv: [1203.1822 \[hep-th\]](https://arxiv.org/abs/1203.1822).
- [33] M. BENEDIKT, D. SCHULTE and F. ZIMMERMANN, *Optimizing integrated luminosity of future hadron colliders*, Phys. Rev. Spec. Top. Accel. Beams 18.CERN-ACC-2015-0166. 10 (2015), 101002. 19 p, doi: [10.1103/PhysRevSTAB.18.101002](https://doi.org/10.1103/PhysRevSTAB.18.101002).
- [34] D. BERTOLINI et al., *Pileup Per Particle Identification*, JHEP 10 (2014), p. 059, doi: [10.1007/JHEP10\(2014\)059](https://doi.org/10.1007/JHEP10(2014)059), arXiv: [1407.6013 \[hep-ph\]](https://arxiv.org/abs/1407.6013).
- [35] P. BHATTACHARJEE et al., *Sizing-up the WIMPs of Milky Way : Deriving the velocity distribution of Galactic Dark Matter particles from the rotation curve data*, Phys. Rev. D87 (2013), p. 083525, doi: [10.1103/PhysRevD.87.083525](https://doi.org/10.1103/PhysRevD.87.083525), arXiv: [1210.2328 \[astro-ph.GA\]](https://arxiv.org/abs/1210.2328).
- [36] P. BLOCH et al., *Changes to CMS ECAL electronics: addendum to the Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 2002.
- [37] A. BOVEIA and C. DOGLIONI, *Dark Matter Searches at Colliders*, Ann. Rev. Nucl. Part. Sci. 68 (2018), pp. 429–459, doi: [10.1146/annurev-nucl-101917-021008](https://doi.org/10.1146/annurev-nucl-101917-021008), arXiv: [1810.12238 \[hep-ex\]](https://arxiv.org/abs/1810.12238).
- [38] A. H. BROEILS, *The mass distribution of the dwarf spiral NGC 1560*, Astron. Astrophys. 256 (Mar. 1992), pp. 19–32.
- [39] O. S. BRÜNING et al., *LHC Design Report*, CERN Yellow Reports: Monographs, CERN, Geneva, 2004, doi: [10.5170/CERN-2004-003-V-1](https://doi.org/10.5170/CERN-2004-003-V-1).
- [40] 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](https://doi.org/10.1140/epjc/s10052-012-1896-2), arXiv: [1111.6097 \[hep-ph\]](https://arxiv.org/abs/1111.6097).
- [41] M. CACCIARI, G. P. SALAM and G. SOYEZ, *The Anti- $k(t)$  jet clustering algorithm*, J. High Energy Phys. 04 (2008), p. 063, doi: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063), arXiv: [0802.1189 \[hep-ph\]](https://arxiv.org/abs/0802.1189).
- [42] A. CANEPA, *Searches for supersymmetry at the Large Hadron Collider*, Reviews in Physics 4 (2019), p. 100033, doi: [10.1016/j.revip.2019.100033](https://doi.org/10.1016/j.revip.2019.100033).

- [43] M. CEPEDA et al., *Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC*, in: *Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC*, ed. by A. DAINESI et al., vol. 7, 2019, pp. 221–584, doi: [10.23731/CYRM-2019-007.221](https://doi.org/10.23731/CYRM-2019-007.221), arXiv: [1902.00134 \[hep-ph\]](https://arxiv.org/abs/1902.00134).
- [44] D. G. CERDEÑO et al., *B anomalies and dark matter: a complex connection*, Eur. Phys. J. C 79.6 (2019), p. 517, doi: [10.1140/epjc/s10052-019-6979-x](https://doi.org/10.1140/epjc/s10052-019-6979-x).
- [45] S. CHATRCHYAN et al., *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, JINST 9.10 (2014), P10009, doi: [10.1088/1748-0221/9/10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009), arXiv: [1405.6569 \[physics.ins-det\]](https://arxiv.org/abs/1405.6569).
- [46] S. CHATRCHYAN et al., *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B 716 (2012), pp. 30–61, doi: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021), arXiv: [1207.7235 \[hep-ex\]](https://arxiv.org/abs/1207.7235).
- [47] S. CITTOLIN, A. RÁCZ and P. SPHICAS, *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*, Technical Design Report CMS, CERN, Geneva, 2002.
- [48] C. J. CLOPPER and E. S. PEARSON, *The use of confidence or fiducial limits illustrated in the case of the binomial*, Biometrika 26.4 (Dec. 1934), pp. 404–413, doi: [10.1093/biomet/26.4.404](https://doi.org/10.1093/biomet/26.4.404).
- [49] D. CLOWE, A. GONZALEZ and M. MARKEVITCH, *Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter*, Astrophys. J. 604.2 (2004), p. 596.
- [50] CMS COLLABORATION, *CMS luminosity measurement for the 2016 data taking period*, tech. rep. CMS-PAS-LUM-17-001, CERN, Geneva, 2017.
- [51] CMS COLLABORATION, *CMS luminosity measurement for the 2017 data-taking period at  $\sqrt{s} = 13$  TeV*, tech. rep. CMS-PAS-LUM-17-004, CERN, Geneva, 2018.
- [52] CMS COLLABORATION, *CMS luminosity measurement for the 2018 data-taking period at  $\sqrt{s} = 13$  TeV*, tech. rep. CMS-PAS-LUM-18-002, CERN, Geneva, 2019.
- [53] CMS COLLABORATION, *CMS Luminosity Public Results*, url: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults> (visited on 22/02/2020).
- [54] CMS COLLABORATION, *Cut Based Electron ID for Run 2*, access may be restricted, url: <https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2> (visited on 18/05/2020).

## BIBLIOGRAPHY

---

- [55] CMS COLLABORATION, *Cut Based Photon ID for Run 2*, access may be restricted, URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedPhotonIdentificationRun2> (visited on 19/05/2020).
- [56] CMS COLLABORATION, *Jet Identification for the 13 TeV data Run2016*, access may be restricted, URL: <https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016> (visited on 18/05/2020).
- [57] CMS COLLABORATION, *Jet Identification for the 13 TeV data Run2017*, access may be restricted, URL: <https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2017> (visited on 18/05/2020).
- [58] CMS COLLABORATION, *Jet Identification for the 13 TeV data Run2018*, access may be restricted, URL: <https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2018> (visited on 18/05/2020).
- [59] CMS COLLABORATION, *Luminosity Physics Object Group (Lumi POG)*, access may be restricted, URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TWikiLUM> (visited on 16/03/2020).
- [60] CMS COLLABORATION, *Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment*, tech. rep. CMS-PAS-JME-18-002, CERN, Geneva, 2019.
- [61] CMS COLLABORATION, *MET Filter Recommendations for Run II*, access may be restricted, URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2> (visited on 19/09/2020).
- [62] CMS COLLABORATION, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*, tech. rep. CMS-PAS-PFT-09-001, CERN, Geneva, Apr. 2009.
- [63] CMS COLLABORATION, *Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data*, J. Instrum. 5.03 (2010), T03012, DOI: [10.1088/1748-0221/5/03/t03012](https://doi.org/10.1088/1748-0221/5/03/t03012).
- [64] CMS COLLABORATION, *Reweighting recipe to emulate Level 1 ECAL prefiring*, access may be restricted, URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1ECALPrefiringWeightRecipe> (visited on 09/07/2020).
- [65] CMS COLLABORATION, *The CMS electromagnetic calorimeter project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.
- [66] CMS COLLABORATION, *The CMS hadron calorimeter project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.

- [67] CMS COLLABORATION, *The CMS magnet project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.
- [68] CMS COLLABORATION, *The CMS tracker: addendum to the Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 2000.
- [69] T. COHEN, M. LISANTI and H. K. LOU, *Semi-visible Jets: Dark Matter Undercover at the LHC*, Phys. Rev. Lett. 115.17 (2015), p. 171804, doi: [10.1103/PhysRevLett.115.171804](https://doi.org/10.1103/PhysRevLett.115.171804), arXiv: [1503.00009 \[hep-ph\]](https://arxiv.org/abs/1503.00009).
- [70] T. COHEN et al., *LHC Searches for Dark Sector Showers*, J. High Energy Phys. 11 (2017), p. 196, doi: [10.1007/JHEP11\(2017\)196](https://doi.org/10.1007/JHEP11(2017)196), arXiv: [1707.05326 \[hep-ph\]](https://arxiv.org/abs/1707.05326).
- [71] J. CONRAD and O. REIMER, *Indirect dark matter searches in gamma and cosmic rays*, Nature Phys. 13.3 (2017), pp. 224–231, doi: [10.1038/nphys4049](https://doi.org/10.1038/nphys4049), arXiv: [1705.11165 \[astro-ph.HE\]](https://arxiv.org/abs/1705.11165).
- [72] G. COWAN et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C 71 (2011), [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554, doi: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0), arXiv: [1007.1727 \[physics.data-an\]](https://arxiv.org/abs/1007.1727).
- [73] X. CUI et al., *Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment*, Phys. Rev. Lett. 119.18 (2017), p. 181302, doi: [10.1103/PhysRevLett.119.181302](https://doi.org/10.1103/PhysRevLett.119.181302), arXiv: [1708.06917 \[astro-ph.CO\]](https://arxiv.org/abs/1708.06917).
- [74] M. DELLA NEGRA et al., *CMS: letter of intent by the CMS Collaboration for a general purpose detector at LHC*, tech. rep. CERN-LHCC-92-003. LHCC-I-1, Open presentation to the LHCC 5 November 1992, M. Della Negra/CERN, CMS Spokesman, CERN, Geneva, 1992.
- [75] N. DEV et al., *The CMS Level-1 electron and photon trigger: for Run II of LHC*, J. Instrum. 12.02 (Feb. 2017), C02014. 10 p, doi: [10.1088/1748-0221/12/02/c02014](https://doi.org/10.1088/1748-0221/12/02/c02014).
- [76] A. DJOUADI et al., *Direct Detection of Higgs-Portal Dark Matter at the LHC*, Eur. Phys. J. C 73.6 (2013), p. 2455, doi: [10.1140/epjc/s10052-013-2455-1](https://doi.org/10.1140/epjc/s10052-013-2455-1), arXiv: [1205.3169 \[hep-ph\]](https://arxiv.org/abs/1205.3169).
- [77] A. DJOUADI et al., *Implications of LHC searches for Higgs-portal dark matter*, Phys. Lett. B 709 (2012), pp. 65–69, doi: [10.1016/j.physletb.2012.01.062](https://doi.org/10.1016/j.physletb.2012.01.062), arXiv: [1112.3299 \[hep-ph\]](https://arxiv.org/abs/1112.3299).
- [78] M. DREWES, *The Phenomenology of Right Handed Neutrinos*, Int. J. Mod. Phys. E 22.08 (2013), p. 1330019, doi: [10.1142/S0218301313300191](https://doi.org/10.1142/S0218301313300191).

## BIBLIOGRAPHY

---

- [79] J. EINASTO, *Dark Matter*, ArXiv e-prints (Jan. 2009), arXiv: [0901.0632 \[astro-ph.CO\]](#).
- [80] R. K. ELLIS, W. J. STIRLING and B. R. WEBBER, *QCD and collider physics*, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. 8 (1996), pp. 1–435.
- [81] F. ENGLERT and R. BROUT, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. 13 (9 Aug. 1964), pp. 321–323, doi: [10.1103/PhysRevLett.13.321](#).
- [82] M. FARINA, M. PERELSTEIN and N. REY-LE LORIER, *Higgs Couplings and Naturalness*, Phys. Rev. D90.1 (2014), p. 015014, doi: [10.1103/PhysRevD.90.015014](#), arXiv: [1305.6068 \[hep-ph\]](#).
- [83] *First constraints on invisible Higgs boson decays using  $t\bar{t}H$  production at  $\sqrt{s} = 13 \text{ TeV}$* , tech. rep. CMS-PAS-HIG-18-008, CERN, Geneva, 2019.
- [84] K. C. FREEMAN, *On the Disks of Spiral and S0 Galaxies*, Astrophys. J. 160 (June 1970), p. 811, doi: [10.1086/150474](#).
- [85] S. FRIXIONE, P. NASON and C. OLEARI, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, J. High Energy Phys. 11 (2007), p. 070, doi: [10.1088/1126-6708/2007/11/070](#), arXiv: [0709.2092 \[hep-ph\]](#).
- [86] S. L. GLASHOW, *Partial-symmetries of weak interactions*, Nuclear Physics 22.4 (1961), pp. 579–588, doi: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [87] P. GONDOLO and G. GELMINI, *Cosmic Abundances of Stable Particles: improved analysis*, Nucl. Phys. B 360 (1991), pp. 145–179, doi: [10.1016/0550-3213\(91\)90438-4](#).
- [88] G. S. GURALNIK, C. R. HAGEN and T. W. B. KIBBLE, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. 13 (20 Nov. 1964), pp. 585–587, doi: [10.1103/PhysRevLett.13.585](#).
- [89] L. J. HALL et al., *Freeze-In Production of FIMP Dark Matter*, J. High Energy Phys. 03 (2010), p. 080, doi: [10.1007/JHEP03\(2010\)080](#), arXiv: [0911.1120 \[hep-ph\]](#).
- [90] T. HAN, Z. LIU and A. NATARAJAN, *Dark matter and Higgs bosons in the MSSM*, J. High Energy Phys. 11 (2013), p. 008, doi: [10.1007/JHEP11\(2013\)008](#), arXiv: [1303.3040 \[hep-ph\]](#).
- [91] T. HAN, J. D. LYKKEN and R.-J. ZHANG, *On Kaluza-Klein states from large extra dimensions*, Phys. Rev. D59 (1999), p. 105006, doi: [10.1103/PhysRevD.59.105006](#), arXiv: [hep-ph/9811350 \[hep-ph\]](#).

- [92] S. HEINEMEYER et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*, ed. by S. HEINEMEYER, CERN Yellow Reports: Monographs, July 2013.
- [93] W. HERR and B. MURATORI, *Concept of luminosity* (2006), doi: [10.5170/CERN-2006-002.361](https://doi.org/10.5170/CERN-2006-002.361).
- [94] J. HERZOG-ARBEITMAN et al., *Empirical Determination of Dark Matter Velocities using Metal-Poor Stars*, Phys. Rev. Lett. 120.4 (2018), p. 041102, doi: [10.1103/PhysRevLett.120.041102](https://doi.org/10.1103/PhysRevLett.120.041102), arXiv: [1704.04499 \[astro-ph.GA\]](https://arxiv.org/abs/1704.04499).
- [95] P. W. HIGGS, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. 13 (16 Oct. 1964), pp. 508–509, doi: [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [96] M. HOFERICHTER et al., *Improved limits for Higgs-portal dark matter from LHC searches*, Phys. Rev. Lett. 119.18 (2017), p. 181803, doi: [10.1103/PhysRevLett.119.181803](https://doi.org/10.1103/PhysRevLett.119.181803), arXiv: [1708.02245 \[hep-ph\]](https://arxiv.org/abs/1708.02245).
- [97] D. HUTERER, *Weak lensing, dark matter and dark energy*, Gen. Relativ. Gravit. 42 (2010), pp. 2177–2195, doi: [10.1007/s10714-010-1051-z](https://doi.org/10.1007/s10714-010-1051-z), arXiv: [1001.1758 \[astro-ph.CO\]](https://arxiv.org/abs/1001.1758).
- [98] Q. INGRAM, *Energy resolution of the barrel of the CMS Electromagnetic Calorimeter*, J. Instrum. 2.04 (2007), P04004–P04004, doi: [10.1088/1748-0221/2/04/p04004](https://doi.org/10.1088/1748-0221/2/04/p04004).
- [99] P. R. KAFLE et al., *On the Shoulders of Giants: Properties of the Stellar Halo and the Milky Way Mass Distribution*, Astrophys. J. 794.1 (2014), p. 59, doi: [10.1088/0004-637X/794/1/59](https://doi.org/10.1088/0004-637X/794/1/59), arXiv: [1408.1787 \[astro-ph.GA\]](https://arxiv.org/abs/1408.1787).
- [100] M. KAKIZAKI et al., *Phenomenological constraints on light mixed sneutrino dark matter scenarios*, Phys. Lett. B 749 (2015), pp. 44–49, doi: [10.1016/j.physletb.2015.07.030](https://doi.org/10.1016/j.physletb.2015.07.030).
- [101] J. C. KAPTEYN, *First Attempt at a Theory of the Arrangement and Motion of the Sidereal System*, Astrophys. J. 55 (May 1922), p. 302, doi: [10.1086/142670](https://doi.org/10.1086/142670).
- [102] V. KARIMÄKI et al., *The CMS tracker system project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.
- [103] V. KHACHATRYAN et al., *A search for new phenomena in  $pp$  collisions at  $\sqrt{s} = 13\text{ TeV}$  in final states with missing transverse momentum and at least one jet using the  $\alpha_T$  variable*, Eur. Phys. J. C 77.5 (2017), p. 294, doi: [10.1140/epjc/s10052-017-4787-8](https://doi.org/10.1140/epjc/s10052-017-4787-8), arXiv: [1611.00338 \[hep-ex\]](https://arxiv.org/abs/1611.00338).

## BIBLIOGRAPHY

---

- [104] V. KHACHATRYAN et al., *Event generator tunes obtained from underlying event and multiparton scattering measurements*, Eur. Phys. J. C 76.3 (2016), p. 155, doi: [10.1140/epjc/s10052-016-3988-x](https://doi.org/10.1140/epjc/s10052-016-3988-x), arXiv: [1512.00815 \[hep-ex\]](https://arxiv.org/abs/1512.00815).
- [105] V. KHACHATRYAN et al., *Search for dark matter, extra dimensions, and unparticles in monojet events in proton-proton collisions at  $\sqrt{s} = 8 \text{ TeV}$* , Eur. Phys. J. C 75.5 (2015), p. 235, doi: [10.1140/epjc/s10052-015-3451-4](https://doi.org/10.1140/epjc/s10052-015-3451-4), arXiv: [1408.3583 \[hep-ex\]](https://arxiv.org/abs/1408.3583).
- [106] V. KHACHATRYAN et al., *Search for Microscopic Black Hole Signatures at the Large Hadron Collider*, Phys. Lett. B 697 (2011), pp. 434–453, doi: [10.1016/j.physletb.2011.02.032](https://doi.org/10.1016/j.physletb.2011.02.032), arXiv: [1012.3375 \[hep-ex\]](https://arxiv.org/abs/1012.3375).
- [107] V. KHACHATRYAN et al., *Searches for invisible decays of the Higgs boson in  $pp$  collisions at  $\sqrt{s} = 7, 8, \text{ and } 13 \text{ TeV}$* , J. High Energy Phys. 02 (2017), p. 135, doi: [10.1007/JHEP02\(2017\)135](https://doi.org/10.1007/JHEP02(2017)135), arXiv: [1610.09218 \[hep-ex\]](https://arxiv.org/abs/1610.09218).
- [108] F. KRAUSS and D. NAPOLETANO, *Towards a fully massive five-flavor scheme*, Phys. Rev. D 98.9 (2018), p. 096002, doi: [10.1103/PhysRevD.98.096002](https://doi.org/10.1103/PhysRevD.98.096002), arXiv: [1712.06832 \[hep-ph\]](https://arxiv.org/abs/1712.06832).
- [109] G. KRNJAC, *Freezing In, Heating Up, and Freezing Out: Predictive Nonthermal Dark Matter and Low-Mass Direct Detection*, J. High Energy Phys. 10 (2018), p. 136, doi: [10.1007/JHEP10\(2018\)136](https://doi.org/10.1007/JHEP10(2018)136), arXiv: [1711.11038 \[hep-ph\]](https://arxiv.org/abs/1711.11038).
- [110] A. J. LARKOSKI et al., *Soft Drop*, JHEP 05 (2014), p. 146, doi: [10.1007/JHEP05\(2014\)146](https://doi.org/10.1007/JHEP05(2014)146), arXiv: [1402.2657 \[hep-ph\]](https://arxiv.org/abs/1402.2657).
- [111] J. G. LAYER, *The CMS muon project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.
- [112] J. M. LINDERT et al., *Precise predictions for  $V + \text{jets}$  dark matter backgrounds*, Eur. Phys. J. C 77.12 (2017), doi: [10.1140/epjc/s10052-017-5389-1](https://doi.org/10.1140/epjc/s10052-017-5389-1).
- [113] M. LISANTI, *Lectures on Dark Matter Physics*, in: *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings (TASI 2015): Boulder, CO, USA, June 1-26, 2015*, 2017, pp. 399–446, doi: [10.1142/9789813149441\\_0007](https://doi.org/10.1142/9789813149441_0007), arXiv: [1603.03797 \[hep-ph\]](https://arxiv.org/abs/1603.03797).
- [114] S. P. MARTIN, *A Supersymmetry primer*, Adv. Ser. Direct. High Energy Phys 18 (1998), pp. 1–98, doi: [10.1142/9789812839657\\_0001](https://doi.org/10.1142/9789812839657_0001), [10.1142/9789814307505\\_0001](https://doi.org/10.1142/9789814307505_0001), arXiv: [hep-ph/9709356 \[hep-ph\]](https://arxiv.org/abs/hep-ph/9709356).
- [115] J. MOUSA et al., *Overview of the CMS Detector Performance at LHC Run 2*, Universe 5 (2019), p. 18, doi: [10.3390/universe5010018](https://doi.org/10.3390/universe5010018).

- [116] P. NASON, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, J. High Energy Phys. 11 (2004), p. 040, doi: [10.1088/1126-6708/2004/11/040](https://doi.org/10.1088/1126-6708/2004/11/040), arXiv: [hep-ph/0409146 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0409146).
- [117] E. NOETHER, *Invariant variation problems*, Transport Theor. Stat. Phys. 1.3 (1971), M. A. Tavel's English translation of Noether's Theorems (1918), pp. 186–207, doi: [10.1080/00411457108231446](https://doi.org/10.1080/00411457108231446).
- [118] M. PERSIC, P. SALUCCI and F. STEL, *The universal rotation curve of spiral galaxies - I. The dark matter connection*, Mon. Not. R. Astron. Soc. 281 (July 1996), pp. 27–47, doi: [10.1093/mnras/281.1.27](https://doi.org/10.1093/mnras/281.1.27), eprint: [astro-ph/9506004](https://arxiv.org/abs/astro-ph/9506004).
- [119] G. RAUCO, *Search for New Physics in All-Hadronic Final States with the CMS Experiment*, PhD thesis, Universität Zürich, 2019.
- [120] A. L. READ, *Presentation of search results: the CL<sub>s</sub> technique*, J. Phys. G 28.10 (2002), pp. 2693–2704, doi: [10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).
- [121] T. SAKUMA, *Cutaway diagrams of CMS detector* (May 2019).
- [122] T. SAKUMA, H. FLAECHER and D. SMITH, *Alternative angular variables for suppression of QCD multijet events in new physics searches with missing transverse momentum at the LHC* (Mar. 2018), arXiv: [1803.07942 \[hep-ph\]](https://arxiv.org/abs/1803.07942).
- [123] A. SALAM, *Weak and Electromagnetic Interactions*, Conf. Proc. C 680519 (1968), pp. 367–377, doi: [10.1142/9789812795915\\_0034](https://doi.org/10.1142/9789812795915_0034).
- [124] S. SCHRAMM, *Searching for Dark Matter with the ATLAS Detector*, Springer Theses, Springer International Publishing, 2016, ISBN: 9783319444536.
- [125] M. SCHUMANN, *Direct Detection of WIMP Dark Matter: Concepts and Status*, J. Phys. G46.10 (2019), p. 103003, doi: [10.1088/1361-6471/ab2ea5](https://doi.org/10.1088/1361-6471/ab2ea5), arXiv: [1903.03026 \[astro-ph.CO\]](https://arxiv.org/abs/1903.03026).
- [126] SCIENCE AND TECHNOLOGY FACILITIES COUNCIL, *CERN Accelerator Complex*, URL: <https://stfc.ukri.org/research/particle-physics-and-particle-astronomy/large-hadron-collider/cern-accelerator-complex/> (visited on 22/03/2020).
- [127] A. M. SIRUNYAN et al., *Observation of Higgs boson decay to bottom quarks*, Phys. Rev. Lett. 121.12 (2018), p. 121801, doi: [10.1103/PhysRevLett.121.121801](https://doi.org/10.1103/PhysRevLett.121.121801), arXiv: [1808.08242 \[hep-ex\]](https://arxiv.org/abs/1808.08242).

## BIBLIOGRAPHY

---

- [128] A. M. SIRUNYAN et al., *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 \[physics.ins-det\]](https://arxiv.org/abs/1706.04965).
- [129] A. M. SIRUNYAN et al., *Search for natural and split supersymmetry in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$  in final states with jets and missing transverse momentum*, J. High Energy Phys. 2018.5 (May 2018), p. 25, doi: [10.1007/JHEP05\(2018\)025](https://doi.org/10.1007/JHEP05(2018)025).
- [130] A. M. SIRUNYAN et al., *Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at  $\sqrt{s} = 13 \text{ TeV}$* , Phys. Rev. D 97.9 (2018), p. 092005, doi: [10.1103/PhysRevD.97.092005](https://doi.org/10.1103/PhysRevD.97.092005), arXiv: [1712.02345 \[hep-ex\]](https://arxiv.org/abs/1712.02345).
- [131] A. M. SIRUNYAN et al., *Search for dark photons in decays of Higgs bosons produced in association with Z bosons in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$* , J. High Energy Phys. 2019.10 (2019), doi: [10.1007/jhep10\(2019\)139](https://doi.org/10.1007/jhep10(2019)139).
- [132] A. SIRUNYAN et al., *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, JINST 13.05 (2018), P05011, doi: [10.1088/1748-0221/13/05/P05011](https://doi.org/10.1088/1748-0221/13/05/P05011), arXiv: [1712.07158 \[physics.ins-det\]](https://arxiv.org/abs/1712.07158).
- [133] A. SIRUNYAN et al., *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$* , JINST 13.06 (2018), P06015, doi: [10.1088/1748-0221/13/06/P06015](https://doi.org/10.1088/1748-0221/13/06/P06015), arXiv: [1804.04528 \[physics.ins-det\]](https://arxiv.org/abs/1804.04528).
- [134] A. SIRUNYAN et al., *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$* , J. Instrum. 13.06 (2018), P06015, doi: [10.1088/1748-0221/13/06/p06015](https://doi.org/10.1088/1748-0221/13/06/p06015).
- [135] A. M. SIRUNYAN et al., *Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements*, Eur. Phys. J. C 80.1 (2020), p. 4, doi: [10.1140/epjc/s10052-019-7499-4](https://doi.org/10.1140/epjc/s10052-019-7499-4), arXiv: [1903.12179 \[hep-ex\]](https://arxiv.org/abs/1903.12179).
- [136] A. M. SIRUNYAN et al., *Measurements of  $t\bar{t}$  differential cross sections in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$  using events containing two leptons*, JHEP 02 (2019), p. 149, doi: [10.1007/JHEP02\(2019\)149](https://doi.org/10.1007/JHEP02(2019)149), arXiv: [1811.06625 \[hep-ex\]](https://arxiv.org/abs/1811.06625).
- [137] A. M. SIRUNYAN et al., *Search for dijet resonances in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$  and constraints on dark matter and other models*, Phys. Lett. B 769 (2017), [Erratum: Phys. Lett.B772,882(2017)], pp. 520–542, doi: [10.1016/j.physletb.2017.09.029](https://doi.org/10.1016/j.physletb.2017.09.029), [10.1016/j.physletb.2017.02.012](https://doi.org/10.1016/j.physletb.2017.02.012), arXiv: [1611.03568 \[hep-ex\]](https://arxiv.org/abs/1611.03568).

- [138] A. M. SIRUNYAN et al., *Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, JHEP 10 (2017), p. 005, DOI: [10.1007/JHEP10\(2017\)005](https://doi.org/10.1007/JHEP10(2017)005), arXiv: [1707.03316 \[hep-ex\]](https://arxiv.org/abs/1707.03316).
- [139] A. M. SIRUNYAN et al., *Search for invisible decays of a Higgs boson produced through vector boson fusion in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, Phys. Lett. B 793 (2019), pp. 520–551, DOI: [10.1016/j.physletb.2019.04.025](https://doi.org/10.1016/j.physletb.2019.04.025), arXiv: [1809.05937 \[hep-ex\]](https://arxiv.org/abs/1809.05937).
- [140] A. M. SIRUNYAN et al., *Search for new particles decaying to a jet and an emerging jet*, J. High Energy Phys. 02 (2019), p. 179, DOI: [10.1007/JHEP02\(2019\)179](https://doi.org/10.1007/JHEP02(2019)179), arXiv: [1810.10069 \[hep-ex\]](https://arxiv.org/abs/1810.10069).
- [141] T. SJÖSTRAND et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. 191 (2015), pp. 159–177, DOI: [10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024), arXiv: [1410.3012 \[hep-ph\]](https://arxiv.org/abs/1410.3012).
- [142] M. J. STRASSLER and K. M. ZUREK, *Echoes of a hidden valley at hadron colliders*, Phys. Lett. B 651 (2007), pp. 374–379, DOI: [10.1016/j.physletb.2007.06.055](https://doi.org/10.1016/j.physletb.2007.06.055), arXiv: [hep-ph/0604261 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0604261).
- [143] M. TANABASHI et al., *Review of Particle Physics*, Phys. Rev. D 98 (3 2018), p. 030001, DOI: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001).
- [144] A. VICENTE, *Anomalies in  $b \rightarrow s$  transitions and dark matter*, Adv. High Energy Phys. 2018 (2018), p. 3905848, DOI: [10.1155/2018/3905848](https://doi.org/10.1155/2018/3905848), arXiv: [1803.04703 \[hep-ph\]](https://arxiv.org/abs/1803.04703).
- [145] S. WEINBERG, *A Model of Leptons*, Phys. Rev. Lett. 19 (21 1967), pp. 1264–1266, DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [146] S. D. M. WHITE and M. J. REES, *Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering*, Mon. Not. R. Astron. Soc. 183.3 (1978), p. 341, DOI: [10.1093/mnras/183.3.341](https://doi.org/10.1093/mnras/183.3.341).
- [147] C. YOZIN and K. BEKKI, *The quenching and survival of ultra-diffuse galaxies in the Coma cluster*, Mon. Not. R. Astron. Soc. 452.1 (2015), pp. 937–943, DOI: [10.1093/mnras/stv1073](https://doi.org/10.1093/mnras/stv1073), arXiv: [1507.05161 \[astro-ph.GA\]](https://arxiv.org/abs/1507.05161).
- [148] A. ZABI et al., *The CMS Level-1 Calorimeter Trigger for the LHC Run II*, J. Instrum. 12.01 (2017), p. C01065, DOI: [10.1088/1748-0221/12/01/c01065](https://doi.org/10.1088/1748-0221/12/01/c01065).



## GLOSSARY

***b*-jet** A jet identified by a given algorithm or classifier as originating from a  $b$  quark.

**PARTICLE FLOW algorithm** An event reconstruction algorithm used in CMS. Information from all subdetectors is combined so that all stable particles can be identified. Jets are clustered, and complex objects such as  $b$ -jets are classified.

**anti- $k_{\mathrm{T}}$  algorithm** A sequential clustering algorithm designed to group hadronic particles into jets with a radius parameter  $R$ . The transverse momentum is symbolised as  $k_{\mathrm{T}}$  instead of  $p_{\mathrm{T}}$ . In CMS, standard radius parameters are  $R = 0.4$  and  $R = 0.8$ , referred to as AK4 and AK8 jets, respectively.

**control region** A region of phase space orthogonal to the signal region, typically by requiring an object that would otherwise be vetoed in the signal region. A control region enriched in a background process present in the signal region (such as  $W(\rightarrow \ell\nu) + \text{jets}$ ) can be used to model its influence more accurately.

**jet** A collimated shower of hadronic particles. High momentum quarks and gluons fragment and radiate; the resulting particles combine into hadrons due to colour confinement and deposit energy in the detector very close to each other and is reconstructed as a single physics object called a jet.

**luminosity** Instantaneous luminosity is a measure of the collision rate in a particle accelerator (given in units of area per unit time). The integrated luminosity is the instantaneous luminosity integrated over time and is a metric for measuring the total amount of data delivered by an accelerator or collected by a detector.

**missing transverse energy** The negative vector sum of the transverse momentum of all particles in a collider event. It is sometimes abbreviated to “MET” with the symbol  $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ , and also referred to in literature as “missing transverse momentum” ( $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ ).

## GLOSSARY

---

**pileup** The term ascribed to additional proton-proton collisions during a bunch crossing. Pileup interactions typically produce a large number of low-momentum particles.

**semi-visible jet** A shower of standard model and dark hadrons from the decay of a leptophobic  $Z'$  or  $\Phi$  mediator that couples the hidden sector to the standard model.

**sideband** A region of phase space orthogonal to the signal region, typically by inverting kinematic requirements. Similar to a control region, a sideband enriched in a background present in the signal region (such as QCD) can be used to model it more accurately.

## ACRONYMS

**CMSSW** CMS SoftWare.

**ALICE** A Large Ion Collider Experiment.

**AOD** Analysis Object Data.

**ATLAS** A Toroidal LHC ApparatuS.

**BSM** beyond the standard model.

**CERN** *Organisation Européenne pour la Recherche Nucléaire* (European Organisation for Nuclear Research).

**CMS** Compact Muon Solenoid.

**DAQ** data acquisition.

**ECAL** electromagnetic calorimeter.

**FASER** ForwArd Search ExpeRiment.

**FCC** Future Circular Collider.

**FNAL** Fermi National Accelerator Laboratory.

**GSF** gaussian sum filter.

**HCAL** hadron calorimeter.

**HF** hadron forward calorimeter.

**HL-LHC** High Luminosity Large Hadron Collider.

## ACRONYMS

---

**HLT** High-Level Trigger.

**ILC** International Linear Collider.

**JEC** jet energy corrections.

**JER** jet energy resolution.

**JES** jet energy scale.

**L1** Level-1.

**L1T** Level-1 Trigger.

**LHC** Large Hadron Collider.

**LO** leading order.

**LSP** lightest supersymmetric particle.

**LZ** LUX-ZEPLIN.

**MC** Monte Carlo.

**MoEDAL** Monopole and Exotics Detector At the LHC.

**NLO** next-to-leading order.

**NNLO** next-to-next-to-leading order.

**PF** PARTICLE FLOW.

**POG** Physics Object Group.

**PS** Proton Synchrotron.

**PSB** Proton Synchrotron Booster.

**QCD** Quantum Chromodynamics.

**SM** standard model.

**SPS** Super Proton Synchrotron.

**SUSY** supersymmetry.

**TOTEM** TOTal Elastic and diffractive cross section Measurement.

**VBF** vector boson fusion.

**WIMP** Weakly Interacting Massive Particle.

