

---

---

# A Higgs in the Dark

*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.

JULY 2020

Word count: number in words



## ABSTRACT

H ere goes the abstract



## 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.

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.

To my friends 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>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Evidence for dark matter . . . . .	1
1.1.1 Alternative theories to dark matter . . . . .	3
1.2 Overview of dark matter searches . . . . .	4
1.2.1 Dark matter searches at the LHC . . . . .	4
<b>2 The standard model and beyond</b>	<b>9</b>
2.1 The standard model of particle physics . . . . .	9
2.1.1 Electroweak symmetry breaking and the Higgs mechanism . . . . .	11
2.1.2 Limitations of the standard model . . . . .	11
2.2 Theoretical motivations for, and descriptions of, dark matter . . . . .	11
2.3 Measuring the branching ratio of invisibly decaying Higgs bosons . . . . .	13
2.3.1 Production modes of the Higgs boson . . . . .	14
2.3.2 Results of previous searches . . . . .	17
2.4 Searches for semi-visible jets . . . . .	17
2.4.1 Kinematics and free parameters of the model . . . . .	19
<b>3 Search for semi-visible jets</b>	<b>23</b>
3.1 Analysis summary . . . . .	23
3.2 Signal simulation overview . . . . .	24
3.3 Generation in PYTHIA . . . . .	25
3.4 Showering in PYTHIA . . . . .	25
3.4.1 Hadronisation of dark quarks . . . . .	26

---

## TABLE OF CONTENTS

---

3.4.2	Jet clustering . . . . .	27
3.4.3	Filtering events . . . . .	28
3.5	Generation in <b>MADGRAPH</b> . . . . .	29
3.5.1	<b>MADGRAPH</b> run settings . . . . .	29
3.5.2	<i>s</i> -channel . . . . .	30
3.5.3	<i>t</i> -channel . . . . .	32
3.6	PYTHIA- <b>MADGRAPH</b> comparisons for <i>s</i> -channel signal . . . . .	32
<b>4</b>	<b>The LHC and the CMS experiment</b>	<b>39</b>
4.1	CERN . . . . .	39
4.2	The Large Hadron Collider . . . . .	41
4.2.1	A proton’s journey . . . . .	41
4.2.2	Luminosity . . . . .	42
4.2.3	Pileup . . . . .	44
4.2.4	Evolution of the LHC . . . . .	45
4.3	The CMS experiment . . . . .	47
4.3.1	The CMS detector . . . . .	48
4.3.2	Data acquisition and triggering . . . . .	54
4.3.3	Simulating CMS data . . . . .	57
4.3.4	Jet energy corrections in the Level-1 Trigger . . . . .	58
<b>5</b>	<b>Definitions of physics objects and observables</b>	<b>63</b>
5.1	Ubiquitous observables and quantities in collider physics . . . . .	63
5.1.1	The electron volt . . . . .	63
5.1.2	Transverse momentum ( $\vec{p}_T$ ) . . . . .	64
5.1.3	$H_T$ . . . . .	64
5.1.4	Missing transverse momentum ( $\vec{p}_T^{\text{miss}}$ ) . . . . .	65
5.1.5	Transverse mass ( $m_T$ ) . . . . .	65
5.2	Classification of analysis-level physics objects . . . . .	66
5.2.1	Jets . . . . .	66
5.2.2	Muons . . . . .	68
5.2.3	Electrons . . . . .	69
5.2.4	Photons . . . . .	70
5.2.5	Tau leptons . . . . .	72
5.2.6	Revised energy sums $\vec{p}_T^{\text{miss}}$ , $H_T$ , and $\vec{H}_T^{\text{miss}}$ . . . . .	72

<b>6 Combined search for invisibly decaying Higgs bosons in hadronic channels</b>	<b>73</b>
6.1 Overview of the analysis	74
6.2 Only tools and forces: software and toolkits	75
6.3 Data and simulation	76
6.3.1 Data	76
6.3.2 Simulated signal processes	77
6.3.3 Simulated background processes	78
6.4 Weights, corrections, and systematic uncertainties for simulation	80
6.4.1 Veto and selection weights	81
6.4.2 Cross section reweighting	82
6.4.3 Pileup reweighting	83
6.4.4 Higher order corrections to $V + \text{jets}$ samples	84
6.4.5 Efficiency of the triggers	85
6.4.6 The $t\bar{t}$ process	85
6.4.7 Object-level scale factors and systematic uncertainties	87
6.4.8 Minor weights and systematic uncertainties	88
6.5 Event selection	89
6.5.1 Preselection	89
6.5.2 Additional filters	90
6.5.3 Mitigating the HEM issue	91
6.6 Categorisation of the non-VBF production modes	91
6.6.1 Classifying boosted topologies from top quarks, and $W$ and $Z$ bosons	94
6.6.2 Optimisation of the categories	95
6.6.3 Binning	98
6.7 Signal region, control region, and sideband definitions	98
6.7.1 The signal region	98
6.7.2 Control regions	99
6.7.3 Sidebands to the signal region	102
6.8 Background estimation	107
6.8.1 Lost lepton ( $W$ and $t\bar{t} + \text{jets}$ )	107
6.8.2 $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$	107
6.8.3 QCD multijet	107
6.9 Statistical model and fit	108
6.10 Analysis of the $t\bar{t}H$ mode	108
6.11 Analysis of the $VH$ mode	108

## TABLE OF CONTENTS

---

6.12 Analysis of the $ggH$ mode . . . . .	108
6.13 Combined results . . . . .	108
6.14 Interpretations in simplified dark matter scenarios . . . . .	114
6.14.1 Higgs portal model with the standard model Higgs boson . . . . .	114
6.14.2 Constraints on a standard model-like Higgs boson . . . . .	115
<b>7 Conclusions</b>	<b>117</b>
<b>A Appendix A</b>	<b>119</b>
<b>Bibliography</b>	<b>121</b>
<b>Glossary</b>	<b>135</b>
<b>Acronyms</b>	<b>137</b>

## LIST OF FIGURES

FIGURE	Page
1.1 A visual representation of the three main types of dark matter detection: direct, indirect, and production . . . . .	5
1.2 The expected masses and interaction cross sections of some dark matter candidates. The LHC, with its centre of mass energy of 13 TeV, is best suited to targeting WIMPs . . . . .	6
2.1 A measure of the comoving number density of WIMP dark matter as a function of time with projections for different particle masses . . . . .	13
2.2 A subset of the Feynman diagrams for the four predominant production mechanisms of the Higgs boson at the LHC . . . . .	15
2.3 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 . . . . .	19
2.4 The constituents of a semi-visible jet as a function of its invisible fraction . . . . .	20
2.5 The typical direction of the missing transverse energy relative to the semi-visible jets as a function of the invisible fraction $r_{\text{inv}}$ . . . . .	21
3.1 A diagram of the mass insertion decay of $\pi_{\text{dark}}$ mesons in the $s$ -channel semi-visible jet model . . . . .	26
3.2 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 . . . . .	27
3.3 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 . . . . .	31

## LIST OF FIGURES

---

3.4	Distributions of the missing transverse momentum 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 . . . . .	33
3.5	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 . . . . .	35
3.6	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 . . . . .	36
4.1	A schematic of the CERN accelerator complex . . . . .	43
4.2	The average number of pileup interactions per bunch crossing at CMS during Run-2 of the LHC . . . . .	45
4.3	The integrated luminosity of $pp$ collision data delivered to CMS during Run-2 of the LHC . . . . .	46
4.4	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	49
4.5	A quadrant of the CMS detector illustrating the main subsystems with their radius $R$ , longitudinal distance $z$ , and pseudorapidity $\eta$ from the interaction point	50
4.6	A transverse slice through the barrel section of the CMS detector with the main subsystems and components visible . . . . .	54
4.7	A summary of the CMS Level-1 Trigger data flow from the hits recorded by the subsystems to the Global Trigger . . . . .	56
4.8	Examples of correction curves used to calibrate the jet energies in two $ \eta $ bins .	60
4.9	The energies of matched pairs of jets in the entire barrel and end cap, in the pileup 40–50 range, before and after jet energy corrections have been applied .	61
4.10	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 . . . . .	61
6.1	NLO QCD and electroweak $k$ -factors used to reweight events in the LO $V + \text{jets}$ background samples . . . . .	85
6.2	The deviations relative to the nominal weight of the combinations of QCD renormalisation and factorisation scale variations. They are presented as a function of $p_T^{\text{miss}}$ in the $t\bar{t}H$ categories for the $t\bar{t}$ samples . . . . .	87
6.3	The azimuthal angle of the $\vec{p}_T^{\text{miss}}$ in the $t\bar{t}H$ and $VH$ categories before applying the selections designed to mitigate the HEM issue in 2018 . . . . .	92

6.4	The psuedorapidity of the electron in the $VH$ category of the $e + \text{jets}$ control region before applying the selection designed to mitigate the HEM issue in 2018	92
6.5	Distributions of $\tilde{\omega}_{\min}$ in the signal region in the $t\bar{t}H$ , $VH$ , and $ggH$ categories, along with the significance—using several figures of merit—if a cut is placed to the right of a given value	97
6.6	Pre-fit yields of the $p_T^{\text{miss}}$ distribution in the signal region for the combined boosted and resolved categories for the $t\bar{t}H$ and $VH$ processes, using simulation derived for each year in Run-2 and scaled to the required luminosity	99
6.7	Pre-fit yields of the $p_T^{\text{miss}}$ distribution in the $\mu + \text{jets}$ control region for the combined boosted and resolved categories for the $t\bar{t}H$ and $VH$ processes, using the full Run-2 dataset for data and simulation	102
6.8	Pre-fit yields of the $p_T^{\text{miss}}$ distribution in the $\mu\mu + \text{jets}$ control region for the combined boosted and resolved categories for the $t\bar{t}H$ and $VH$ processes, using the full Run-2 dataset for data and simulation	102
6.9	Pre-fit yields of the $p_T^{\text{miss}}$ distribution in the $e+\text{jets}$ control region for the combined boosted and resolved categories for the $t\bar{t}H$ and $VH$ processes, using the full Run-2 dataset for data and simulation	103
6.10	Pre-fit yields of the $p_T^{\text{miss}}$ distribution in the $ee + \text{jets}$ control region for the combined boosted and resolved categories for the $t\bar{t}H$ and $VH$ processes, using the full Run-2 dataset for data and simulation	103
6.11	Pre-fit yields of the $p_T^{\text{miss}}$ distribution in the $\gamma+\text{jets}$ control region for the combined boosted and resolved categories for the $t\bar{t}H$ and $VH$ processes, using the full Run-2 dataset for data and simulation	104
6.12	Pre-fit yields of the $p_T^{\text{miss}}$ distribution in the tight double sideband for the combined boosted and resolved categories for the $t\bar{t}H$ and $VH$ processes, using the full Run-2 dataset for data and simulation	105
6.13	Pre-fit yields of the $p_T^{\text{miss}}$ distribution in the loose double sideband for the combined boosted and resolved categories for the $t\bar{t}H$ and $VH$ processes, using the full Run-2 dataset for data and simulation	105
6.14	Pre-fit yields of the $p_T^{\text{miss}}$ distribution in the $H_T^{\text{miss}}/p_T^{\text{miss}}$ sideband for the combined boosted and resolved categories for the $t\bar{t}H$ and $VH$ processes, using the full Run-2 dataset for data and simulation	106
6.15	Pre-fit yields of the $p_T^{\text{miss}}$ distribution in the tight $\tilde{\omega}_{\min}$ sideband for the combined boosted and resolved categories for the $t\bar{t}H$ and $VH$ processes, using the full Run-2 dataset for data and simulation	106

---

## LIST OF FIGURES

---

6.16 Pre-fit yields of the $p_T^{\text{miss}}$ distribution in the loose $\tilde{\omega}_{\min}$ sideband for the combined boosted and resolved categories for the $t\bar{t}H$ and $VH$ processes, using the full Run-2 dataset for data and simulation . . . . .	106
6.17 An infographic showcasing the role of each analysis region in the final fit . . . . .	107
6.18 Expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the $t\bar{t}H$ category, for both the individual subcategories, and the combination of them, for each data-taking year in Run-2 . . . . .	109
6.19 Expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the $VH$ category, for both the individual subcategories, and the combination of them, for each data-taking year in Run-2 . . . . .	110
6.20 Expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the $ggH$ category, for both the individual subcategories, and the combination of them, for each data-taking year in Run-2 . . . . .	111
6.21 Expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the monojet category, for both the individual subcategories, and the combination of them, for each data-taking year in Run-2 . . . . .	112
6.22 Expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction $\mathcal{B}(H \rightarrow \text{inv.})$ and the corresponding profile likelihood ratio as a function of it, for both the individual categories that target a specific production mode, as well as the combination of them, for the 2016 dataset . . . . .	113
6.23 Expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction $\mathcal{B}(H \rightarrow \text{inv.})$ and the corresponding profile likelihood ratio as a function of it, for both the individual categories that target a specific production mode, as well as the combination of them, for the 2017 dataset . . . . .	113
6.24 Expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction $\mathcal{B}(H \rightarrow \text{inv.})$ and the corresponding profile likelihood ratio as a function of it, for both the individual categories that target a specific production mode, as well as the combination of them, for the 2018 dataset . . . . .	114
6.25 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 . . . . .	116

## LIST OF TABLES

TABLE	Page
2.1 The most recent searches for invisibly decaying Higgs bosons with 2016 data from CMS, and the upper limits on the $H \rightarrow \text{inv.}$ branching ratio at 95 % confidence level achieved . . . . .	18
4.1 The integrated luminosity delivered by the LHC during Run-2 which were recorded and certified by CMS . . . . .	46
4.2 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 . . . . .	47
5.1 The requirements for a jet to pass tight identification for data taken in 2016, and to Monte Carlo events emulating that year . . . . .	67
5.2 The requirements for a jet to pass tight identification for data taken in 2017, and to Monte Carlo events emulating that year . . . . .	68
5.3 The requirements for a jet to pass tight identification for data taken in 2018, and to Monte Carlo events emulating that year . . . . .	69
5.4 Requirements used to define the veto and tight electron IDs in the barrel region (supercluster $ \eta  \leq 1.479$ ) . . . . .	69
5.5 Requirements used to define the veto and tight electron IDs in the end cap region (supercluster $ \eta  > 1.479$ ) . . . . .	70
5.6 Requirements used to define the loose and medium photon IDs in the barrel region of the detector ( $ \eta  \leq 1.479$ ) . . . . .	71
5.7 Requirements used to define the loose and medium photon IDs in the end cap region of the detector ( $ \eta  > 1.479$ ) . . . . .	71
6.1 Cross sections of the $H \rightarrow \text{inv.}$ signal processes in the analysis . . . . .	77

## LIST OF TABLES

---

6.2	The values of factor $f_{\text{var}}$ for each QCD renormalisation and factorisation scale variation as calculated for the $t\bar{t}$ background samples. These factors remove the effect of a change in the normalisation of the event yields. . . . .	86
6.3	Categorisation of the $t\bar{t}H$ , $VH$ and $ggH$ production modes in the analysis . . . . .	93
6.4	The trigger thresholds required for events to enter the signal region in each data taking year . . . . .	99
6.5	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 . . . . .	101
6.6	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 . . . . .	101
6.7	Definitions of the data sidebands used to determine the QCD multijet background in the signal region for the $t\bar{t}H$ and $VH$ categories . . . . .	104
6.8	Definitions of the data sidebands used to determine the QCD multijet background in the signal region for the $ggH$ categories . . . . .	104
6.9	Definitions of the data sidebands used to determine the QCD multijet background in the signal region for the monojet categories . . . . .	104

## 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.

### 1.1 Evidence for dark matter

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 [100]. 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. 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  $400 M_\odot/L_\odot$  [143]. Further evidence is that the rotation curves of most galaxies are roughly flat [117], 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 [75] to compensate for the decline expressed by visible matter [40, 83]. 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 [31, 32, 98]. Though these approximations include non-visible identifiable matter such as dim stars, black holes and neutron stars, the term “dark matter” 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 of particle physics (SM). 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 [37, 94], since the combination of visible and dark matter drives the measured rotation curves. One small caveat is that 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 current age of the universe. Dark matter production is postulated to have occurred only in the early universe via a thermal freeze-out mechanism (see Chpt. 2.2). 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 provided by coalescing dark matter, then merge to form larger structures [142].
- 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 substantial dark matter present, interaction with itself and the visible matter surrounding it was minimal at most [50]. 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 stellar system, visible matter frequently collides, dissipating angular momentum and collapsing into a disc.

### 1.1.1 Alternative theories to dark matter

Though little is intrinsically known about dark matter since all evidence stems from its gravitational influence, there are alternative theories that may explain the observations presented above. However, the scientific community can exclude many of these. Mismeasurements of the amount of baryonic matter such as neutrinos, neutrons, and interstellar gas are among the simpler propositions.

The neutrino flux from stars [REF] and the cosmic neutrino background [141] are precise and well-tested. Even considering the upper limits on neutrino masses [114], they cannot make a significant contribution to the dark matter content in the universe. This is even discounting their highly relativistic nature, where myriad experimental evidence suggests dark matter is cold.

One can also use the Cosmic Microwave Background to calculate the average photon and neutrino densities, and Big Bang Nucleosynthesis calculations to determine the baryonic matter density (see Ref. 80 for results with the latest Planck mission data). These can be compared to other measurements, e.g., mass-to-light ratios averaged across the universe, to reveal a discrepancy [68].

Neutrons cannot contribute to dark matter because isolated neutrons are unstable, decaying in a matter of minutes [138]. Transforming into charged protons and electrons, they interact strongly with light and therefore contribute to the luminous matter content.

Modified Newtonian dynamics (MOND) is a hypothesis that aims to explain phenomena typically associated with dark matter instead by modifications to Newton's laws of motion. There exist many theories and interpretations derived from this principle, though any one strand that tries to explain an observation usually fails to satisfy other phenomena or apply to length scales that general relativity may predict well. For example, observations of the Bullet Cluster [50] have discredited many popular MOND models.

## 1.2 Overview of dark matter searches

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. The detection of dark matter can be classified into three distinct methods with unique signatures (paired with a visual summary in Fig. 1.1):

- **Direct:** dark matter may interact with visible matter on small scales, scattering SM particles [122]. The recoil these SM particles experience could be detected by highly-sensitive, low background experiments such as LUX-ZEPLIN (LZ) [11] that specialises in the search for WIMP dark matter at a large 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 [66]. 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 (see Chpt. 2.2) reinforcing the idea that dark matter may exist in these accessible mass ranges. Many beyond the standard model (BSM) theories accommodate dark matter candidates with a diverse spectrum of final states that can be investigated by analysing LHC data.

### 1.2.1 Dark matter searches at the LHC

Since detection at the CMS experiment from the production mechanism is the focus of this thesis, it is important to establish the current state of dark matter searches at the Large Hadron Collider (LHC); the world’s most powerful particle accelerator that provides the infrastructure for CMS to collect data. Both machines are described in detail in Chpt. 4, and as such only a summary will be provided here. 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 supersymmetry (SUSY) [113],

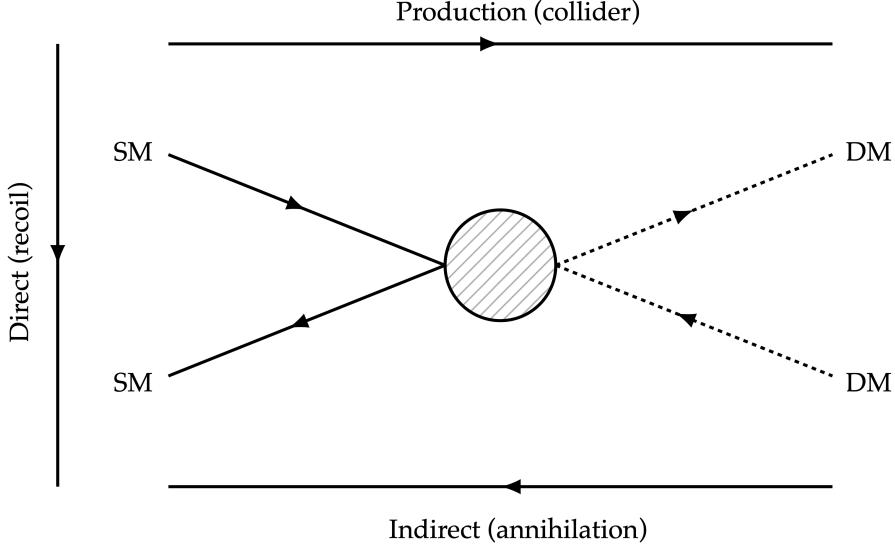


Figure 1.1: A visual representation of the three main types of dark matter detection: direct (dark matter recoiling from standard model particles); indirect (annihilation of dark matter); and production (dark matter created in high energy physics collisions).

sterile neutrinos [74], and Kaluza-Klein states [89] 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. 1.2 illustrates the masses and interaction cross sections of many dark matter candidates. Weakly Interacting Massive Particles (WIMPs) (highlighted by the purple rectangle) are the subject of many searches at the LHC since the expected mass ranges and cross sections are accessible there.

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 (in which I made contributions) [102, 126]. Searches for specific supersymmetric particles in a variety of decay modes have been conducted by both experiments [44]. 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 [113], 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 inter-

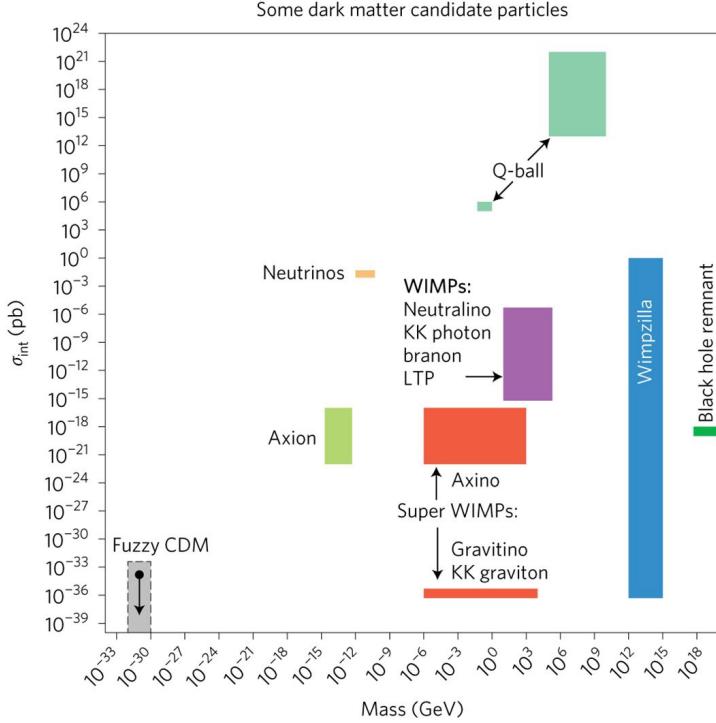


Figure 1.2: The expected masses and interaction cross sections of some dark matter candidates. The LHC, with its centre of mass energy of 13 TeV, is best suited to targeting WIMPs. Figure acquired from Ref. 66.

pretations and approaches for discovery, many others have also been explored at the LHC. From microscopic black holes [105], to [OTHERS], there are extensive propositions that have the potential to uproot the standard model. The analyses above are usually characterised by large “missing” transverse momentum (explained further in Chpt. 5.1.4), a quantity that represents the momenta of particles invisible to the detector, such as dark matter. In Chpt. 3, I discuss in detail a search for dark matter that utilises this variable.

Precision measurements of standard model parameters is the other method often consulted in the hopes of attributing anomalies to new physics. For example, attempts to explain anomalies in the  $b \rightarrow s$  transition include dark matter candidates [45, 140]. In Chpt. 6, I investigate how the measurement of the Higgs boson to invisible state branching ratio can accommodate dark matter.<sup>1</sup>

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

<sup>1</sup>Maybe add another example of anomalies in precision measurements

## 1.2. OVERVIEW OF DARK MATTER SEARCHES

---

to discover, and add to humanity’s collective wisdom. With  $137\text{ fb}^{-1}$  at  $\sqrt{s} = 13\text{ TeV}$  from CMS during Run-2 of the LHC, there is great potential to constrain some of the properties of dark matter. This thesis showcases the motivation for, and results of, searches for invisibly decaying Higgs bosons and semi-visible jets with the full Run-2 dataset collected by the CMS experiment.<sup>2</sup>

---

<sup>2</sup>Add another figure or two somewhere to help separate the text? Maybe an astronomical image of the Bullet Cluster/dark matter distribution in a cosmological object, or a gravitational lensing ring from dark matter



## 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. The experimental chapter 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. 3. Before either of which, the theoretical and phenomenological motivations must be understood to corroborate the need for them 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. Theoretical descriptions of dark matter that best fit the relic density and astrophysical observations will then be discussed. Specific interpretations in the forms of semi-visible jets and invisibly decaying Higgs bosons are examined that provide the background for the respective analysis chapters.

### 2.1 The standard model of particle physics

The standard model of particle physics (SM) 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 aspects of a single *electroweak* force, unified above an energy  $\mathcal{O}(100 \text{ GeV})$ . All

of the elementary particles—the quarks, leptons, gauge bosons, and the Higgs boson—and their interactions with each other are considered in the standard model. The mathematical representation is often [given] in terms of the Lagrangian density

$$(2.1) \quad \mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{Gauge}} + \mathcal{L}_{\text{Fermions}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{Higgs}}$$

before the manifestation of electroweak symmetry breaking, where each term can be expanded to yield

$$(2.2) \quad \begin{aligned} \mathcal{L}_{\text{SM}} = & -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ & + i\bar{\psi} \not{D} \psi + \text{h.c.} \\ & + \bar{\psi}_i y_{ij} \psi_j \phi + \text{h.c.} \\ & + |D_\mu \phi|^2 - V(\phi) \end{aligned}$$

The boson fields and their couplings are represented by  $F_{\mu\nu} F^{\mu\nu}$  and  $D$ , respectively. Feynman slash notation is used, i.e.,  $\not{D} \equiv \gamma^\mu D_\mu$ , with  $\gamma^\mu$  as the Dirac “gamma” matrices. The fermion fields are denoted by  $\psi$ , with the Yukawa coupling  $y_{ij}$ .<sup>1</sup> Finally, the Higgs field is represented with  $\phi$  and potential  $V(\phi)$ .

- Give an overview of the fundamental forces and particles in a QFT/group theory context.
- May have to mention chirality and helicity, briefly, in relation to “handedness” of particles.
- Discuss the standard model in some amount of detail, emphasizing certain aspects as they relate to the Higgs field & boson.
- Mention resonances and widths. Each decay mode of a particle contributes a partial width (which determines its branching ratio:  $\mathcal{B}$  = partial width/total width).
- Talk in some detail about the Higgs mechanism as that will inform the Higgs to invisible section. Mention Yukawa coupling to fermions, and that it’s dependent on the squared mass of the decay products. Lends credence to suppression of direct decay to neutrinos (assuming they even couple to Higgs).

---

<sup>1</sup>Make sure this is correct.

### 2.1.1 Electroweak symmetry breaking and the Higgs mechanism

### 2.1.2 Limitations of the standard model

Despite the standard model providing precise predictions of three of the four fundamental forces and the particles that they interact with, there are many experimental observations that it cannot currently explain. Neutrino masses, dark matter, dark energy, and gravity all escape its description.

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 intermediate vector, and Higgs, bosons of  $\mathcal{O}(100 \text{ GeV}/c^2)$  are much smaller than than the Planck mass of  $\mathcal{O}(10^{19} \text{ GeV}/c^2)$ . The mass term for the Higgs boson is  $m^2\phi^\dagger\phi$  in the SM. Invariance under a gauge or global symmetry in the Higgs field  $H$  leads to the mass being open to radiative corrections up to the Plank scale.<sup>2</sup> It appears that, in nature, these very large terms cancel to give the familiar  $m_H = 125 \text{ GeV}/c^2$  [4, 47]. 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 BSM theories like supersymmetry 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  $\mathcal{O}(\text{TeV})$  scale is required, with new particles coupling to the Higgs field to prevent these corrections from being unreasonably large [79]. Arguments such as this give credence to new physics being discoverable at particle accelerators such as the Large Hadron Collider.

## 2.2 Theoretical motivations for, and descriptions of, dark matter

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

---

<sup>2</sup>Justify this more?

annihilated seeing that the cosmos was still small. Inevitably, the universe expanded and cooled; the temperature became too low to allow significant production [29]. Matter was further separated and the dark matter annihilation rate decreased, leaving 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<sup>3</sup> can be found in literature [34, 85], with the WIMP Miracle as a consequence: with relatively few assumptions, the correct dark matter relic abundance can be recovered by requiring a WIMP mass of  $\mathcal{O}(\text{GeV-TeV})$ , dependent on the annihilation cross section. This is a range 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. 2.1.

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 time-dependent  $n_\chi$ —where dark matter is modelled as a weakly-interacting, diffuse gas of particles—suggest  $x_f \sim 20$  [34, 111]. 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.

Another popular mechanism, targeting low-mass dark matter, is the *freeze-in* process [87, 108]. 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 energy to produce dark matter in any meaningful quantity. The relic abundance can therefore be claimed 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.<sup>4</sup>

---

<sup>3</sup>Is it worth going through the derivation, maybe as an appendix? It might be superfluous considering it's already described in literature.

<sup>4</sup>Could also briefly describe other interpretations of dark matter relic abundance? For all of these cases, not sure how much detail I should go into, in terms of quantitative derivations.

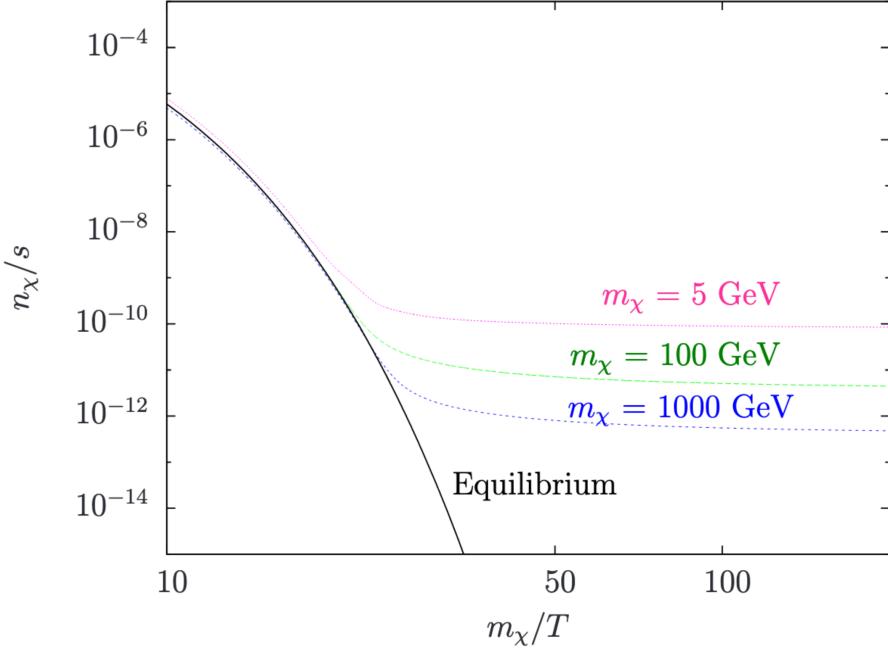


Figure 2.1: A measure of the comoving number density of WIMP dark matter as a function of time with projections for different particle masses. A higher mass must be balanced by a larger annihilation cross section to achieve the correct relic density, to which it tends asymptotically from the point of decoupling. The black curve represents the scenario in which dark matter remains in equilibrium with the standard model. Figure taken from Ref. 88.

## 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  $H \rightarrow \gamma\gamma$  channel in 2012—*independently by both CMS [47] 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 [77, 86, 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 [124] or ATLAS [3]. Constraints on its other properties have also been placed, such as its resonance width and branching ratios  $\mathcal{B}$  to several final states. 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  $H \rightarrow \text{inv.}$  branching ratio aims to do just that.

The only SM process in which Higgs boson can decay invisibly<sup>5</sup> is  $H \rightarrow ZZ \rightarrow 4\nu$  with a branching ratio of  $\mathcal{O}(0.1\%)$  [92]. The leading observed experimental upper limits on this measurement are 19 % from CMS [133] and 26 % from ATLAS [2], far higher than the predicted value. If undiscovered invisible particles, perhaps dark matter, couple to the Higgs field the branching ratio will be enhanced.<sup>6</sup> Experimental evidence shows the coupling strength to proportionally follow the mass of the particle, as verified in ATLAS and CMS' latest measurements [136]. A considerably large enhancement may allow for this process to be observed at the LHC. At the very least, a more accurate constraint on the branching ratio is able to exclude some models of dark matter, such as those described in Refs. 72, 99.

There is no reason to assume dark matter does *not* interact with the Higgs field, since it bestows mass to all known elementary particles (a small caveat, perhaps, being neutrinos).<sup>7</sup> Higgs “portal” models have been theorised that connect the visible sector of the standard model to a dark sector where particle dark matter resides [22, 26]. Certain models also predict a detectable presence at the LHC from a sufficient production rate [39], perhaps even with data obtained during Run-2 [7].

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

---

<sup>5</sup>A direct decay to neutrinos is possible if they acquire their mass from the Higgs field. But as the coupling is of a Yukawa form and the upper limit on the SM neutrino masses is very small, the branching ratio is expected to be heavily suppressed.

<sup>6</sup>Do I need to give a more mathematical motivation for the BR being enhanced/what kind of values the BR is expected to be from various DM models?

<sup>7</sup>Do I need to give some mathematical motivation as to *why* dark matter would couple to the Higgs? Or is the fact that it has mass enough justification?

## 2.3. MEASURING THE BRANCHING RATIO OF INVISIBLY DECAYING HIGGS BOSONS

processes are shown in Fig. 2.2.<sup>8</sup> They have very different characteristics, production rates, and event signatures, complementing each other and allowing a single analysis 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. 4.3.2) can reliably connect these back to the individual quark decays so one has some certainty of the process they are observing.

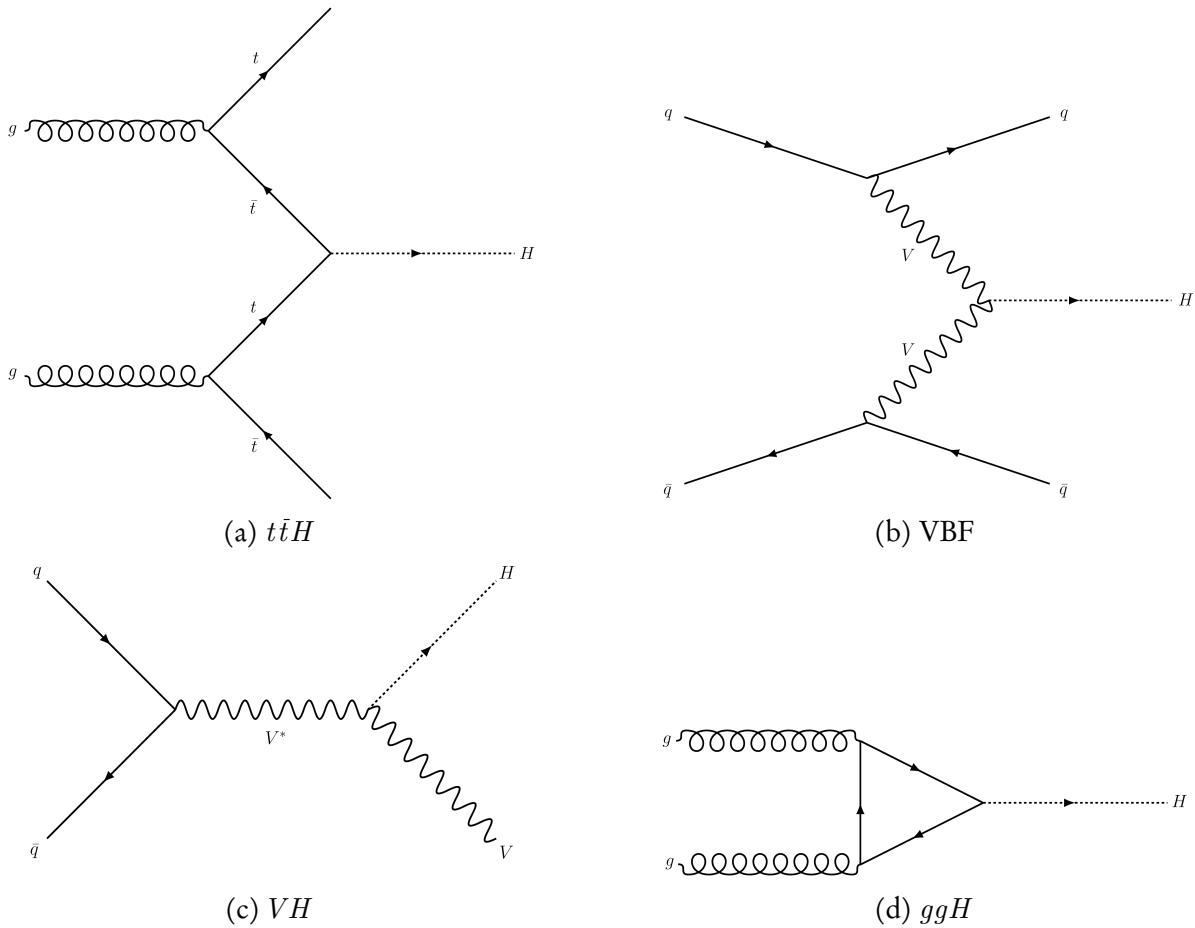


Figure 2.2: A subset of the Feynman diagrams for the four predominant production mechanisms of the Higgs boson at the LHC.

<sup>8</sup> Additional diagrams for ggF are a square top loop (gluons join the two vertices on the left), where the Higgs exits at one of the vertices on the right and a gluon exits at the other. For ZH, can also have  $gg \rightarrow ZH$  as well as  $pp \rightarrow ZH$ . Not sure if it's worth showing them or just describing them.

### 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 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>9</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  $t\bar{t}H$ , two  $t\bar{t}$  pairs are produced. A top quark  $t$  from one pair and antiquark  $\bar{t}$  from the other 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  $\bar{t} \rightarrow \bar{b} W^-$ ) [138]. 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 is 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  $p_T^{\text{miss}}$  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 two leptons (the *dileptonic* channel). The magnitude and direction of the  $p_T^{\text{miss}}$  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. Recently-developed algorithms can assist in this case by inspecting AK8 jets—objects clustered with the anti- $k_T$  algorithm using a radius parameter of 0.8 instead of the usual 0.4. The DEEPAK8 tagger, explained in Chpt. 6.6.1, is used in the analysis to classify 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

---

<sup>9</sup>These coordinates are described in Chpt. 4.3.1.1.

of the parent boson would distinguish the hadronic channel.  $b$ -taggers can be exploited if the decay is to a  $b$  quark, i.e.,  $W^+ \rightarrow \bar{b} q_u$ ,<sup>10</sup>  $W^- \rightarrow b \bar{q}_u$ , or  $Z \rightarrow b \bar{b}$ . Single lepton channels are possible for  $WH$  and dilepton for  $ZH$ . For a boosted  $V$ , one expects the products to be collimated into a single AK8 jet, at least in the hadronic channel. As with  $t\bar{t}H$ , we take advantage of DEEPAK8 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  $\mathcal{B}(H \rightarrow \text{inv.})$  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 at 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  $p_T^{\text{miss}}$ .

### 2.3.2 Results of previous searches

Many previous analyses have investigated the  $H \rightarrow \text{inv.}$  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. 2.1<sup>11</sup> by the upper limits attained compared to the other mechanisms. In Ref. 133, 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 [106].

## 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 [104] and dijet [132] searches are able to exploit this, for example. However, no sign of WIMPs have

---

<sup>10</sup>I'm not sure what the symbol is (if there even is one) for an up-type quark. Maybe  $q_{u,c}$  to be more specific, or  $q_{\uparrow}$ ?

<sup>11</sup>Might be good to add a row with the combined (bar  $t\bar{t}H$ ) result with only 2016 data.

Targeted mode	Analysis	Final state	Observed UL	Expected UL
VBF	Ref. 133	VBF-jets + $p_T^{\text{miss}}$	33 %	25 %
$ZH(Z \rightarrow \ell\ell)$	Ref. 130	$Z(\rightarrow \ell\ell) + p_T^{\text{miss}}$	40 %	42 %
$t\bar{t}H$	Ref. 82	$t\bar{t}(\rightarrow \text{jets}/\ell/\ell\ell) + p_T^{\text{miss}}$	46 %	48 %
$VH(V \rightarrow q\bar{q})$	Ref. 127	$V(\rightarrow q\bar{q}) + p_T^{\text{miss}}$	50 %	48 %
$ggH$	Ref. 127	jets + $p_T^{\text{miss}}$	66 %	59 %

Table 2.1: The most recent searches for invisibly decaying Higgs bosons with 2016 data from CMS, and the upper limits (UL) on the  $H \rightarrow \text{inv.}$  branching ratio at 95 % confidence level achieved. Only the VBF analysis is a dedicated search, while the others are interpretations of the results obtained in their respective primary analyses. For the  $t\bar{t}H$  analysis, the hadronic (jets), semi-leptonic ( $\ell$ ), and dileptonic ( $\ell\ell$ ) channels were combined.

been observed yet. Thankfully, a boundless supply of alternative theories exist, with possible signatures equally as varied. Though the  $p_T^{\text{miss}}$  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<sup>12</sup> scenarios [137], but communicate with the visible sector through a portal interaction. 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 can be explained for the dark matter presence and relic density arising from a baryon-like asymmetry.

Proposed in Refs. 64, 65, a strongly-coupled dark sector in a Hidden Valley is imagined with interactions analogous to QCD.<sup>13</sup> 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) mediator. 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. 2.3. In the LHC, protons could collide at energies high enough to access the dark sector. From either

<sup>12</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.

<sup>13</sup>Do I need to mention that this dark sector is  $SU(2)_{\text{dark}}$ , and write the lagrangian for how it couples to SM via dark weak force?

the resonant production of a  $Z'$  or exchange of a  $\Phi$ , dark quarks  $\chi^{14}$  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*.

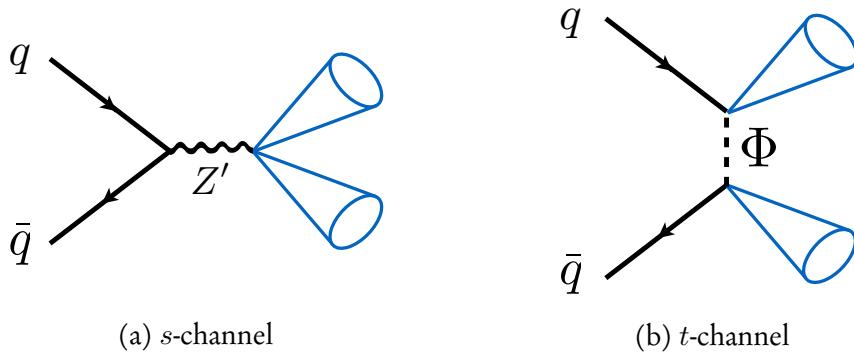


Figure 2.3: Example Feynman diagrams for the two main production modes of semi-visible jets. A  $Z'$  boson mediates the *s*-channel process while a bi-fundamental  $\Phi$  mediates the *t*-channel process. Figure from Ref. 65.

### 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  $m_{\text{jj}}$  or transverse mass  $m_{\text{T}}$ .
- $\alpha_{\text{dark}}$ : In Ref. 65, 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,

$$(2.3) \quad \Lambda_{\text{dark}} = 1000 \text{ [GeV]} \exp\left(\frac{-2\pi}{\alpha_{\text{dark}} b}\right)$$

---

<sup>14</sup>Not to be confused with stable dark matter that is often denoted by the symbol  $\chi$ .

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. 65 describes some of these to be insensitive, a parameter scan over these two variables are necessary in the search described in Chpt. 3.
- $r_{\text{inv}}$ : This is defined as the fraction of produced invisible particles that remain stable, at least over timescales where they interact with a detector. When generating simulated samples,  $r_{\text{inv}}$  can be interpreted as the *probability* of a dark hadron remaining 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  $\vec{p}_{\text{T}}^{\text{miss}}$  is much more intuitive, as demonstrated in Figs. 2.4 and 2.5, respectively. A large value of  $r_{\text{inv}}$  would yield a similar final state to a WIMP search.

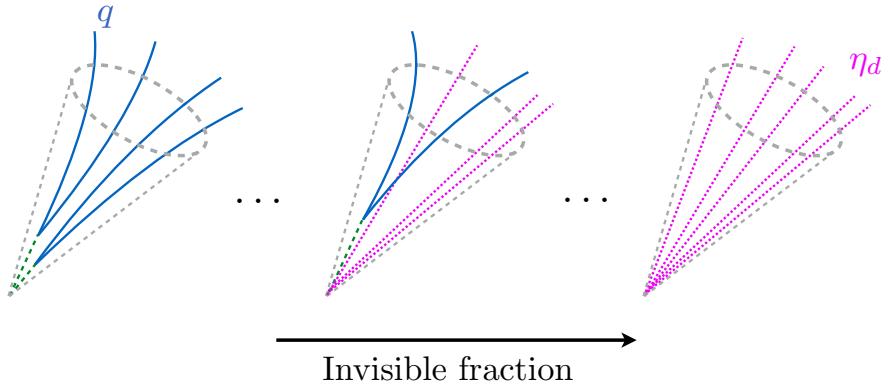


Figure 2.4: The constituents of a semi-visible jet as a function of its invisible fraction  $r_{\text{inv}}$ . Figure taken from Ref. 65.

In the search for semi-visible jets in Chpt. 3, only the  $s$ -channel process has been analysed with LHC data. 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  $\vec{p}_{\text{T}}^{\text{miss}}$  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  $\vec{p}_{\text{T}}^{\text{miss}}$  is also larger and possibly 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

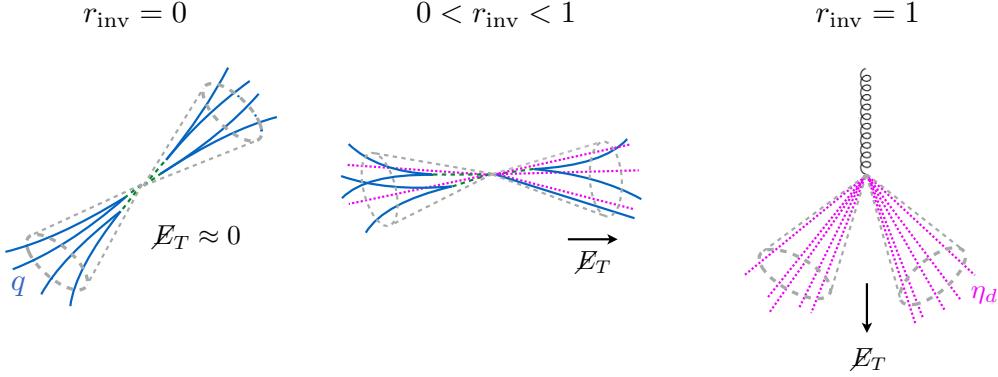


Figure 2.5: The typical direction of the missing transverse energy  $\cancel{E}_T$  (or  $p_T^{\text{miss}}$ ) relative to the semi-visible jets as a function of their invisible fraction  $r_{\text{inv}}$ . Figure from Ref. 65.

it is possible to disentangle signal and background with some certainty.

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 [134] to supersymmetry contexts [126]. These are not considered in Chpt. 3, so the dark hadrons are assumed to decay promptly. Long-lived interpretations have been noted as possible extensions to the search, however.



## SEARCH FOR 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 a novel search for semi-visible jets, characterised by hadronic final states and moderate missing transverse momentum.<sup>1</sup>

### 3.1 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. A dijet final state with moderate  $p_T^{\text{miss}}$  is expected. The transverse mass of the dijet system  $m_T$ —calculated with Eq. 5.6—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

<sup>1</sup>Not 100% sure whether this should go between theory and detector chapters, between detector and objects, or between objects and Hinv. With a fair bit of material and discussion, probably deserves its own (short) chapter rather than maybe being tacked on to theory chapter.

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. 65 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. 118.

My contribution is predominantly the development and understanding of signal simulation in `MADGRAPH5_AMC@NLO` [17] for both the  $s$ -channel and  $t$ -channel aspects of the model. As such, the remainder of the chapter is focused on its discussion. The code I developed 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.

## 3.2 Signal simulation overview

In the main analysis, `PYTHIA8` [135] 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 `CMS`. 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 we favour `MADGRAPH`. 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. 3.3 and 3.4. 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. 3.5. A description of the  $t$ -channel process is also present in the section. The parameters defining the shower were synchronised between myself and the FNAL colleagues. Comparisons between the `PYTHIA` and `MADGRAPH` implementations are discussed in Chpt. 3.6.

For all spectra, no additional object corrections or selections were performed and were direct from the simulation pipeline. Reconstructed jets satisfy the tight identification criteria given in Tabs. 5.1, 5.2, or 5.2 depending on the data taking year. The missing transverse

momentum  $p_T^{\text{miss}}$  contains the type-I correction as outlined in Chpt. 5.2.6. 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.<sup>2</sup>

### 3.3 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 we expect a small percentage of them 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. 3.4. For both the complete event generation, and for hadronisation in the case of MADGRAPH generation, PYTHIA 8.230 was used.

### 3.4 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*.

---

<sup>2</sup>I could just scale the 2016 plots to  $137 \text{ fb}^{-1}$  to avoid writing the above. Or, for realism, also generate 2017 and 2018 samples for the mass points and scale them to their respective lumis, then combine them for a full Run-2 picture.

### 3.4.1 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>3</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. 76, 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. 3.1.

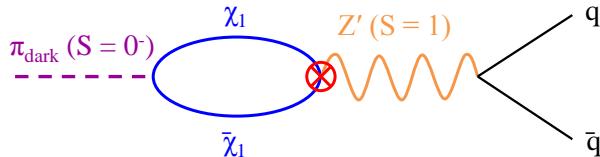


Figure 3.1: A diagram of the mass insertion decay of  $\pi_{\text{dark}}$  mesons in the  $s$ -channel semi-visible jet model.

We encode the dark confinement scale  $\Lambda_{\text{dark}}$  and running dark coupling  $\alpha_{\text{dark}}$ , 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.

#### 3.4.1.1 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 was very

<sup>3</sup>Top quarks are excluded in all cases, since in the scan of model parameters where we predict the greatest sensitivity, the dark mesons are always too light to decay on shell to a  $t\bar{t}$  pair.

small on  $m_{Z'}$ , and instead was 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

$$(3.1) \quad \Lambda_{\text{dark}}^{\text{peak}} = 3.2m_{\text{dark}}^{0.8}$$

which gives the parameter  $\alpha_{\text{dark}}^{\text{peak}} = \alpha_{\text{dark}}(\Lambda_{\text{dark}}^{\text{peak}})$  calculated according to Eq. 2.3. 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, I have introduced an additional point by which to vary it:  $\alpha_{\text{dark}}^{\text{v. high}} = 2\alpha_{\text{dark}}^{\text{peak}}$ . 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. 3.1 also means that we can simply vary  $\Lambda_{\text{dark}}$  (and therefore  $\alpha_{\text{dark}}$ ) as a function of  $m_{\text{dark}}$  for each configuration of parameters in a scan of the model. In the cases where a numerical value of  $\alpha_{\text{dark}}$  is preferred, my signal production repository supports that as well as the string identifiers. Fig. 3.2 illustrates the evolution of  $\Lambda_{\text{dark}}$  as a function of  $m_{\text{dark}}$  for each value of  $\alpha_{\text{dark}}$  we consider.

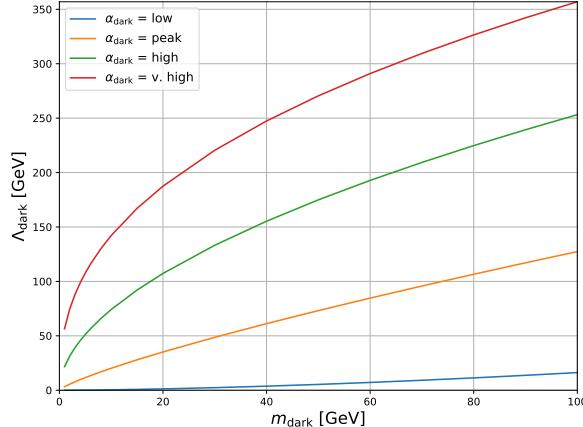


Figure 3.2: The dependence of the dark force scale  $\Lambda_{\text{dark}}$  on the dark hadron mass  $m_{\text{dark}}$  for each value of  $\alpha_{\text{dark}}$  used in the thesis.

### 3.4.2 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 qCut, which we 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  $k_T$  is traditionally used in naming algorithms such as the anti- $k_T$  algorithm. For consistency within this thesis, however,  $p_T$  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. 121:

$$(3.2) \quad d_{ij} = \min(p_{Ti}^{2k}, p_{Tj}^{2k}) \frac{\Delta R_{ij}}{R^2}, \\ d_{iB} = p_{Ti}^{2k}$$

where, in the rapidity-azimuthal ( $y$ - $\phi$ ) 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>4</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. Our choice is the latter 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, we are left with 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.

### 3.4.3 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.

---

<sup>4</sup>This is calculated with Eq. 4.6 where  $\eta$  is replaced with  $y$  from Eq. 4.5.

## 3.5 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  $s$ -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 [15]. 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 [27]. The  $t$ -channel description is completely custom. Both were initially acquired by the analysis team from one of the authors of Ref. 65 at <https://github.com/smsharma/SemivisibleJets>. At the time, it contained only one set of mass points. I have since greatly expanded on both models, allowing any combination of  $Z'$ ,  $\Phi$ , and  $\chi$  masses and couplings. All signal samples produced with **MADGRAPH** that feature in this thesis were 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. 3.4). 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. 2.3. Individually, former parameter allows for the production of two types of dark hadron, described in Chpt. 3.4, while the latter sets the dark colour composition of the dark gluon  $g_{\text{dark}}$ .

### 3.5.1 **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

$$(3.3) \quad k_T^2 = 2 \min(p_{Ti}, p_{Tj}) [\cosh(\eta_i - \eta_j) - \cos(\phi_i - \phi_j)]$$

where  $\eta$  and  $\phi$  are the pseudorapidity and azimuthal angle, respectively. Complete definitions are presented in Chpt. 4.3.1.1. Events with  $k_T < \text{xqcut}$  are not simulated. In essence, the `xqcut` is a measure of the required separation between partons in the event. We set a value of 100 GeV in the analysis for consistency with Ref. 65. This interpretation of a merging scale is similar, but not quite the same, as `PYTHIA`'s discussed in Chpt. 3.4.

The choice of parton distribution function also lives in the run card. As stated above, this changes to reflected the data taking period we emulate. 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.

### 3.5.2 *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 our implementation. Consideration of these supplementary species of particle are beyond the scope of this thesis.

Cross sections for this process are calculated as a function of  $m_{Z'}$ , assuming resonant production and a universal quark coupling.<sup>5</sup> Those given by `MADGRAPH` are far too large and appear unrealistic.

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 `SVJ_<mmediator>-<mdark>-<rinv>-<αdark>`. 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 `SVJ_3000_-20_0.3_peak`. The choice is motivated by Ref. 64 and considerations of the other attributes

---

<sup>5</sup>Do I need to explicitly mention events are weighted by cross section of the model, then all scaled to the integrated lumi?

of the signal provided in this chapter. Fig. 3.3 shows the  $m_T$  in the case where a single free parameter was varied with respect to the benchmark point. A glimpse into the sensitivity of the parameter on the kinematics is evident.<sup>6</sup>

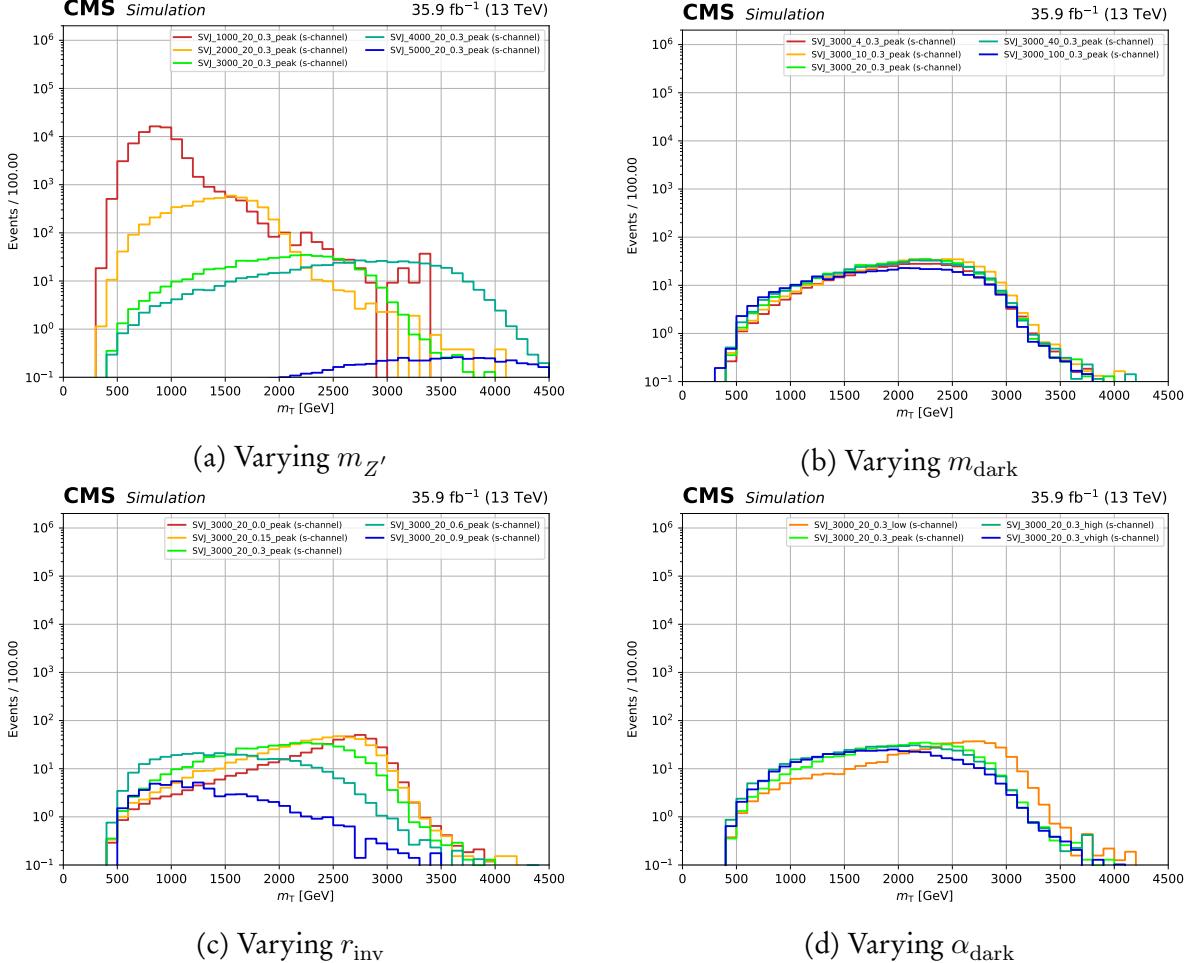


Figure 3.3: Distributions of the transverse mass of the dijet system  $m_T$  for  $s$ -channel semi-visible jet samples emulating the 2016 data taking period. In each panel, one of the free parameters of the model is varied with respect to the benchmark point SVJ\_3000\_20\_0\_3\_peak (bright green line).

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 did 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

<sup>6</sup>Not sure if the legends are too small for the plots in this chapter. Can enlarge them if need be.

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  $r_{\text{inv}}$ .

### 3.5.3 $t$ -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.<sup>7</sup>

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. 65, 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  $p_T^{\text{miss}}$  distribution are shown in Fig. 3.4.<sup>8</sup>

## 3.6 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

---

<sup>7</sup>Is this for charge conservation?

<sup>8</sup>In these plots, just using the MadGraph cross sections for the moment. 4000\_20\_0.5\_peak also looks a bit janky so remaking that MC. Remake plots once that's fixed and settled on cross section. Then, discuss. Also, only looking at MET as it seems the most discriminatory of all the ones I've made so far.

### 3.6. PYTHIA-MADGRAPH COMPARISONS FOR $S$ -CHANNEL SIGNAL

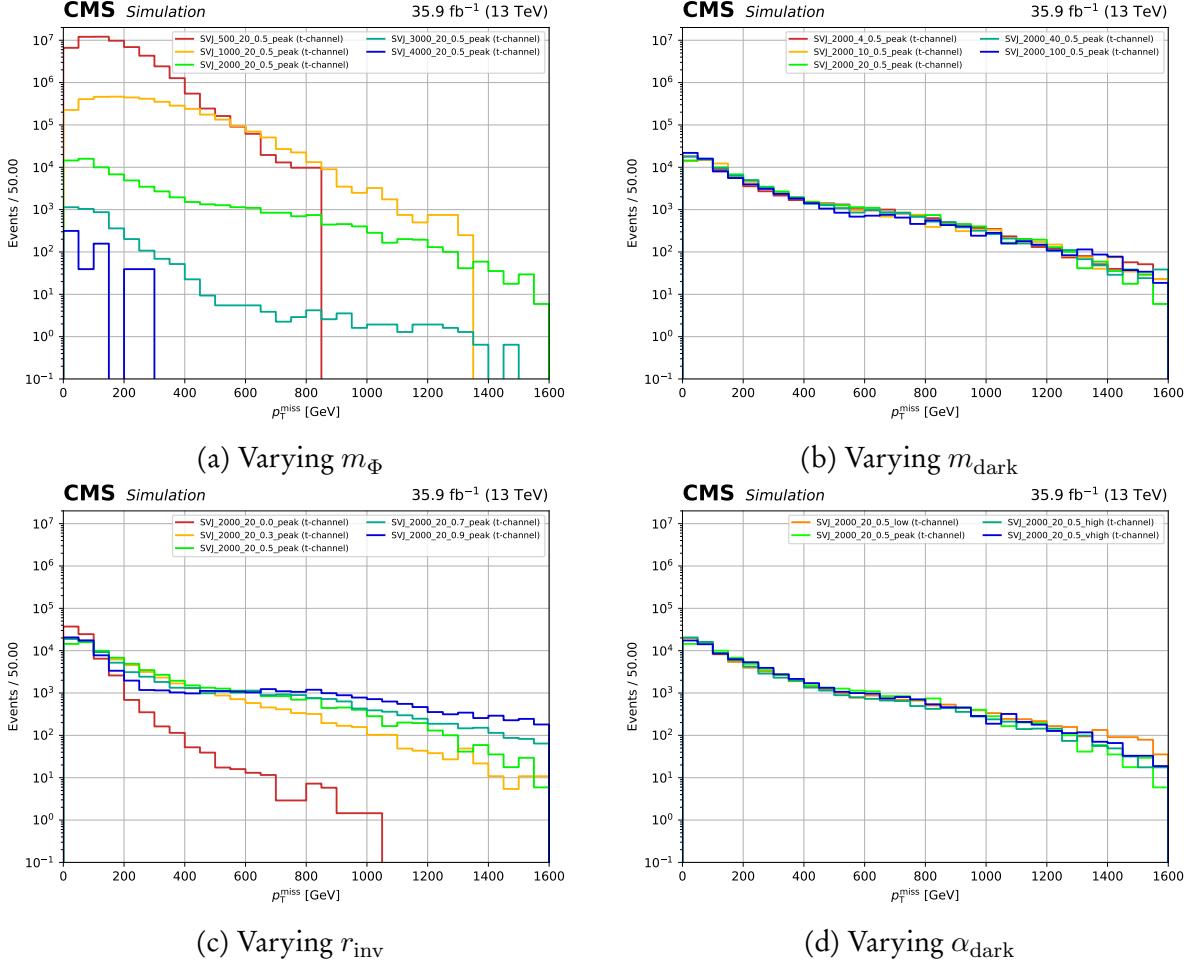


Figure 3.4: Distributions of the missing transverse momentum for  $t$ -channel semi-visible jet samples emulating the 2016 data taking period. In each panel, one of the free parameters of the model is varied with respect to the benchmark point SVJ\_2000\_20\_0.5\_peak (bright green line).

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>9</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

<sup>9</sup>With our semi-visible jets model, around 50 subprocesses exist for the  $s$ -channel mode, while approximately 820 exist for  $t$ -channel.

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. 4.3.3), 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  $p_T^{\text{miss}}$  and two leading jets  $\Delta\phi_{\min}(j_{1,2}, E_T^{\text{miss}})$ . 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. 3.5 showcases SVJ\_1000\_20\_0.3\_peak, SVJ\_3000\_20\_0.1\_peak, and SVJ\_3000\_20\_0.9\_peak, while Fig. 3.6 shows SVJ\_3000\_20\_0.5\_peak, SVJ\_3000\_50\_0.3\_peak, and SVJ\_4000\_20\_0.3\_peak.

In general, the distributions above—as well as the others investigated such as  $p_T^{\text{miss}}$ ,  $H_T^{\text{miss}}$ , and jet  $p_T$ —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  $m_T$  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

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

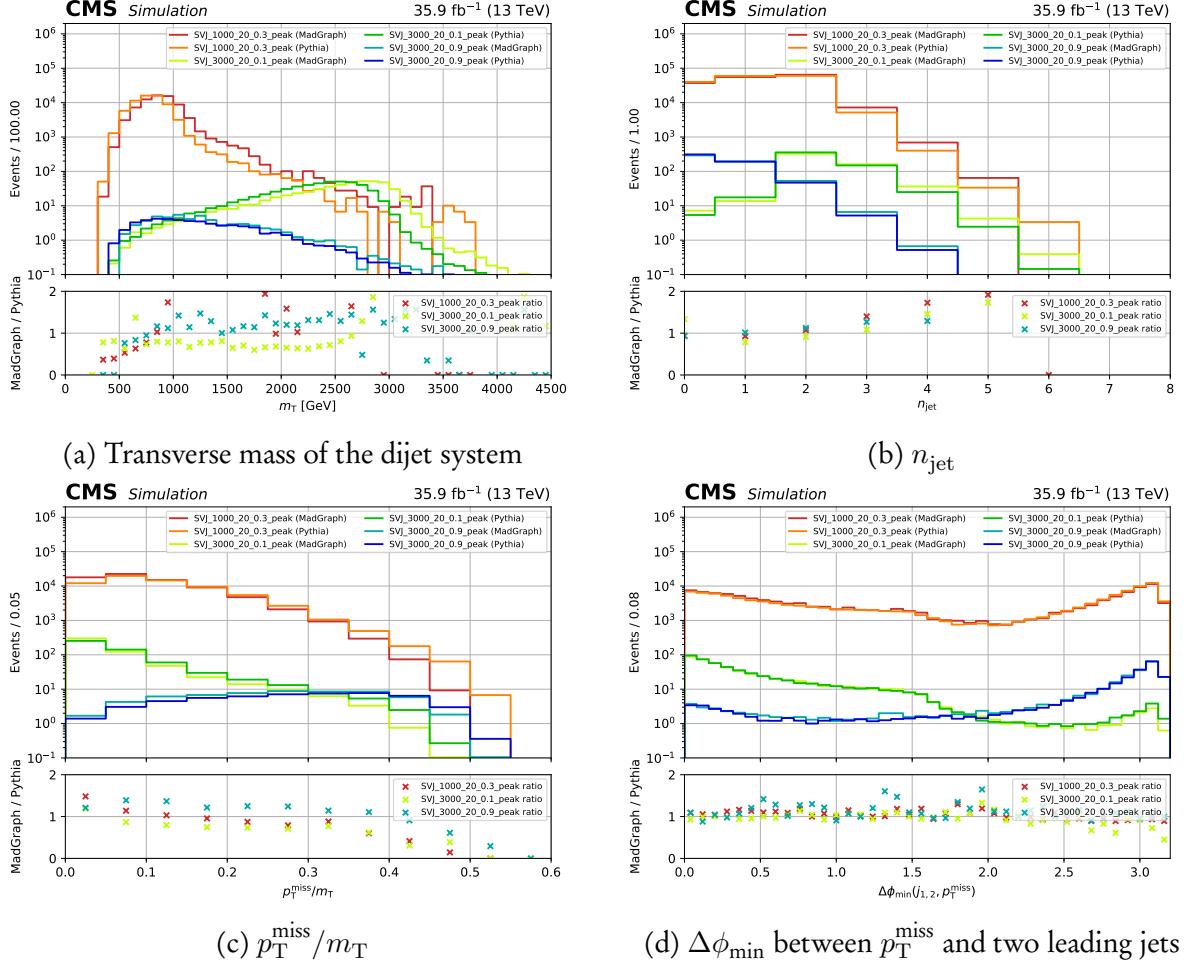


Figure 3.5: Distributions of several observables for the models SVJ\_1000\_20\_0.3\_peak, SVJ\_3000\_20\_0.1\_peak, and SVJ\_3000\_20\_0.9\_peak. Generation in `MADGRAPH5_AMC@NLO` is compared to `PYTHIA8`, with the ratios between them for each model displayed in the respective subplot.

the dark shower prompting 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. 3.5 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  $p_{\text{T}}^{\text{miss}}$  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  $p_{\text{T}}^{\text{miss}}$  would appear to recoil from

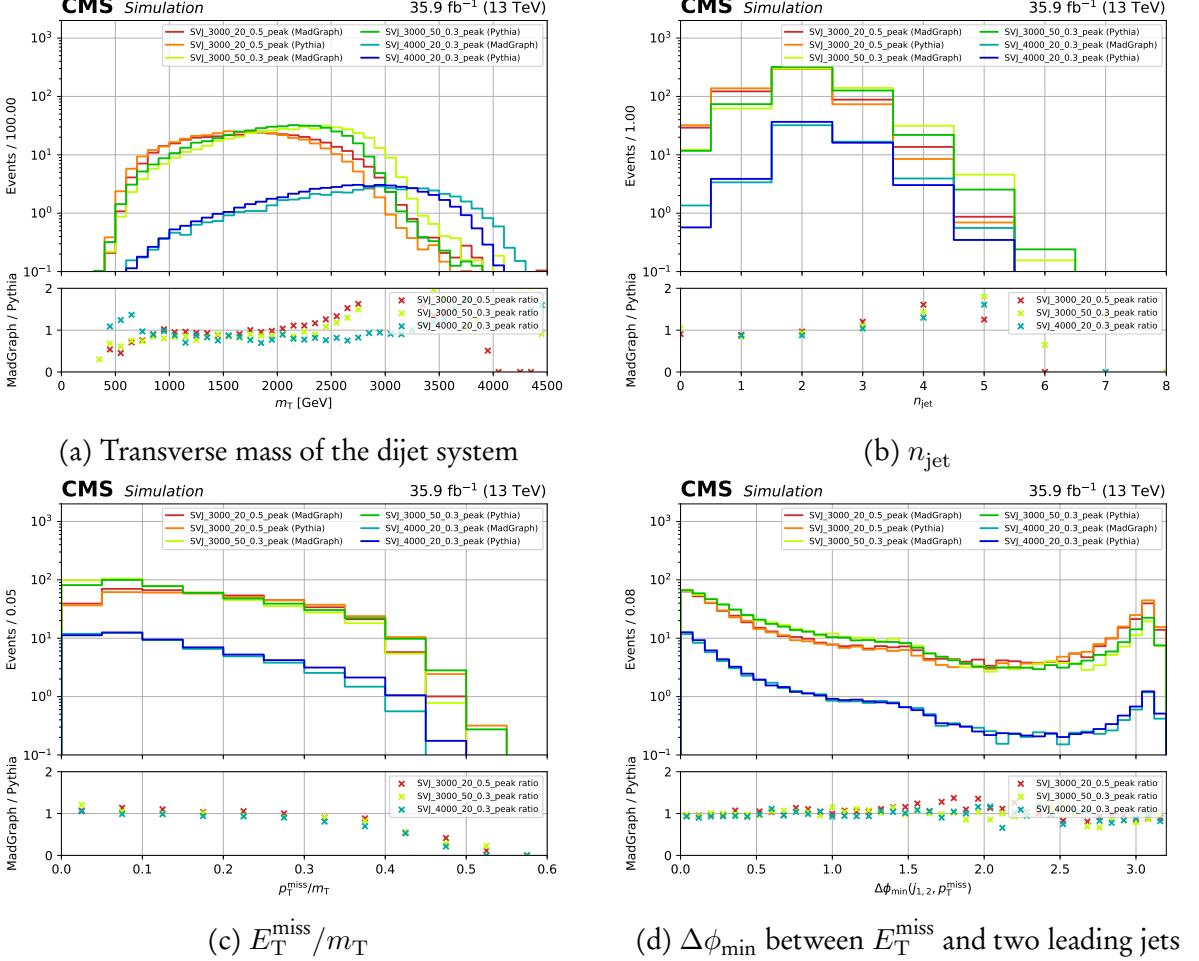


Figure 3.6: Distributions of several observables for the models SVJ\_3000\_20\_0.5\_peak, SVJ\_3000\_50\_0.3\_peak, and SVJ\_4000\_20\_0.3\_peak. Generation in `MADGRAPH5_amc@NLO` is compared to `PYTHIA8`, with the ratios between them for each model displayed in the respective subplot.

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  $m_{\text{T}}$  distribution as the  $p_{\text{T}}^{\text{miss}}$  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. Tuning the merging scale and matching parameters based on  $r_{\text{inv}}$  may yield in better efficiency. Combined with the filters in Chpt. 3.4.3, efficiencies on the order of 15–20 % are expected for most `MADGRAPH`

samples.



## 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 is the organisation that manages the machines discussed and is a pioneer in the high energy physics community. As such, it will be given a short overview. The Large Hadron Collider provides the CMS experiment with proton-proton collision data that is then stored, corrected, and then used by physicists for analysis. Ranging from standard model precision measurements, searches for new physics, 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 I have been a part of that effort during my PhD.

### 4.1 CERN

CERN, the European Organisation for Nuclear Research (*Organisation Européenne pour la Recherche Nucléaire*), is the body responsible for large scale particle and high energy physics projects in Europe. It was founded in 1954 under the *Conseil Européen pour la Recherche Nucléaire* (European Council for Nuclear Research), from where its acronym is derived. CERN's primary site is situated in the canton of Geneva and creeps over the Franco-Swiss

border. Its main purpose today is to provide physical and technological infrastructure for particle and high energy physics experiments. From large scale accelerators to extensive computing farms, CERN has grown into the largest laboratory for particle physics in the world. The organisation also provides a central community for the many researchers, engineers, and technicians to share ideas and collaborate effectively.

The organisation was founded by twelve European member states, the United Kingdom being one of them. It has since expanded to twenty three, encompassing most of western Europe and some of the continent’s east. Many more countries from across the globe are affiliated with CERN in some way, providing researchers, computing resources, and more. The member states and associated members contribute to CERN’s budget, to the tune of 1.2 billion Swiss Francs for the year 2020 [81].

Many important inventions and discoveries can be attributed to CERN and its personnel. Physics accomplishments include observations of weak neutral current interactions in 1973 [90, 91], paving the way for the  $W$  [23, 30] and  $Z$  [24, 28] boson discoveries with the UA1 and UA2 experiments in 1983. The number of light neutrino generations at the Large Electron-Positron Collider (LEP) in 1989 [139], and direct CP violation with the NA48 experiment in 1999 [78], were also observed. Tim Berners-Lee and Robert Cailliau are credited with inventing the World Wide Web—the ubiquitous service for accessing the internet—in 1989/90.

CERN is perhaps most widely known as the home of the Large Hadron Collider, the particle accelerator involved in the discovery of the Higgs boson [4, 47]. More details regarding the machine are discussed in Chpt. 4.2. But CERN is involved in many more undertakings. There are many fixed target experiments that use beams from the PS and SPS such as COMPASS, that studies hadronic structure, and NA62, investigating rare decays of kaons. Experiments like ALPHA and AEGIS use antiprotons from the Antiproton Decelerator to study antimatter in detail. The ISOLDE facility at CERN delivers beams of radioactive ions to perform many nuclear physics experiments.

Concern is given not just to contemporary science, but also to the physics of tomorrow. The experiments and accelerators are frequently upgraded, in particular the LHC improvements documented in Chpt. 4.2.4. Advanced accelerators are also being discussed, such as the Future Circular Collider (FCC). It would use the LHC as a booster, with the final ring having a 90–100 km circumference and up to 100 TeV centre of mass energy. Proposals for the injectants are electron-positron beams (FCC-ee,  $\sqrt{s} = 90\text{--}350 \text{ GeV}$ ) or proton-proton beams (FCC-hh,  $\sqrt{s} = 100 \text{ TeV}$ ). Each option boasts its own merits, and in an integrated scenario the former may be used as an intermediate step toward the latter.

## 4.2 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. 41, detailing the ring itself, infrastructure, general services, and the injector chain.

### 4.2.1 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 [sent] to the linear accelerator LINAC2.<sup>1</sup> This is the start of a long voyage through the accelerator complex (visualised in Fig. 4.1).

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

---

<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.

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 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.

### 4.2.2 Luminosity

The *luminosity* of a particle accelerator is, along with centre of mass energy, a quantitative measure of its performance. It is 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

$$(4.1) \quad \mathcal{L} = \frac{1}{\sigma_p} \frac{dN}{dt}$$

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

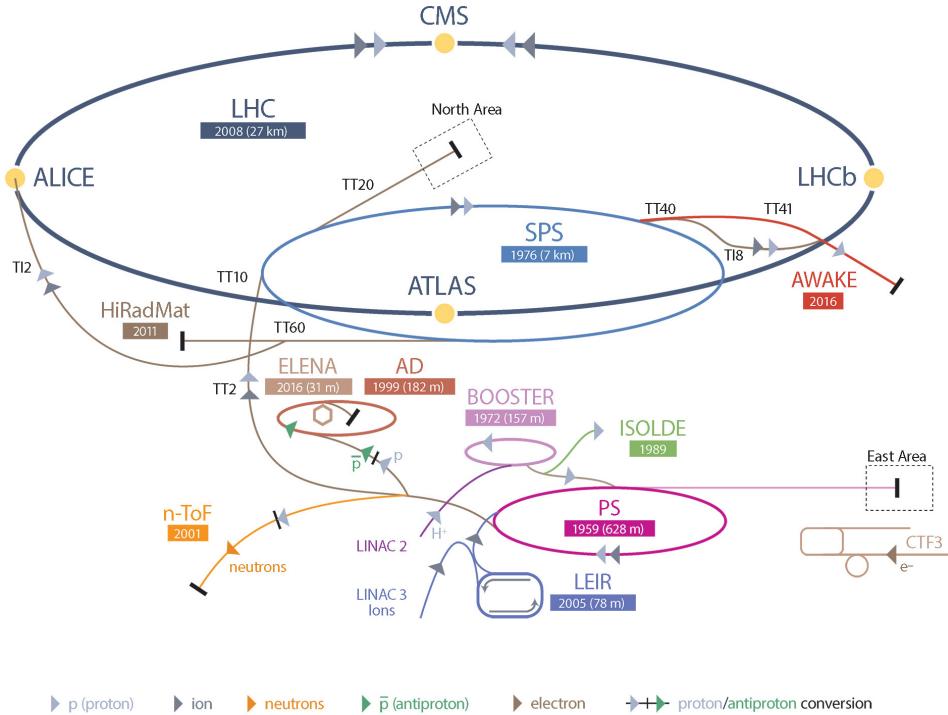


Figure 4.1: A schematic of the CERN accelerator complex. Various particles are shown from their sources to the detectors they are observed at. Figure obtained from Ref. 123.

in each bunch distributed according to a gaussian profile,

$$(4.2) \quad \mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y} \cdot F$$

where  $N_1$  and  $N_2$  are the number of particles in a bunch in beams 1 and 2, respectively,  $N_b$  is the number of bunches per beam,  $f$  is the revolution frequency (11,245 Hz for protons in the LHC), and  $\sigma_x$  and  $\sigma_y$  are the horizontal and vertical sizes of the beam, respectively.<sup>2</sup>

The geometric reduction factor  $F$  is unity for beams colliding head on.<sup>3</sup> 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  $\mu$ rad. As the runs progress, more protons collide and the luminosity decreases. 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  $\mu$ 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

<sup>2</sup>Add the  $H$  factor from the hourglass effect and explain?

<sup>3</sup>Add mathematical formula for F?

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*  $\mathcal{L}_{\text{int.}}$ :

$$(4.3) \quad \mathcal{L}_{\text{int.}} = \int \mathcal{L} dt$$

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 colloquially 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. 4.1 gives the integrated luminosities over Run-2 delivered by the LHC and collected by CMS. Taking  $137.19 \text{ fb}^{-1}$  from this table, and assuming the inelastic  $p_T$  cross section above gives an estimate of  $10^{16}$  collisions recorded by CMS over Run-2 (excluding pileup interactions).

### 4.2.3 Pileup

At the high instantaneous luminosity of the LHC, multiple interactions per bunch crossing—known as *pileup*—are frequent. The expected, or nominal, number of collisions per bunch crossing is one, but because of all the factors that increase the luminosity, many “softer” collisions also take place. 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. 4.2.

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, 46]. At CMS, various methods are available in the calorimeter triggers to subtract the effects of pileup.

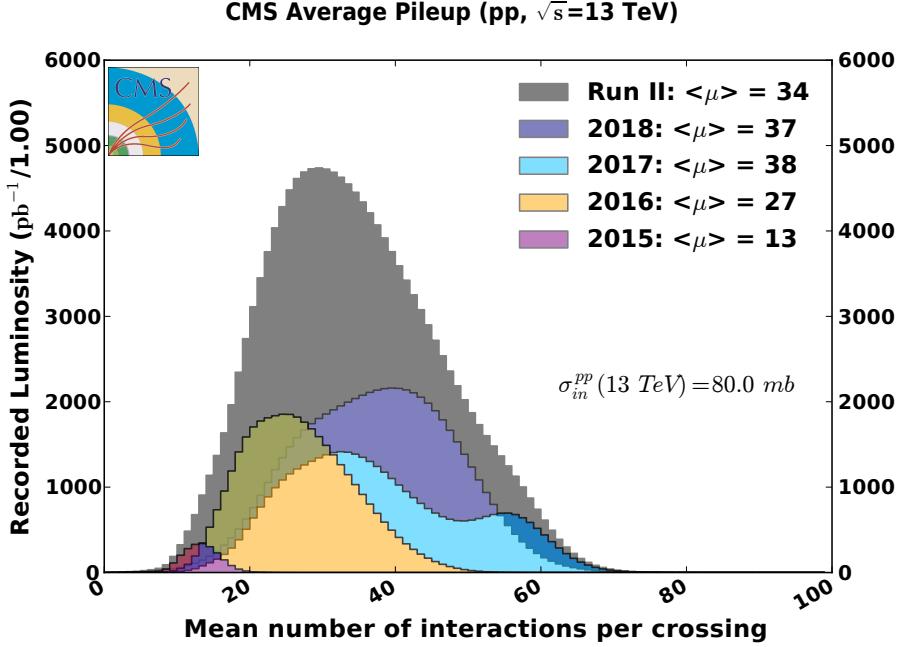


Figure 4.2: The average number of pileup interactions per bunch crossing at CMS during Run-2 of the LHC. Figure obtained from Ref. 51.

#### 4.2.4 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 by 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.

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  $\sqrt{s} = 13 \text{ TeV}$ . While valuable data was taken that year, it was not until 2016 when substantial datasets were amassed. Therefore, when analysts 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— $158.64 \text{ fb}^{-1}$  of  $pp$  collisions delivered,  $146.45 \text{ fb}^{-1}$  of which were recorded by CMS who certified  $137.19 \text{ fb}^{-1}$  suitable for analysis [51, 57]. A breakdown by year is presented in Tab. 4.1<sup>4</sup> with a visual

<sup>4</sup> Maybe add Run-1 column(s)? One for  $7 \text{ TeV}$  (or 2011) and one for  $8 \text{ TeV}$  (or 2012)? Though, it does risk becoming too cluttered.

chart in Fig. 4.3.<sup>5</sup> 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 ( $\text{fb}^{-1}$ )	4.21	40.99	49.79	67.86	162.85
Recorded by CMS ( $\text{fb}^{-1}$ )	3.80	37.80	44.98	63.67	150.25
Certified by CMS ( $\text{fb}^{-1}$ )	2.26	35.92	41.53	59.74	139.45

Table 4.1: 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. 51, 57.

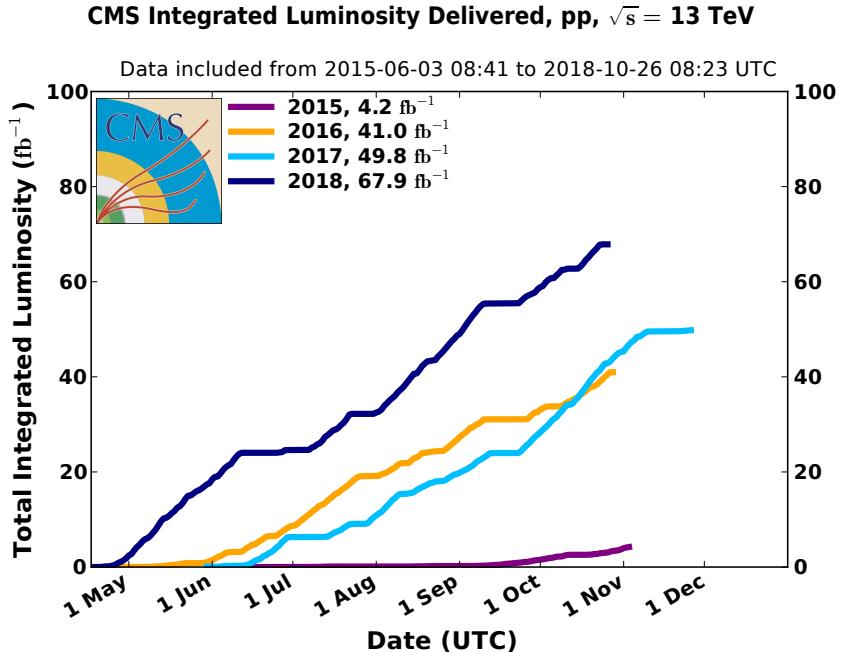


Figure 4.3: The integrated luminosity of  $pp$  collision data delivered to CMS during Run-2 of the LHC. Figure obtained from Ref. 51.

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

<sup>5</sup>As I'm considering with the table, should I replace this figure with the one that also includes Run-1? But then there are lots of lines and things can look less clean.

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. 4.2.<sup>6</sup>

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 4.2: 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. 19, 35.

\* Does not account for the large increase to virtual luminosity from the introduction of crab cavities.

## 4.3 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 [70]. 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 detector, and collaboration as a whole.

---

<sup>6</sup>Should I add a column for the Tevatron, as a comparison to/sense of scale for the LHC’s numbers? Or maybe discuss the Tevatron qualitatively somewhere.

### 4.3.1 The CMS detector

The machine itself—aptly named the CMS Detector—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 with an incredibly small tolerance of a few centimetres. Here, they were joined together to assemble the 14,000 tonne goliath. The finished detector is a hermetic cylinder enclosing the LHC beam pipe, measuring 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 used to generate the magnetic field.

A kaleidoscope through which we 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. 4.4 and a transverse slice through it in Fig. 4.6.

#### 4.3.1.1 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, i.e., the angle subtended from the  $0^\circ$  direction on the polar axis to the particle's position, 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):

$$(4.4) \quad \eta = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right]$$

where  $\theta$  is the angle between the particle's three-momentum  $\vec{p}$  and the positive direction of the beam axis. So as  $\theta \rightarrow 0^\circ$  (the beam line),  $\eta \rightarrow \infty$ . 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 defined:

$$(4.5) \quad y = \frac{1}{2} \ln \left( \frac{E + p_L}{E - p_L} \right)$$

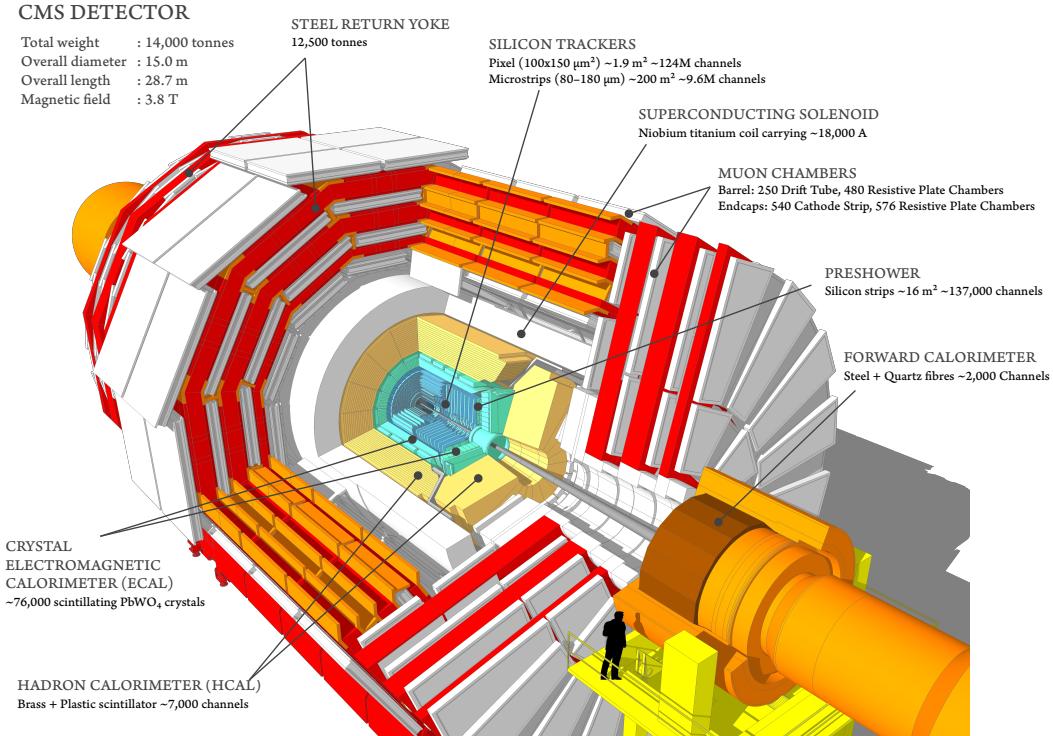


Figure 4.4: A cutaway diagram of the CMS detector with all of the principal components labelled. This detector configuration was used for the 2017–18 data taking years, where the coverage of the pixel detectors in the silicon tracker were upgraded. Image taken from Ref. 119.

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. 4.5.

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

$$(4.6) \quad \Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

It is often used as a distance parameter in jet clustering, or between individual objects in highly-boosted decays.

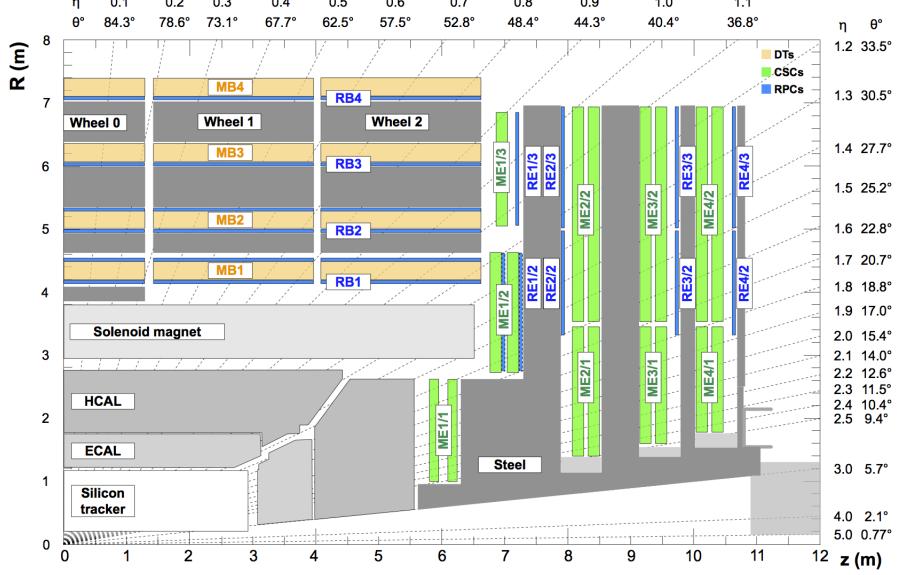


Figure 4.5: A quadrant of the CMS detector illustrating the main subsystems with their radius  $R$ , longitudinal distance  $z$ , and pseudorapidity  $\eta$  from the interaction point. The grey box at  $3 < |\eta| < 5$  and  $11 < z < 12$  m is the hadron forward calorimeter. Figure taken from Ref. 115.

#### 4.3.1.2 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 [63, 101]. 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  $\mu\text{m}$ . 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 pummelling them every collision.

The hit is registered by a pixel or strip via electron-hole pairs created by the ionisation caused from a particle interacting with the silicon. An electric field within the sensor separates them and they drift toward the electrodes at the boundaries. The activity is then read out, where pixels record electrons and strips record holes.

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 indi-

vidual particles. Each is suited to different types of track,  $p_T$  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. The momentum of particles can be measured from the curvature of the track from the magnetic field induced by the solenoid (see Chpt. 4.3.1.5).<sup>7</sup> One of the algorithms, the gaussian sum filter (GSF) method, is used extensively in the HLT for triggering on electrons and photons [18].

#### 4.3.1.3 The ECAL

After the tracker, the particles can interact with the ECAL [38, 60]. 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. 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. 4.3.2).

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. Muons and tau leptons ( $\tau$ ), however, usually penetrate farther. Measuring the energy of electrons and photons is performed as follows:

$$(4.7) \quad E_{e,\gamma} = G(\eta) \cdot F_{e,\gamma} \cdot \sum_i A_i S_i(t) c_i$$

where  $G$  is the global energy scale,  $F$  is the cluster correction (as several hits in a cluster can belong to one object, especially when they electromagnetically shower),  $A$  is the amplitude of the pulse from the readout electronics,  $S$  is the time-dependent response correction from the laser monitoring system that measures the crystal transparency, and  $c$  an intercalibration correction that accounts for the non-uniform response of the crystals and photodetectors.<sup>8</sup> Charged hadrons, muons and taus only deposit a sliver of energy in the ECAL, and so information from the other subdetectors is required in their cases.

---

<sup>7</sup>Should I mention something about resolution as well?

<sup>8</sup>Should I mention something about energy resolution as well?

#### 4.3.1.4 The HCAL

Hadronic particles propagate past the ECAL into the HCAL [61]. The barrel ( $|\eta| < 1.30$ ) and end cap ( $1.30 < |\eta| < 2.96$ ) regions both consist of brass absorber<sup>9</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  $2.96 < |\eta| < 5.19$ ,<sup>10</sup> 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. 4.3.2). The towers project mostly-radially from the interaction point. Their granularity is greatest at low  $\eta$ , and decreases with increasing  $\eta$ . Information from both subdetectors is used 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.

#### 4.3.1.5 The magnet and muon chambers

The three subsystems described above are encased in a massive superconducting solenoid [62], 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,

---

<sup>9</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.

<sup>10</sup>I'm not entirely sure what the bounds of the HF are. Literature seems to define it as  $3.0 < |\eta| < 5.0$ , but in the JEC the edge of the final  $|\eta|$  bin is 5.191.

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 [110] 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. Consequently, constructing a dedicated subdetector is therefore essential.

There are four muon “stations,” as can be seen by the orange strips in Fig. 4.6, 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.<sup>11</sup>

---

<sup>11</sup> Maybe mention the upgrades done between Run-1 and Run-2 (each subsystem), and possibly even the planned upgrades for Run-3 and Phase-2.

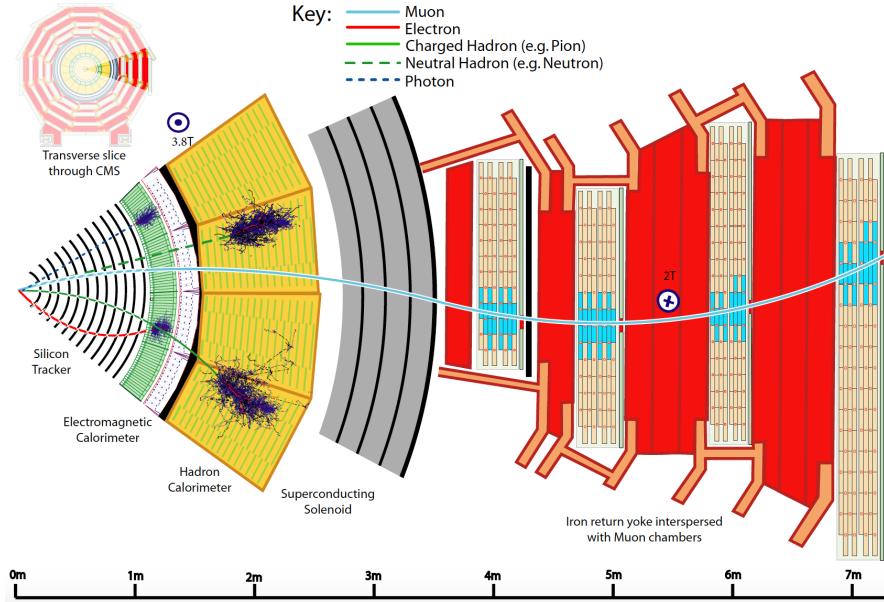


Figure 4.6: A transverse slice through the barrel section of the CMS detector with the main subsystems and components visible. Several particles produced at the primary vertex and their interactions with the detector are also depicted. Figure obtained from Ref. 125.

### 4.3.2 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,  $\mathcal{O}(1 \text{ petabyte s}^{-1})$ . As such, a *trigger* is used to select the events that may be of use to analysers. In CMS, a two stage trigger [33] is used: the Level-1 Trigger (L1T), implemented in the detector hardware; and the High-Level Trigger (HLT) [49], 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.

#### 4.3.2.1 The Level-1 Trigger

The Level-1 Trigger is a set of algorithms (a trigger menu)<sup>12</sup> 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

<sup>12</sup>Do I need to say that different menus are used throughout the data-taking periods and for different injectants (heavy ions, etc.)?

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 bandwidths. The information from the two subsystems are combined into calorimeter towers, and some simple position- and energy-dependent calibrations are applied.<sup>13</sup> 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,  $e$ ,  $\gamma$ ,  $\tau$ ).<sup>14</sup> Additional calibrations are applied to them (for example, in Chpt. 4.3.4), and simple pileup subtraction is performed.<sup>15</sup> 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 is then combined with some of the calorimeter information 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 information gathered—object candidates, energy sums, beam conditions—is provided to it. These triggers may be dependent on the presence a single object or the number of objects of a single class (e.g., one muon, two jets), multiple classes 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  $\mu\text{s}$ .<sup>16</sup> A diagram of this data flow is given in Fig. 4.7.

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 different domains, collecting information from various sources. Data acquisition, data quality, access to the experimental cavern, the L1T (which I have performed on numerous occasions), and the HLT are among them. Experts in these, and more specific systems, are on call on a rotating period. I, myself, have been an on call for Layer-2 of the Calorimeter Trigger several times.

---

<sup>13</sup>Mention something about trigger primitives here?

<sup>14</sup>An overview of the latest Run-2 algorithms for object identification can be found in Ref. 144.

<sup>15</sup>Do I need to describe in detail how jets are formed? Jet seed, chunky donut PU subtraction, etc... Since it may be relevant for JEC.

<sup>16</sup>Mention pre-scaling of triggers, and that as  $\mathcal{L}$  decreases over a fill, the pre-scales change?

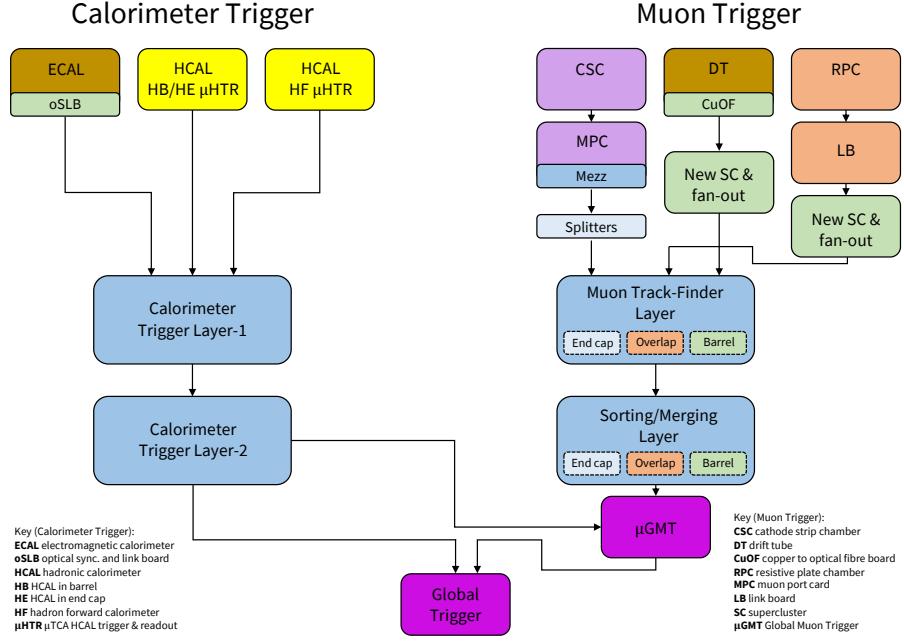


Figure 4.7: A summary of the CMS Level-1 Trigger data flow from the hits recorded by the subsystems to the Global Trigger. Figure reproduced from Ref. 71.

### 4.3.2.2 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 back-log, 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, see Chpt. [REF]) is executed. A larger and more complex trigger menu is available, including the possibility analysis-specific triggers (such as those that target VBF topologies in Chpt. 6). Complex variables such as  $\alpha_T$  can even be calculated and triggered on.

Physics objects are reconstructed further, with algorithms such as anti- $k_T$  used to cluster jets [43]. Additional classification algorithms are also applied to objects, such as the DEEPCSV neural network [128] to identify  $b$ -jets. A global event reconstruction from the PARTICLE FLOW algorithm (PF) [58, 125] 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 of these kind of algorithms are re-run on the retained

events in later stages of postprocessing.<sup>17</sup>

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, back ups 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.

#### 4.3.2.3 Data reduction and compression

Several data “tiers” are used in CMS to serve different purposes. All sorted in ROOT file containers,<sup>18</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 designed 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 used to shrink 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 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 our search for invisibly decaying Higgs bosons.

#### 4.3.3 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

---

<sup>17</sup>If more detail is required here, I should make a separate subsubsection and discuss.

<sup>18</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  $\eta$ .

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 [16] or POWHEG [84, 116] models the hard scattering process, usually at leading order (LO) but sometimes at higher orders. Events then pass through a hadroniser (usually PYTHIA [135] 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 in different years. Material interactions and emulation of the triggers are included. GEANT4 [10, 13, 14] 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.

#### 4.3.4 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. 4.3.2.1, 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>.

[com/eshwen/L1JetEnergyCorrections](https://github.com/eshwen/L1JetEnergyCorrections).<sup>19</sup>

#### 4.3.4.1 The procedure

QCD multijet MC datasets with a large  $p_T$  range used 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. 4.3.3, the jets these calibrations are derived for are post-hadronisation, but before interaction with the detector. They will be referred to as “Level-1 (L1) jets.” The reference jets directly from the generator (“GenJets” as colloquialised in CMS) are important for matching to our L1 jets to ensure we are not mistakenly using jets from the parton shower that have no well-defined source.

The reference and L1 jets are matched using the variable  $\Delta R$  (see Eq. 4.6). The algorithm used to match the jets does so 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 ( $p_T^{\text{ref.}}$ ).<sup>20</sup> 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. Our metric for measuring the detector response is the mean of these ratios:

$$(4.8) \quad r_j = \left\langle p_T^{\text{L1}} / p_T^{\text{ref.}} \right\rangle$$

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<sup>21</sup> is used for the tail:

$$(4.9) \quad p_T^{\text{L1, corr.}} = p_T^{\text{L1}} \cdot \left( p_0 + \frac{p_1}{\left( \log_{10} p_T^{\text{L1}} \right)^2 + p_2} + p_3 \cdot \exp \left( -p_4 \left( \log_{10} p_T^{\text{L1}} - p_5 \right)^2 \right) \right)$$

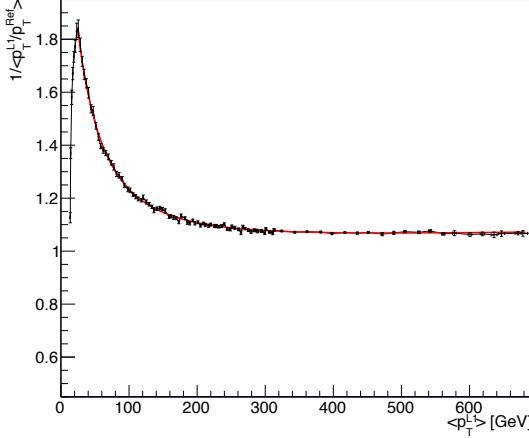
---

<sup>19</sup>Note: I have only forked it from the original developers and made modifications on top.

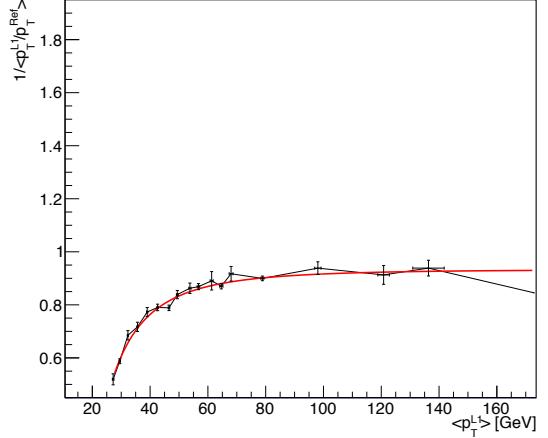
<sup>20</sup>Should I add tables somewhere of the  $p_T^{\text{ref.}}$  and  $|\eta|$  bins for the calibrations?

<sup>21</sup>Go into more detail regarding the equation? Reasoning, etc.

The input 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. 4.8 for an example). The value of this fit function in each  $p_T^{\text{ref}}$  bin is exported.



(a)  $0.435 < |\eta| < 0.783$  (Barrel)



(b)  $4.191 < |\eta| < 5.191$  (HF)

Figure 4.8: 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. 4.8) 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 we have just 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. 4.9 showing scatter plots of  $p_T^{\text{ref}}$  vs.  $p_T^{\text{L1}}$  before and after JEC are applied, and Fig. 4.10 illustrating the response.

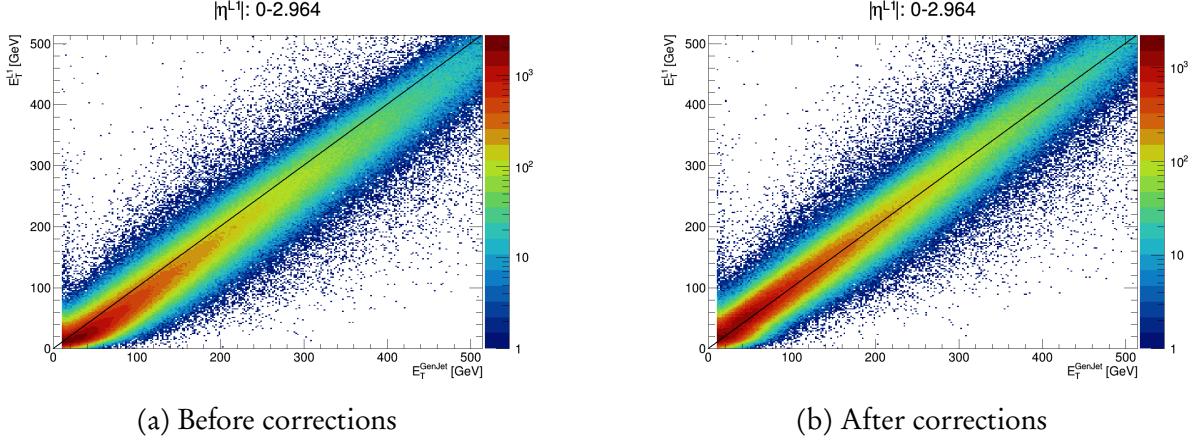


Figure 4.9: The energies of matched pairs of jets in the entire barrel and end cap, in the pileup 40–50 range, before and after jet energy corrections have been applied. After calibrations, the distribution is much more symmetrical. An equivalent plot using jets from LHC data is expected to look similar after applying these calibrations.

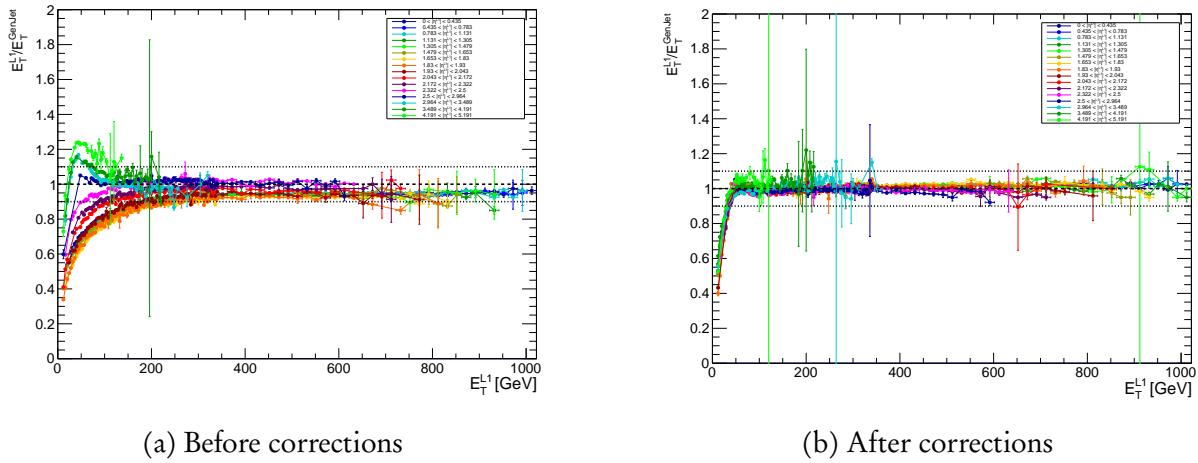


Figure 4.10: The response curves in each  $|\eta|$  bin as a function of  $p_T^{L1}$ , in the pileup 40–50 range, before and after JEC are applied. Note that in panel b, the  $x$ -axis is the corrected  $p_T^{L1}$ .



## 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 and units in high energy physics, particularly in the context of colliders. 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 analysis in Chpt. 6.

### 5.1 Ubiquitous observables and quantities in collider physics

#### 5.1.1 The electron volt

In highly relativistic systems, such as beams of particles in accelerators, the ability to simply equate mass, energy, and momentum is desirable. In the LHC, when protons are accelerated to an enormous Lorentz boost factor, their invariant mass  $m_0$  contributes little to their total energy  $E$ . With Einstein's energy-momentum relation from the special theory of relativity,

one can express the energy as functions of intrinsic mass and relativistic momentum:

$$(5.1) \quad E^2 = (pc)^2 + (m_0 c^2)^2$$

where  $p$  is the magnitude of the momentum and  $c$  is the speed of light. For highly relativistic objects,  $pc \gg m_0 c^2$  and so  $E \approx pc$ . At rest,  $E = m_0 c^2$ . The electron volt (eV) unit is common in high energy physics. Its value is the energy supplied to (or removed from) an electron accelerated through a potential difference of 1 V:  $1.6 \times 10^{-19}$  J. The momentum gained is then  $1 \text{ eV}/c$  and relativistic mass  $1 \text{ eV}/c^2$ . The factors of  $c$  and  $c^2$  are often dropped in less formal contexts, or when using natural units (where  $c = 1$ ).

An electron volt is a minute quantity of energy, so when discussing properties of high energy particles and accelerators, a long string of digits may be required to express them. SI prefixes mitigate this problem and provide an intuitive sense of scale to scientists. The most frequently used in the context of LHC physics are *mega* (M,  $10^6$ ), *giga* (G,  $10^9$ ), and *tera* (T,  $10^{12}$ ). For example, the mass of a proton is  $0.93 \text{ GeV}/c^2$  and the present centre of mass energy of the LHC is  $13 \text{ TeV}$ , which are much more natural and understandable numbers than  $1.78 \times 10^{-27} \text{ kg}$  or  $1.6 \times 10^{-7} \text{ J}$ , respectively.

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

In the LHC (or any other collider), the longitudinal momentum of the initial state particles is typically unknown. However, the momentum transverse to the beam is zero before the collision, and must be zero afterward 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.<sup>1</sup>

### 5.1.3 $H_T$

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  $H_T$ :

$$(5.2) \quad H_T \equiv \sum_{\text{jets}} p_T$$

Typically, a lower limit on the  $p_T$  is used 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. 4.2.3), and those from the primary vertex that can often be mismeasured.

---

<sup>1</sup>Do I need a more formal/mathematical definition? Could just reference geometry section in detector chapter.

### 5.1.4 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 used interchangeably 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.<sup>2</sup> Formally,

$$(5.3) \quad \begin{aligned} \vec{p}_T^{\text{miss}} &\equiv - \sum_i^{N_{\text{particles}}} \vec{p}_{T,i} \\ p_T^{\text{miss}} &\equiv \left| \vec{p}_T^{\text{miss}} \right| \end{aligned}$$

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:

$$(5.4) \quad \begin{aligned} \vec{H}_T^{\text{miss}} &\equiv - \sum_j^{N_{\text{jets}}} \vec{p}_{T,j} \\ H_T^{\text{miss}} &\equiv \left| \vec{H}_T^{\text{miss}} \right| \end{aligned}$$

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

### 5.1.5 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 often 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. Eq. 5.1 can be reformulated for each daughter particle in a decay as

$$(5.5) \quad E_T^2 = m_0^2 + p_T^2$$

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

$$(5.6) \quad \begin{aligned} m_T^2 &= (E_{T,1} + E_{T,2})^2 - (\vec{p}_{T,1} + \vec{p}_{T,2})^2 \\ &= m_1^2 + m_2^2 + 2(E_{T,1}E_{T,2} - \vec{p}_{T,1} \cdot \vec{p}_{T,2}) \end{aligned}$$

---

<sup>2</sup>A depiction of the MET (e.g., SM particles recoiling off dark matter) might be useful for the reader.

In the case of one unidentified daughter, the  $\vec{p}_T^{\text{miss}}$  may substitute a particle in the first line of Eq. 5.6. If one also treats the daughters as massless, i.e., if  $p_T \gg m_0$  like with boosted leptons, the formula approximates to

$$(5.7) \quad m_T^2 = 2p_T p_T^{\text{miss}}(1 - \cos(\Delta\phi))$$

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

## 5.2 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 we use, 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.

Different working points for the identification of many objects are available from the respective Physics Object Groups (POGs), trading inclusivity for purity the tighter the restrictions are. “Veto” and “loose” criteria are used to specify the secondary leptons in the dilepton control regions, and when vetoing leptons and photons (or applying their corresponding veto weights) in the signal region. The “medium” and “tight” selections are used to construct the primary objects in the control regions. We denote these particles with a subscript that signifies the working point. The requirements elaborated upon in the ensuing subsections.<sup>3</sup>

### 5.2.1 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, we then apply the *latest* JEC and jet energy resolution (JER) treatments 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 uncertainty attributed to the JER—and

---

<sup>3</sup>The following subsections take up a good few pages. I’m not sure if I need to go into this much detail regarding the object definitions. For the “basic” requirements for each object, should I instead just write a summary paragraph or two, like they do in CMS papers?

the several sources from the jet energy scale (JES)—are propagated through the analysis and explained further in Chpt. 6.4.8.

For the jets that remain, we explicitly restrict the pseudorapidity range to  $|\eta| < 5.0$ . Either the tight, or tight + 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).<sup>4</sup> These are documented in Tabs. 5.1, 5.2, and 5.3. Finally, jets must not overlap (i.e., exhibit  $\Delta R > 0.4$ ) with any  $e_{\text{veto}}$ ,  $\mu_{\text{loose}}$ , and  $\gamma_{\text{loose}}$  in the event to avoid potential misidentification.

AK8 jets follow the same reconstruction and identification prescription as AK4 jets, only differing in the cone size.<sup>5,6</sup>

A  $b$ -jet is described by a nanoAOD AK4 jet with corrected  $p_{\text{T}} > 20 \text{ GeV}$ ,  $|\eta| < 2.5$  (2.4 in 2016), and to satisfy the medium working point (a mis-tag rate of 1 %) from the DEEPCSV algorithm [128]. The systematic uncertainty associated with the algorithm’s classification is described in Chpt. 6.4.7.2.

### 5.2.1.1 2016

The tight and tight + lepton veto ID specifications are detailed in Tab. 5.1. 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.

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	$< 0.99$ (0.90*)	—	—	—
# neutral particles	—	—	$> 2$	10

Table 5.1: 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. 54.

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

<sup>4</sup>For all objects, should I also give the background rejection efficiency?

<sup>5</sup>I’m not what/if the minimum  $p_{\text{T}}$  of an AK8 jet is, and whether they have an  $\eta$  restriction (I’ve seen  $|\eta| < 2.4$  in a few places, but nothing confirmed).

<sup>6</sup>We don’t currently apply the JES and JER corrections and uncertainties to AK8 jets. Will need to update this bit if we do get around to it.

### 5.2.1.2 2017

The tight and tight + lepton veto ID conditions are detailed in Tab. 5.2. 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 5.2: 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. 55.

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

### 5.2.1.3 2018

The tight and tight + lepton veto ID specifications are detailed in Tab. 5.3.<sup>7</sup> As in 2017, jets with  $p_T < 50 \text{ GeV}$  must fulfill the tight pileup ID conditions.<sup>8</sup>

## 5.2.2 Muons

The reconstruction and identification criteria for muons is described in significant detail in Ref. 129. A loose muon  $\mu_{\text{loose}}$  satisfies the “loose muon identification” requirements from the paper,  $|\eta| < 2.4$ ,  $p_T > 10 \text{ GeV}$ , and the loose working point of the PARTICLE FLOW relative isolation parameter (Eq. 5.4 in Ref. 125)  $I_{\text{PF}} < 0.25$ . A tight muon  $\mu_{\text{tight}}$  meets the “tight muon ID” requirements from the same paper,  $|\eta| < 2.4$ ,  $p_T > 20 \text{ GeV}$ , 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.

<sup>7</sup>There’s probably a more elegant way to write the ID criteria, since the same variables are used for all years and the same thresholds in many cases. But until I figure it out, separate tables for each year will have to do

<sup>8</sup>Is it worth going into the pileup ID spec.?

## 5.2. CLASSIFICATION OF ANALYSIS-LEVEL PHYSICS OBJECTS

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 5.3: 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. 56.

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

### 5.2.3 Electrons

In the analysis, 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. 5.4, and end cap region in Tab. 5.5. 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 $A_{\text{eff}}$ .	$< 0.198 + \frac{0.506}{p_{\text{T}}}$	$< 0.0287 + \frac{0.506}{p_{\text{T}}}$
$ 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 5.4: Requirements used to define the veto and tight electron IDs in the barrel region (supercluster  $|\eta| \leq 1.479$ ). Information taken from Ref. 52.

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. They are taken at the point of closest approach to the supercluster, extrapolated from the innermost track

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 5.5: Requirements used to define the veto and tight electron IDs in the end cap region (supercluster  $|\eta| > 1.479$ ). Information taken from Ref. 52.

state.<sup>9</sup>  $H/E$  is the ratio of the candidate’s central energy deposit in the HCAL to the ECAL. The relative isolation parameter is computed in the same way as muons (Eq. 5.4 in Ref. 125 ), 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 [42] 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, we place a cut 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:  $d_0 < 0.05 \text{ cm}$  and  $dz < 0.1 \text{ cm}$  in the barrel, with  $d_0 < 0.1 \text{ cm}$  and  $dz < 0.2 \text{ cm}$  in the end cap. Cross cleaning is also performed against the  $\mu_{\text{loose}}$  collection of objects, so electrons are not counted in the event if they are within  $\Delta R < 0.3$  of one. The  $e_{\text{veto}}$  and  $e_{\text{tight}}$  sets are only separated by the identification prerequisites above, and transverse momentum. A veto electron requires  $p_{\text{T}} > 10 \text{ GeV}$ , and a tight electron  $p_{\text{T}} > 40 \text{ GeV}$ .

### 5.2.4 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. We define loose  $\gamma_{\text{loose}}$  and medium photons  $\gamma_{\text{medium}}$  that fulfill the corresponding working point criteria for identification. For inclusivity, the latter is used instead of the tight working point. These are detailed in

<sup>9</sup>Consider rewording this sentence. Currently lifted verbatim from a public twiki page.

## 5.2. CLASSIFICATION OF ANALYSIS-LEVEL PHYSICS OBJECTS

---

Tab. 5.6 and Tab. 5.7 for the barrel and end cap regions, respectively. The loose photon ID specification is designed to be 90 % efficient, and the medium ID 80 %.<sup>10</sup> Both the  $\gamma_{\text{loose}}$  and  $\gamma_{\text{medium}}$  collections are cross-cleaned against  $\mu_{\text{loose}}$  and  $e_{\text{veto}}$  objects in the event with  $\Delta R < 0.3$ . They must also pass the electron conversion veto to ensure they originate promptly from the primary vertex.

Criterion	Loose	Medium
$H/E$	$< 0.04596$	$< 0.02197$
$\sigma_{i\eta i\eta}$	$< 0.0106$	$< 0.01015$
$\rho$ -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 5.6: Requirements used to define 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. 53.

Criterion	Loose	Medium
$H/E$	$< 0.0590$	$< 0.0326$
$\sigma_{i\eta i\eta}$	$< 0.0272$	$< 0.0272$
$\rho$ -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 5.7: Requirements used to define 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. 53.

The variables  $\sigma_{i\eta i\eta}$  and  $\rho$  follow the same definitions as in Chpt. 5.2.3. 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:

$$(5.8) \quad I_{\text{PF}}^{\text{corr.}} = \max(I_{\text{PF}} - \rho A_{\text{eff.}}, 0)$$

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  $p_T > 230 \text{ GeV}$ .

---

<sup>10</sup>Should I give the background rejection efficiencies as well?

### 5.2.5 Tau leptons

The only purpose tau leptons  $\tau$  serve is to veto (or apply veto weights to) events that contain them. We denote these as “very loose” taus  $\tau_{v.\;loose}$ .<sup>11</sup>

### 5.2.6 Revised energy sums $\vec{p}_T^{\text{miss}}$ , $H_T$ , and $\vec{H}_T^{\text{miss}}$

Most CMS analyses, including this one, we define the missing transverse momentum with the “type-I correction”: corrections to the JER and JES are applied as in Chpt. 5.2.1, 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 each of the control regions detailed in Chpt. 6.7.2, the  $\vec{p}_T^{\text{miss}}$  is recalculated without the objects used to define said region to model the process it is predicting. Conditions in the event selection and binning that refer to  $p_T^{\text{miss}}$  use the recalculated quantity when applied to the control regions.

In 2017, significant noise in the ECAL end cap affected jets and  $p_T^{\text{miss}}$ , 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. 5.1.3 and 5.1.4, respectively. They are calculated from the analysis-level AK4 jets with  $p_T > 30\text{ GeV}$ , as detailed in Chpt. 5.2.1.

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.

---

<sup>11</sup>Tau definition still not finalised so leave out this subsection for now.

## COMBINED SEARCH FOR INVISIBLY DECAYING HIGGS BOSONS IN HADRONIC CHANNELS

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 troublesome 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 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.

- Emphasize my contributions: control region construction and studies, background estimation, ttbar systematics, and other studies I will have conducted by the time I write up (check AN, chip and HToInv-nanoAOD-tools MRs, and git commit history).
- Maintaining CMS internal analysis note, documenting all aspects of the analysis. I will

first add all relevant information there which I can subsequently use when writing this chapter.

- Since it's my thesis, I can talk about  $t\bar{t}H$ ,  $VH$  and  $ggH$ /monojet, even though the Bristol contribution to the final, public result may only be  $t\bar{t}H$  and resolved  $VH$ . Would need to be able to run the fit for all three modes simultaneously, ensuring we have complete (and correct) systematics for  $ggH$ .

## 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 all four of the production modes from the outset. In contrast to many of the previous analyses that separately set limits on  $\mathcal{B}(H \rightarrow \text{inv.})$  by reinterpretation, our approach has many benefits. A simultaneous search for all of the production modes allows us to construct search regions to target each one, building in 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.

Direct collaboration amongst several institutes divides up the task appropriately. Researchers at the University of Bristol, including as myself, and LLR (*Laboratoire Leprince-Ringuet, École Polytechnique, Université Paris-Saclay*) focus on  $t\bar{t}H$  and the resolved  $VH$  modes. Those at Imperial College London, who have a long history with the VBF search [48, 133], assume responsibility of this process. Researchers at Boston University—who have previously been involved in BSM analyses with interpretations in invisibly decaying Higgs bosons [103, 127]—specialize in monojet and mono- $V$  searches. This translates into the boosted aspect of  $VH$ , and  $ggH$  modes of the analysis.

Emphasis will be placed on the non-VBF modes, particularly  $t\bar{t}H$  and resolved  $VH$  as they are novel dedicated searches. The boosted  $VH$  and  $ggH$  mechanisms also under our charge, drawing on expertise from Boston. Accordingly, they will also be given some attention. Definitions of the physics objects used analysis-wide have already been discussed in Chpt. 5.2. The data from CMS and from simulation, along with required corrections and systematic uncertainties, are acknowledged in Chpt. 6.3. These are followed by the event selection in Chpt. 6.5. Characterisation of the categorisation of production modes and phase space

regions are illustrated in Chpts. 6.6 and 6.7, respectively. Methods for accurately estimating the background processes in the signal region are delineated in Chpt. 6.8. The incorporation of these and the aforementioned sections in the statistical treatment are presented in Chpt. 6.9, before the results of the analysis of the  $t\bar{t}H$ ,  $VH$ , and  $ggH$  modes in Chpts. 6.10, 6.11 and 6.12, respectively.

## 6.2 Only tools and forces: software and toolkits

The high energy physics community is well-acquainted with the software package `ROOT`. The extensive suite of functions and tools it offers makes it the de facto standard in many analyses. Even though it is entrenched in particle physics, the native C++ implementation and difficulty in integrating it with Python make `ROOT` often cumbersome for newcomers to learn, and can be difficult to develop an analysis around it. In recent years, focus has started to shift toward industry-standard Python libraries such as `numpy` and `pandas` for analysis, and `matplotlib` or `seaborn` for plotting.

The FAST set of tools utilise these, along with vectorisation, to perform an analysis very quickly with simple YAML-controlled config files. Jobs can be submitted on the batch system at a given site, running the full analysis on the upwards of 15 billion events in as little as thirty minutes. Developed by colleagues at Bristol, a set of packages harmoniously work together to run most components of an analysis. Dataset information is gathered with `fast-curator` and stored in YAML configs. With `fast-carpenter`, modules (both custom and internal) for the different stages of the analysis process the datasets. Output is in the form of binned pandas dataframes of user-specified variables. They may be subsequently be fed into `fast-plotter` for visualisation as 1D histograms. All of the stages are controlled by YAML configs and are interpreted by `fast-flow`. Only a simple interface to `ROOT` is required to extract the data. Further information on these packages are available at <http://fast-hep.web.cern.ch/fast-hep/>.

Before the main analysis is run with the above tools, a light skim—or reduction of events—of the remotely-available datasets is effectuated with `nanoAOD-tools`.<sup>1</sup> Operating on the nanoAOD data tier, the repository contains corrections, scale factors, and systematic uncertainties for CMS data and simulation. They are maintained by members of the Collaboration, so it is ensured they remain up to date. Our custom modules are applied on top of this: introducing additional scale factors and systematic uncertainties, classification of some physics objects, and add supplementary variables. Running over the datasets on

---

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

the Worldwide LHC Computing Grid, output is stored on networked university storage elements. The primary benefit of this is a quicker input-output stream that improves the performance of the later stages of the analysis.<sup>2,3</sup>

## 6.3 Data and simulation

### 6.3.1 Data

The data collected by the CMS experiment that is used in the analysis corresponds to an integrated luminosity of  $137.19 \text{ fb}^{-1}$ , and recorded from 2016–2018. A breakdown by year is given in Tab. 4.1. At over three times the volume at  $\sqrt{s} = 13 \text{ TeV}$  analysed in the previous search, there is potential to substantially lower the ceiling of  $\mathcal{B}(H \rightarrow \text{inv.})$ . 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  $E_T^{\text{miss}}$  and  $H_T^{\text{miss}}$  combination triggers,<sup>4</sup> 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: *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. 6.4.8) to account for the effect.

---

<sup>2</sup>Link to our HTToInv-nanoAOD-tools and chip repos? Would have to say they’re restricted.

<sup>3</sup>Also not sure if I need to mention the later steps—fast-datacard, and Combine for the fit. Perhaps in the statistical treatment section?

<sup>4</sup>I try to be consistent in most places and use  $p_T^{\text{miss}}$  instead of  $E_T^{\text{miss}}$ . But since the HLT paths refer to  $E_T^{\text{miss}}$ , which is the more correct symbol here? And should I update Tab. 6.4 accordingly?

In 2018, a sector of the HCAL end cap lost power, leaving the approximate area bound by  $\eta \in [-3.0, -1.4]$  and  $\phi \in [-1.57, -0.87]$  uncovered 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 is more difficult to accurately record the properties of jets, and therefore reliably calculate  $p_T^{\text{miss}}$ . Studies performed showed an excess in data in the affected area of the  $\phi(p_T^{\text{miss}})$  distribution of many of our regions and categories. To mitigate this issue, the cuts in Chpt. 6.5.3 were added to the preselection.

### 6.3.2 Simulated signal processes

All Monte Carlo datasets used in this analysis were produced centrally by experts in CMS as outlined in Chpt. 4.3.3. 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 NNLO. The samples are produced with a Higgs boson mass of  $m_H = 125 \text{ GeV}$ . The vector bosons in each of the  $VH$  channels decay to  $q\bar{q}$ . When simulating 2016 data, the tuning of the shower parameters in PYTHIA was labelled CUETP8M1 (the latest available), whilst for 2017 and 2018 datasets the newer CP5 version was used.

The cross section of each mechanism at  $\sqrt{s} = 13 \text{ TeV}$  is detailed in Tab. 6.1. The signal simulation assumes a 100 % branching ratio to invisible states.

Dataset ( $H \rightarrow \text{inv.}$ )	Cross section (pb)	Accuracy
$t\bar{t}H$	$5.07 \times 10^{-1}$	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 6.1: Cross sections of the  $H \rightarrow \text{inv.}$  signal processes in the analysis. They are calculated at  $\sqrt{s} = 13 \text{ TeV}$  at the highest orders available. The simulation datasets are generated with a 100 % branching ratio of the Higgs boson to invisible states, and in the  $VH$  processes the  $V$  decays in all cases to  $q\bar{q}$ .

### 6.3.3 Simulated background processes

Monte Carlo datasets for many processes<sup>5</sup> are used 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:<sup>6</sup>

- 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, is the dominant background in the dilepton control regions used to model 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.
- Multiboson: This background encompasses the production of two (diboson) or three (triboson) electroweak bosons from QCD(?) vertices. They may decay leptonically, also producing neutrinos in the case of  $W$  bosons, 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  $V + 2 \text{jets}$ : 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.
- $\gamma + \text{jets}$ : Events with photons are vetoed in the signal region. However, this is the dominant background in 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 LO with MADGRAPH5\_AMC@NLO, requiring the  $\Delta R$  between a photon and jet to be at least 0.4.
- QCD multijet: The most common type of event produced in  $pp$  collisions at the

---

<sup>5</sup>In this subsection, I use the word “process” a lot. What alternatives can I use? Decay, dataset, mechanism, background?

<sup>6</sup>Not sure if I should use a long, sprawling list, or just have a separate paragraph for each process, maybe with the process name in bold.

LHC is several jets from QCD vertices. These typically do not have large  $p_T^{\text{miss}}$ , but are prone to mismeasurements 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`.

- Single top: Events with one final-state top quark are a subdominant background in the analysis, but are important to consider, especially in the  $t\bar{t}H$  categories. These electroweak-induced processes include  $s$ -channel and  $t$ -channel production where a *four-flavour scheme* [107] 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` [25] to include spin correlations.
- $t\bar{t} + \text{jets}$ : The QCD-induced process presents a major background, largely in the  $t\bar{t}H$  categories. Large  $p_T^{\text{miss}}$  can arise in scenarios where the leptons in the decay are “lost” ( $\ell_{\text{lost}}$ ) from misidentification or are outside the bounds of the detector acceptance. The three channels (hadronic, semi-leptonic, and dileptonic) are modelled with `POWHEG` at NLO accuracy.
- $t\bar{t}X + \text{jets}$ : 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` with the “FxFx” jet matching scheme for the hard scatter, and decay the particles with `MADSPIN`.  $t\bar{t}Z + \text{jets}$  uses `MADGRAPH5_amc@NLO` without the two aforementioned additions. Meanwhile,  $t\bar{t}H + \text{jets}$  (where the  $H$  decays to visible states) is generated with `POWHEG`.
- $W(\rightarrow \ell\nu) + \text{jets}$ : A QCD-induced process, this background 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`.
- $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ : From the presence of neutrinos, the genuine  $p_T^{\text{miss}}$  from this channel is a dominant background in the signal region. Its contribution is estimated using the

double lepton and  $\gamma + \text{jets}$  control regions. As with many of the above, the datasets are generated at LO accuracy with `MADGRAPH5_amc@NLO`.<sup>7,8</sup>

Cross sections are specified at the highest order available, except for the LO processes that require reweighting expounded in Chpt. 6.4.4. The dominant backgrounds that appear in the signal region— $t\bar{t} + \text{jets}$ ,  $W(\rightarrow \ell\nu) + \text{jets}$ , and  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ —are estimated using transfer factors from the control regions they enrich (see Chpt. 6.7.2). Contributions from QCD multijet events should be adequately suppressed by the analysis-level selection requirements. However, it is still a process that must be accurately simulated 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:

$$(6.1) \quad \mathcal{L}_{\text{eq.}} = \frac{N_{\text{events}}}{\sigma}$$

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. A dedicated method to predict its presence in the signal region is implemented in Chpt. 6.8.3. The contributions from the remaining backgrounds are simply taken from their yield in the signal region.

## 6.4 Weights, corrections, and systematic uncertainties for simulation

In order for simulated events, particularly for background processes, to resemble LHC data as closely as possible, many corrections and weights are applied. These are discussed in more detail in the subsequent sections. All weights and associated systematic uncertainties are applied to all samples for all data taking years, unless stated otherwise. A final event weight  $w_{\text{event}}$  is the product of the weights from all of the individual sources  $i$  that provide one:

$$(6.2) \quad w_{\text{event}} = \prod_i w_i$$

---

<sup>7</sup>The gridpacks/hard process are exactly the same for all samples in 2017 and 2018 (i.e., the same files), I think. But perhaps don't need to mention.

<sup>8</sup>In all of the processes where the datasets are split by generator-level  $H_T$  (with the exception of  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ ), the “MLM” jet matching scheme applied in `MADGRAPH`. Not sure if I need to mention either of these.

## 6.4. WEIGHTS, CORRECTIONS, AND SYSTEMATIC UNCERTAINTIES FOR SIMULATION

---

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

$$(6.3) \quad N_{\text{corr.}} = \sum_j^{N_{\text{MC}}} w_{\text{event } j}$$

where  $N_{\text{MC}}$  is the number of unweighted, simulated events in the bin. The statistical uncertainty ascribed to the yield in a bin is given as

$$(6.4) \quad \Delta N_{\text{corr.}} = \pm \frac{N_{\text{corr.}}}{\sqrt{N_{\text{MC}}}}$$

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

$$(6.5) \quad \Delta N_{\text{data}} = \pm \frac{N_{\text{data}}}{\sqrt{N_{\text{data}}}}$$

The standard prescription for error propagation is to approximate the uncertainty as the square root of the sum of the weights squared [36]:

$$(6.6) \quad \Delta N_{\text{corr.}} = \pm \left( \sum_j^{N_{\text{MC}}} w_{\text{event } j}^2 \right)^{1/2}$$

Our reasoning for using Eq. 6.4 for MC instead of Eq. 6.6 is that the error should be determined purely from the integer number of events we select ( $k$  in a Poisson statistical treatment), regardless of whether they are weighted or not. This often reduces the uncertainty for MC compared to Eq. 6.6 since many more events are generated to predict a given equivalent luminosity. Further justification is that it is a good approximation in our assumed regime where we expect a large number of events from our MC samples before any cuts are applied (say  $N$ ), and a large enough number of events after the cuts such that we do not encounter the low- $k$  or low- $N$  limits of Poissonian error propagation.

### 6.4.1 Veto and selection weights

In an analysis, events are often rejected by placing kinematic or object-based requirements. This type of selection strictly removes an event from the analysis if the condition is not met. While kinematic requirements are either fulfilled or not, a different approach can be used when selecting the number of objects, i.e., when defining control regions. For a set of objects, the selection weight at event level is defined as

$$(6.7) \quad w_{\text{sel.}} = \prod_i^{N_{\text{objects}}} \epsilon_i$$

where  $\epsilon_i$  is the efficiency/scale factor applied to object  $i$ . Only reconstructed (“reco level”) objects that have been matched to a generator level object are considered. For leptons ( $e, \mu, \tau$ ) and photons, these scale factors are typically 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-MC scale factor at the given working point from the algorithm used to identify them. These weights are calculated individually for each type of object in an event, and individually for each source since they also introduce systematic variations that cannot be trivially aggregated. A veto weight is represented as

$$(6.8) \quad w_{\text{veto}} = \prod_i^{N_{\text{objects}}} 1 - \epsilon_i$$

which is essentially the probability to mis-tag the objects and allow it to enter the signal region as a consequence. The uncertainties/systematic variations follow the same prescription. With these quantities defined, an event that meets the object criteria is given the selection weight, otherwise it is given the veto weight. For example, if an event with one muon meets the criteria for the  $\mu + \text{jets}$  region, it will enter that region with its muon-related selection weights. That same event can also enter the signal region or one of the sidebands (depending on the event kinematics) with the muon-related veto weights. The veto weights for the other leptonic objects will just be unity, as per Eq. 6.8. This “migration” of events, where they are able to contribute to more than one region, and the fact that weights are applied instead of event rejection, provides a noticeable decrease to the Monte Carlo statistical uncertainty in a given bin of a distribution.

One thing must be noted about the migration of events, since we have many different regions of phase space in the analysis. The signal region and the sidebands have orthogonal kinematic prerequisites, so an event cannot enter the signal region *and* one of the sidebands. The same is true amongst the control regions, i.e., an event cannot enter more than one of them due to the designed orthogonality. An event *is* able to enter the signal region or a sideband with  $w_{\text{veto}}$ , and also one of the control regions with  $w_{\text{sel}}$ . Since events in data are not weighted, they may only enter a single region.

#### 6.4.2 Cross section reweighting

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

$$(6.9) \quad w_\sigma = \frac{\sigma \mathcal{L}_{\text{int.}}}{N\varepsilon}$$

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 (assumed to be unity for all datasets since no generator-level cuts are applied).<sup>9</sup> If a dataset is generated at leading order, higher order corrections are usually applied on an event-by-event basis that changes the shapes of distributions (see Chpt. 6.4.4). In some circumstances, “flat”  $k$ -factors can be applied to a dataset that only alters the normalisation (i.e., its cross section).

### 6.4.3 Pileup reweighting

Pileup interactions at the LHC are frequent (see Chpt. 4.2.3) 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.

In the trees of the simulated samples, the branch `Pileup_nTrueInt` is the mean of the Poisson distribution from which random numbers are drawn. In each simulated event, these random numbers (all from the same distribution) are used to set the number of in-time pileup interactions as well as the number of the interactions in each neighbouring bunch crossing to simulate the out-of-time pileup. In data, the same branch gives the average number of pileup interactions for a colliding bunch pair in a luminosity section.<sup>10</sup> The distribution of `Pileup_nTrueInt` in the data is derived from the measured instantaneous luminosity for each colliding bunch pair in each lumi section and the cross section of the total inelastic  $pp$  interaction.

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. The pileup distributions, and therefore

---

<sup>9</sup>This statement on filter efficiency might not be strictly true, since the Drell-Yan datasets do apply a generator-level cut/requirement.

<sup>10</sup>A luminosity section is the duration (approximately 23 seconds) over which the instantaneous luminosity in colliding bunches remains constant.

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.

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

Vector boson +jets processes present a sizeable background in some of the categories and regions of the analysis. At the dataset sizes required to provide sufficient statistical coverage of the phase space, NLO MC samples are not available with the complete detector simulation applied. To circumvent this, LO samples are used in the analysis 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.

Two types of reweighting are applied: QCD corrections to QCD and electroweak processes, and electroweak corrections to QCD processes. For both cases, NLO  $W + \text{jets}$ , Drell-Yan  $Z(\rightarrow \ell\bar{\ell}) + \text{jets}$ , and  $\gamma + \text{jets}$  MC are used to derive  $k$ -factors for the respective LO processes. One or two additional jets in the matrix element calculations are permitted. The NLO Drell-Yan MC is also used to derive corrections for LO  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ , where the cross sections are modified for that process. The NLO samples are processed through the following generator-level event selection to ensure they are in a similar phase space to the analysis:

- $p_T^{j_1} > 80 \text{ GeV}$
- $p_T^V > 200 \text{ GeV}$
- $H_T^{\text{Gen}} > 200 \text{ GeV}$
- $H_T^{\text{Gen}}/p_T^V < 1.2^{11}$

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 final selection in the list mimics the  $H_T^{\text{miss}}/p_T^{\text{miss}}$  cut in Chpt. 6.5.1. The  $k$ -factors are binned in two dimensions for the QCD corrections to  $W$  and  $Z$  processes,  $p_T^V$  and  $p_T^{j_1}$ . For those corrections to  $\gamma$  process and for all electroweak corrections, they are binned solely in  $p_T^V$  as they were derived by Boston University in the monojet phase space.

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. Distributions of all of the  $k$ -factors are presented in Fig. 6.1.

---

<sup>11</sup>Not sure if we'll need updated k-factors if we switch to another scenario where we don't use the MHT/MET cut.

## 6.4. WEIGHTS, CORRECTIONS, AND SYSTEMATIC UNCERTAINTIES FOR SIMULATION

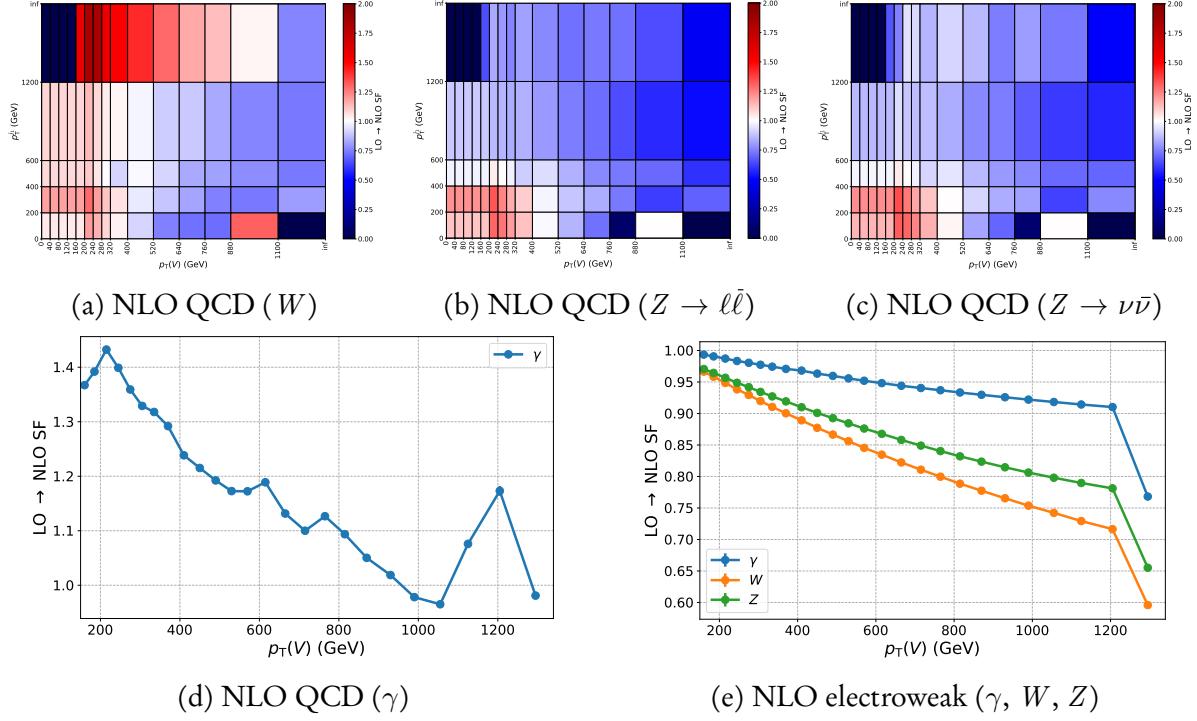


Figure 6.1: NLO QCD and electroweak  $k$ -factors used to reweight events in the LO  $V + \text{jets}$  background samples.

### 6.4.5 Efficiency of the triggers

### 6.4.6 The $t\bar{t}$ process

$t\bar{t} + \text{jets}$  is the dominant background for the  $t\bar{t}H$  process. As such, much effort ensures it is modelled correctly. Variations of the QCD renormalisation and factorisation scales provide an uncertainty only to the shape of the  $p_{\text{T}}^{\text{miss}}$  spectrum of  $t\bar{t}$ . Reweighting of the top quark  $p_{\text{T}}$  distribution to NNLO accuracy is also performed, yielding a shape correction and associated uncertainty.

#### 6.4.6.1 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, weights exist in the ntuple for each combination of these scales being varied independently up and down (and together) 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  $f_{\text{var}}$ :

$$(6.10) \quad f_{\text{var}} = \frac{\sum_{\text{unskimmed events}} w_{\mu_{R\text{ nom.}} \& \mu_{F\text{ nom.}}}}{\sum_{\text{unskimmed events}} w_{\text{var}}}$$

The table below documents the values of  $f_{\text{var}}$ . The values did not differ until the fourth decimal place between the final states or datasets.

Variation	$f_{\text{var}}$
$\mu_{R\downarrow} \& \mu_{F\downarrow}$	0.880
$\mu_{R\downarrow} \& \mu_{F\text{ nom.}}$	0.897
$\mu_{R\downarrow} \& \mu_{F\uparrow}$	0.910
$\mu_{R\text{ nom.}} \& \mu_{F\downarrow}$	0.974
$\mu_{R\text{ nom.}} \& \mu_{F\uparrow}$	1.021
$\mu_{R\uparrow} \& \mu_{F\downarrow}$	1.082
$\mu_{R\uparrow} \& \mu_{F\text{ nom.}}$	1.116
$\mu_{R\uparrow} \& \mu_{F\uparrow}$	1.145

Table 6.2: The values of factor  $f_{\text{var}}$  for each QCD renormalisation and factorisation scale variation as calculated for the  $t\bar{t}$  background samples. These factors remove the effect of a change in the normalisation of the event yields.

The envelope is characterised by the correlated variations  $\mu_{R\uparrow} \& \mu_{F\uparrow}$ , and  $\mu_{R\downarrow} \& \mu_{F\downarrow}$ . Confirmed by Fig. 6.2.

#### 6.4.6.2 Top quark $p_{\text{T}}$ reweighting

The top quark  $p_{\text{T}}$  distribution has been shown to be harder in simulation than data [131]. Developed by members of the CMS Collaboration, the distribution of the NNLO to NLO ratio against the generator-level top  $p_{\text{T}}$  is fit with an exponential function.<sup>12</sup> The same function is used for all  $t\bar{t} + \text{jets}$  samples for all data taking years.

<sup>12</sup>If the paper/PAS for HIG-19-008 with this fit function in, reference it.

## 6.4. WEIGHTS, CORRECTIONS, AND SYSTEMATIC UNCERTAINTIES FOR SIMULATION

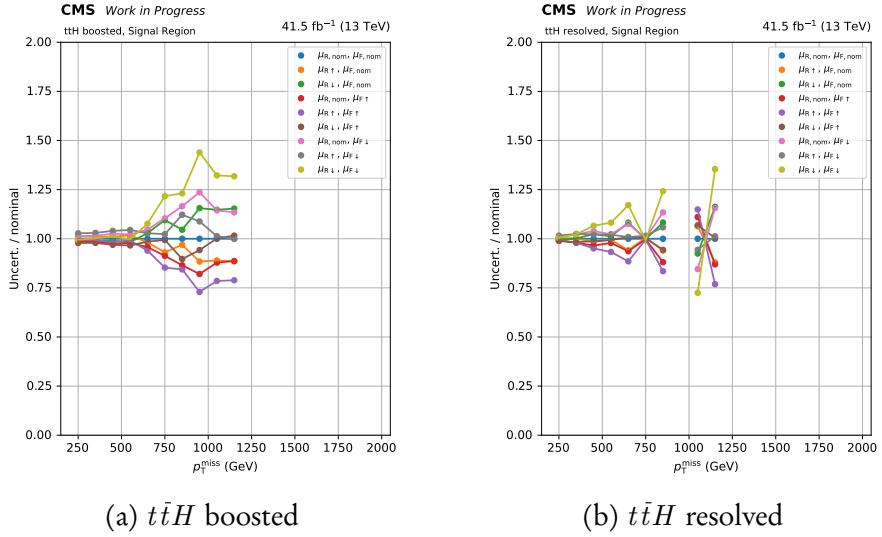


Figure 6.2: The deviations relative to the nominal weight of the combinations of QCD renormalisation and factorisation scale variations. They are presented as a function of  $p_T^{\text{miss}}$  in the  $t\bar{t}H$  categories for the  $t\bar{t}$  samples. Events are from the signal region in the 2017 dataset after the analysis-level selection.

For a top or antitop, the scale factor is given by the value of the fit function. The event weight is then the square root of the product of the top and antitop scale factors. We assume the uncertainty on the weight is equal to it, making the downward variation equal to unity and the upward variation the square of the weight.<sup>13</sup>

#### 6.4.7 Object-level scale factors and systematic 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. 6.4.6.2. In the following sections, weights attributed to objects themselves are referred to as *scale factors*, while the event weight is the product of the scale factors as in Eq. 6.2. Systematic uncertainties are aggregated in the same manner. Unless stated otherwise, the following scale factors and systematics are derived separately for each year in Run-2.

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

Electrons and photons are identified using the cut-based identification described in Chpts. 5.2.3 and 5.2.4, 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

<sup>13</sup>Not sure if I should actually give the equation of the fit function.

are derived as a function of  $\eta$  and ECAL supercluster  $p_T$  and binned as such. 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.<sup>14</sup> 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.

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. An equivalent reconstruction efficiency is not given as the dedicated muon chambers and reconstruction algorithms are more than sufficient.<sup>15</sup>

#### 6.4.7.2 $b$ -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. 128. 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.<sup>16</sup>

#### 6.4.7.3 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 uncertainties associated with the scale factors are also derived separately, according to Sec. 8.1 of Ref. 112.

### 6.4.8 Minor weights and systematic uncertainties

Derived by members of the Collaboration, the nominal weight applied to events in simulation that emulates the effect of pre-firing in 2016 and 2017 data, is given in Ref. 59:

$$(6.11) \quad w_{\text{pre-firing}} = \prod_{i=\gamma, \text{jet}} 1 - \epsilon_i^{\text{pre-firing}}(\eta, p_T^{\text{EM}})$$

---

<sup>14</sup>Not sure how we correlate between years.

<sup>15</sup>For electrons, muons, and photons, not really sure of the actual procedures for getting the scale factors for each source. Twikis do a pretty poor job of explaining them.

<sup>16</sup>We don't seem to take any explicit systematics at the skimming stage where our  $b$ -jets are defined and SFs taken. So I don't really know where the up and down variations come from.

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). The systematic uncertainty in the event weight is found by taking the uncertainty on  $\epsilon$ : the maximum of  $0.2\epsilon$  and the statistical uncertainty in the  $(\eta, p_T^{\text{EM}})$  bin of the pre-firing map from which  $\epsilon$  is extracted. Pre-firing probabilities are assumed to be uncorrelated between photons and jets.<sup>17</sup>

## 6.5 Event selection

The event selection aims to strike a balance between rejecting as many background events while maintaining as great a signal sensitivity as possible. The preselection, in Chpt. 6.5.1, is applied to data and simulation in all regions and categories to do just that. Filters to reject potentially-mismeasured events and those that lead to incorrect  $p_T^{\text{miss}}$  calculations are documented in Chpt. 6.5.2. A strategy to combat the HEM issue faced in 2018 (detailed in Chpt. 6.3.1) is given in Chpt. 6.5.3.

### 6.5.1 Preselection

The preselection is characterised by applying the following cuts:<sup>18</sup>

- $p_T^{j_1} > 80 \text{ GeV}$
- $H_T > 200 \text{ GeV}$
- $H_T^{\text{miss}} > 200 \text{ GeV}$
- $p_T^{\text{miss}} > 200 \text{ GeV}$
- $H_T^{\text{miss}}/p_T^{\text{miss}} < 1.2$  (inverted for some of the QCD-enriched sidebands)
- Filters defined in Chpt. 6.5.2

To ensure orthogonality with the phase space occupied by their subset of the analysis, we also require  $|\eta_{j_1}| < 2.4$ ,  $|\eta_{j_2}| < 2.4$  if  $n_{\text{jet}} > 1$ , and a veto of the VBF kinematic selection:

- $p_T^{j_1} > 80 \text{ GeV}$
- $p_T^{j_2} > 40 \text{ GeV}$
- $|\eta_{j_1}| < 5.0$
- $|\eta_{j_2}| < 5.0$

---

<sup>17</sup>Do we correlate the systematic between years?

<sup>18</sup>Do I have to explain the reasoning behind each cut in the preselection?

- $\eta_{j_1} \cdot \eta_{j_2} < 0$
- $p_T^{\text{miss}} \geq 250 \text{ GeV}$
- $|\Delta\eta_{j_1 j_2}| > 1$
- $m_{jj} > 200 \text{ GeV}$
- $\Delta\phi(j_1, j_2) < 1.5$
- $\Delta\phi_{\min}(j, p_T^{\text{miss}}) > 0.5$

### 6.5.2 Additional filters

Further selections are applied to all years, regions and categories to filter poorly measured or mis-reconstructed events in both data and MC. 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  $f_E^\mu < 0.5$ , and  $\Delta\phi(j, p_T^{\text{miss}}) < \pi - 0.4$ .

Charged ( $f_E^\pm$ ) and neutral hadron energy fraction ( $f_E^0$ ) requirements are applied to all jets via fulfillment of the “tight” jet ID criteria (see Chpt. 5.2.1). Furthermore, stricter selections are placed on the leading two jets—applied in all categories except the monojet category, where they only apply to the leading jet—as follows:

- $f_E^\pm(j_1) > 0.1$
- $f_E^0(j_1) < 0.8$
- $f_E^\pm(j_2) > 0.1$
- $f_E^0(j_2) < 0.8$

Events with anomalously energetic jets have been observed for data in the HF between  $|\eta| \in [3.0, 3.1]$  that can lead to excess events, particularly in the QCD-enriched sidebands. As such, a stipulation is placed on the ratio of the  $H_T$  formed from jets within  $|\eta| < 5$  ( $H_T^5$ ) and  $|\eta| < 2.4$  ( $H_T^{2.4}$ ), in OR with the difference in  $\phi$  between the leading jet and  $H_T^{\text{miss}}$ :

$$\circ \frac{H_T^5}{H_T^{2.4}} < 1.2 \text{ or } \Delta\phi(j_1, H_T^{\text{miss}}) \geq \left( 5.3 \frac{H_T^5}{H_T^{2.4}} - 4.78 \right)$$

where, if the former requirement is not met, the latter is constructed such that the  $\Delta\phi \geq \pi/2$ . The  $H_T$  used in the analysis (i.e, in the preselection) only places  $|\eta|$  circumscriptions on the two hardest jets from to the VBF orthogonality constraints in Chpt. 6.5.1.

To avoid any potential horns (spikes in data and/or MC) for jets at large pseudorapidity, we require the agreement between the charged hadron-subtracted  $E_T^{\text{miss}}$  and calorimeter-calculated  $E_T^{\text{miss}}$  to be within 20 %:<sup>19</sup>

$$\circ 1 - \frac{E_T^{\text{miss}}(\text{CHS})}{E_T^{\text{miss}}(\text{calo})} < 0.8$$

The filters described below were recommended to remove events with miscalculated  $p_T^{\text{miss}}$ :

- 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.

### 6.5.3 Mitigating the HEM issue

In 2018, requirements were placed on electrons and the azimuthal angle of the  $\vec{p}_T^{\text{miss}}$  to mitigate the HEM issue (see Chpt. 6.3.1). As such, the region-dependent selections for 2018 were updated to remove events that met the following:

- $\circ -1.8 < \phi(\vec{p}_T^{\text{miss}}) < -0.6$  in the signal region and sidebands
- $\circ$  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  $e + \text{jets}$  control region
- $\circ$  Any loose photon  $\gamma_{\text{loose}}$  with  $p_T > 15 \text{ GeV}$ ,  $-3.0 < \eta < -1.4$ , and  $-1.57 < \phi < -0.87$  in the  $\gamma + \text{jets}$  control region

The effect of these cuts can be seen in Figs. 6.3 and 6.4.

## 6.6 Categorisation of the non-VBF production modes

To extract maximum sensitivity to the  $H \rightarrow \text{inv. decay}$  from the analysis, categories are established to target each of the production modes. From first principles, the topologies

---

<sup>19</sup> Again, I try to use the symbol  $p_T^{\text{miss}}$  everywhere. But every instance of the CHS and calo MET has used  $E_T^{\text{miss}}$  symbol, as far as I can tell. Should I swap to  $p_T^{\text{miss}}$ ?

## CHAPTER 6. COMBINED SEARCH FOR INVISIBLY DECAYING HIGGS BOSONS IN HADRONIC CHANNELS

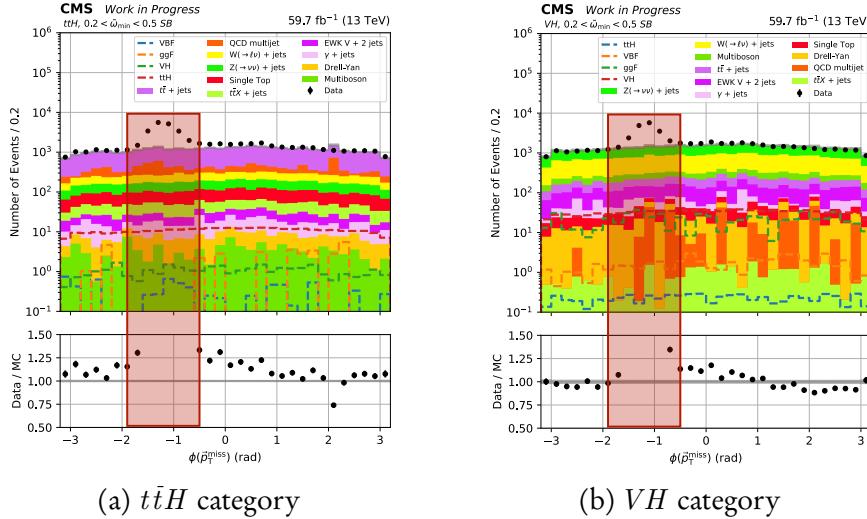


Figure 6.3: The azimuthal angle of the  $\vec{p}_T^{\text{miss}}$  in the  $t\bar{t}H$  and  $VH$  categories before applying the selections designed to mitigate the HEM issue in 2018. The loose  $\tilde{\omega}_{\min}$  sideband is used to demonstrate the effect since it kinematically resembles the signal region, and the data-simulation discrepancy can be removed while still blind in said region. A red box encloses the sector that is removed by the selection applied in the signal region and sidebands.

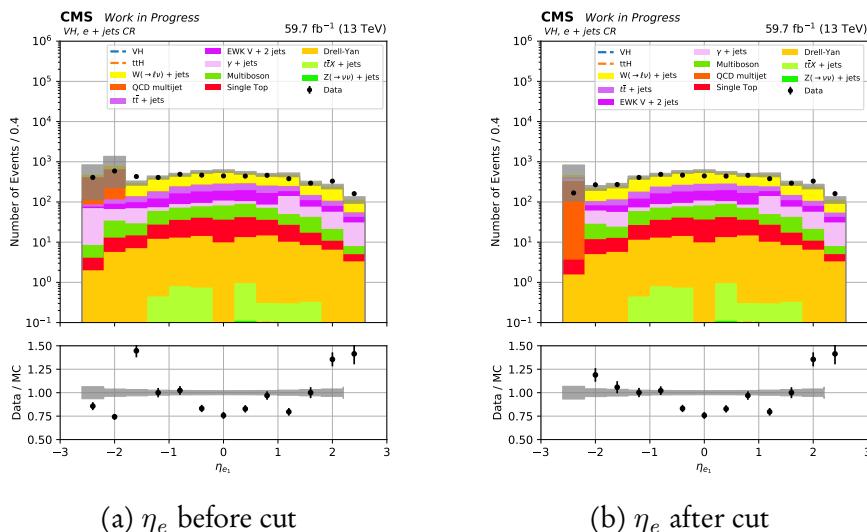


Figure 6.4: The psuedorapidity of the electron in the  $VH$  category of the  $e + \text{jets}$  control region before applying the selection designed to mitigate the HEM issue in 2018.

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 subcategories that capitalise on specific topologies within the mechanism—are orthogonal. An additional angular variable cut based on optimisation studies in Chpt. 6.6.2 is also implemented, devising the categories in Tab. 6.3.

Category	Subcategory	$n_{\text{jet}}$	$n_b$	$n_t$	$n_V$	$m_{jj}$	Angular variable
$t\bar{t}H$	2Boosted	$\geq 0$	$\geq 0$		2		
	1t0b	$\geq 3$	0	1	0		
	1t1b	$\geq 3$	1	1	0		
	1W1b	$\geq 3$	1	0	1		
	1W2b	$\geq 3$	2	0	1	—	$\tilde{\omega}_{\min} > 0.5$
	5j1b	5	1	0	0		
	6j1b	$\geq 6$	1	0	0		
	5j2b	5	$\geq 2$	0	0		
	6j2b	$\geq 6$	$\geq 2$	0	0		
$VH$	2j0b	2	0	0	0	$\in [65, 105)$	
	2j1b	2	1	0	0	$\in [65, 105)$	
	2j2b	2	2	0	0	$\in [65, 105)$	$\tilde{\omega}_{\min} > 0.5$
	1V	0	0	0	1	—	
$ggH$	2jM	2	0	0	0	$\notin [65, 105)$	
	3j	3	0	0	0	—	
	4j	4	0	0	0	—	
	5j	$\geq 5$	0	0	0	—	
Monojet	0b	1	0	0	0	—	
	1b	1	1	0	0	—	$\Delta\phi_{\min}(j, p_T^{\text{miss}}) > 2.5$

Table 6.3: Categorisation of the  $t\bar{t}H$ ,  $VH$  and  $ggH$  production modes in the analysis. Each subcategory highlights one of the possible final states of the mechanism, accounting for inefficiencies in object tagging or reconstruction. In the  $ggH$  2jM subcategory, the dijet mass requirement ensures orthogonality with  $VH$  2j0b.

The monojet category resembles the search in Ref. 127 and primarily targets the  $ggH$  mode, but is separated to distinguish it from the categories we have developed specifically for the analysis.

The number of boosted top quark- and vector boson-tagged jets ( $n_t$  and  $n_V$ , respectively) are defined in Chpt. 6.6.1. In the  $t\bar{t}H$  2Boosted category, if  $n_V = 2$  we also require  $n_b \geq 1$  for confidence that the  $V$  boson originates from a decaying  $t$  quark.

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 subcategory definition. The same treatment is applied to  $b$ -jets as well. This only applies to the categorisation and does not affect selections made on jets or  $b$ -jets in Chpt. 6.5.

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.

In the  $t\bar{t}H$  and  $VH$  categories, ancillary groupings of subcategories may be of interest: 2Boosted, 1t0b, 1t1b, 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 subcategories, appropriately named the “ $t\bar{t}H$  resolved” category; then, assembling the 2j0b, 2j1b, and 2j2b  $VH$  subcategories elicits the “ $VH$  resolved” moniker. Low yields may be observed in individual subcategories, so using these grouped alternatives is useful to quickly inspect the wider effect of changes to the analysis.

### 6.6.1 Classifying boosted topologies from top quarks, and $W$ and $Z$ 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 particles. In this analysis, we pay specific interest 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 number of tagged jets in an event inform the boosted subcategories, as seen in Tab. 6.3. The design and optimisation of the algorithm is described in Ref. 112.<sup>20</sup>

We employ the ensuing selections 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 “loose” working point (2.5 % mis-tag rate)
- Tight ID requirement for AK8 jets

– Boosted  $V$ :

---

<sup>20</sup>Not sure if I should give a summary of the development and features of the algorithm here.

- $p_T > 200 \text{ GeV}$
- $65 < m_{\text{SD}} < 105 \text{ GeV}$
- $W$  vs. QCD discriminator score with the “loose” working point (5 % mis-tag rate)
- Tight ID requirement for AK8 jets

Orthogonality is ensured between the two classifications through the soft drop mass  $m_{\text{SD}}$  [109] window. We also use the loose working points to maximise the number of events that remain in the relevant categories. Otherwise, we suffer from very poor statistics in some of the control regions. The boosted  $V$  jets are not further separated into  $W$ - and  $Z$ -tagged since the mass degeneracy makes this inherently difficult. After correspondence with the DEEPAK8 developers, we were reassured that we could use the  $W$  vs. QCD discriminator to tag both  $W$  and  $Z$  jets. An additional recommendation was to use the nominal version of the tagger as opposed to the mass-decorrelated version. The performance improvements are noticeable, as shown in Ref. 112, and we do not use a mass dimension for our fitting variable which would otherwise preclude the use of the latter version. Data-MC scale factors to weight events containing these objects are described in Chpt. 6.4.7.3.

### 6.6.2 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  $t\bar{t}H$  and  $ggH$  categories. Historically, variables such as  $\Delta\phi_{\min}^*$  and  $\alpha_T$  have been used to suppress it, as in the previous supersymmetry analysis I was a part of.

Recently, more elaborate variables have been developed to better remove multijet background events in analyses with hadronic final states. Colleagues at the University of Bristol have designed a family of them:  $\chi_{\min}$ ,  $\hat{\omega}_{\min}$ , and  $\tilde{\omega}_{\min}$  [120].<sup>21</sup> These show noticeable improvement over previous variables (particularly  $\Delta\phi_{\min}^*$ , which is used as a benchmark for comparisons). Optimisation was performed for these variables on a per-category, over a per-subcategory, basis to avoid excessive fine tuning for the many subcategories in the analysis. After investigating each variable,  $\tilde{\omega}_{\min}$  was chosen for the  $t\bar{t}H$ ,  $VH$ , and  $ggH$  categories. A more traditional  $\Delta\phi_{\min}(j, p_T^{\text{miss}})$  was selected for the monojet category, as the aforementioned ones typically perform better at higher jet multiplicity.

---

<sup>21</sup>How much detail do I need to go into regarding the explanations of these variables (or just the one we end up using)?

Various metrics (or figures of merit) were considered for the threshold to cut on the given angular variable. 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

$$(6.12) \quad Z_{\text{Poisson}} = \frac{S}{\sqrt{B}}$$

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

$$(6.13) \quad Z_{\text{Poisson}} = \frac{S}{\sqrt{B + (\sigma_{B,\text{syst.}} B)^2}}$$

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 [67]. The median significance can then be extracted. Colloquially ascribed the *Asimov significance*. 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. 6.9. 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  $Z_{\text{Asimov}}$  can be expressed as

$$(6.14) \quad Z_{\text{Asimov}} = \sqrt{2 \left( (S + B) \cdot \ln \left( 1 + \frac{S}{B} \right) - S \right)}$$

reducing to Eq. 6.12 for  $S \ll B$ . Including a systematic uncertainty expands Eq. 6.14 to

$$(6.15) \quad Z_{\text{Asimov}} = \sqrt{2 \left( (S + B)c_1 - \frac{B^2}{\sigma_{B,\text{syst.}}^2}c_2 \right)}, \text{ where}$$

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

While these are quantitative measures of the sensitivity with an 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

## 6.6. CATEGORISATION OF THE NON-VBF PRODUCTION MODES

cut for each category in the final column of Tab. 6.3, we first processed all simulated signal and background events through the analysis. They were categorised according to the aforementioned table (of course, excluding the angular variable cut). For each subcategory, 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 category, the significances for each subcategory were summed in quadrature. Fig. 6.5 demonstrates the results as a function of  $\tilde{\omega}_{\min}$  for the  $t\bar{t}H$ ,  $VH$ , and  $ggH$  categories with the 2017 datasets.

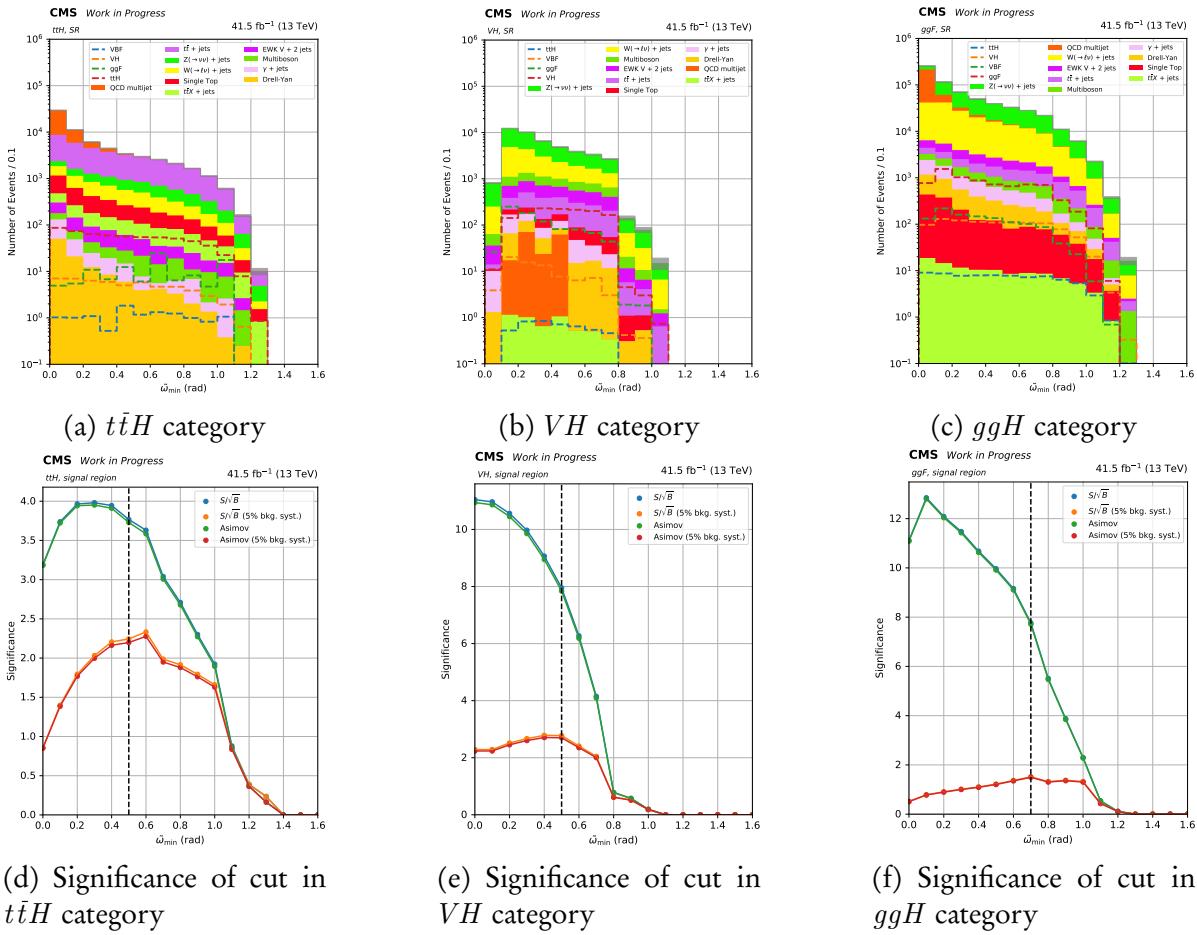


Figure 6.5: Top row: distributions of  $\tilde{\omega}_{\min}$  in the signal region in the  $t\bar{t}H$ ,  $VH$ , and  $ggH$  categories. Bottom row: the significance—using several figures of merit—if a cut is placed to the right of a given value of  $\tilde{\omega}_{\min}$  in each category. The black 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.

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. The control regions in the analysis perform better estimations of their responsible backgrounds when adequately populated. Reducing the background too greatly with the  $\tilde{\omega}_{\min}$  cut can therefore cause problems in multiple aspects of the analysis.

### 6.6.3 Binning

In addition to the categories above, events are placed in bins of  $p_T^{\text{miss}}$  as that distribution is expected to maximally differentiate signal and background in the final fit (see Chpt. 6.9). Since the number of events can significantly differ between categories in the different regions of the analysis, a global binning configuration is inadequate. For each category, 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 transfer factors from the background estimation methods are calculated on a bin-by-bin basis.<sup>22</sup>

## 6.7 Signal region, control region, and sideband definitions

Several regions of phase space are explored in the analysis. The signal region is accompanied by control regions and sidebands to aid in the prediction of certain backgrounds, as explained below. The preselection is applied to every region, so they are separated by the CMS datasets used, the High-Level Trigger requirements, and object and/or kinematic selections.

### 6.7.1 The signal region

The signal region is the area of parameter space where we expect the signal to manifest, and the background to be reduced sufficiently that the potential presence of signal in data can be statistically verified. In this analysis, only hadronic final states are permitted. Events with leptons, photons and taus meeting the “loose” or “veto” criteria in Chpt. 5.2 are not explicitly vetoed, however, but are given a weight as described in Chpt. 6.4.1.

---

<sup>22</sup>Not sure if this subsection should instead go in the fit discussion. But as I haven’t written any of that, I’ll leave it here for now.

## 6.7. SIGNAL REGION, CONTROL REGION, AND SIDEBAND DEFINITIONS

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, and at least one jet fulfilling the tight ID criteria (see Chpt. 5.2). The triggers are illustrated in Tab. 6.4.

Year	$p_T^{\text{miss}, \mu}$ (GeV)	$H_T^{\text{miss}, \mu}$ (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 6.4: 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 variables  $p_T^{\text{miss}}$  and  $H_T^{\text{miss}}$  do not include muons in the calculations.

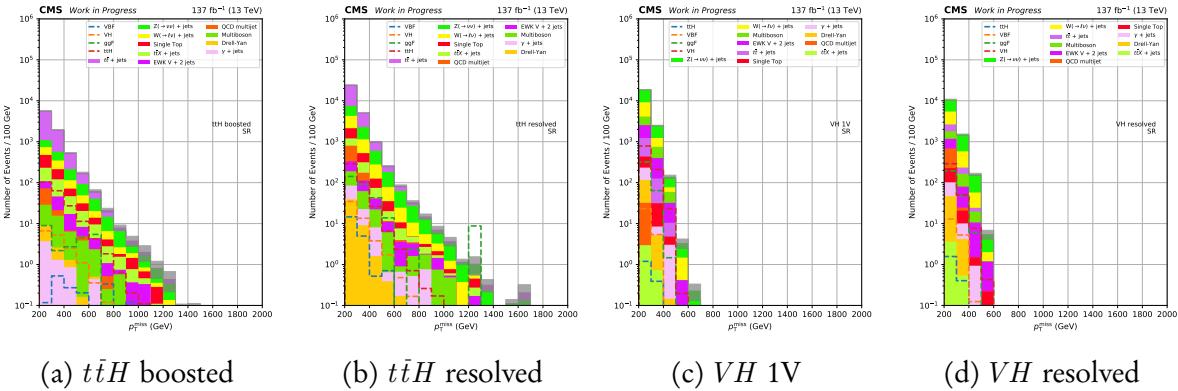


Figure 6.6: Pre-fit yields of the  $p_T^{\text{miss}}$  distribution in the  $\mu + \text{jets}$  control region for the combined boosted and resolved categories for the  $t\bar{t}H$  and  $VH$  processes, using simulation derived for each year in Run-2 and scaled to the required luminosity.

### 6.7.2 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 or weights 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 also designed to, ideally, be devoid of signal. 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. 5.2—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. 5.7) 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  $60 < m_{\mu\mu} < 120 \text{ GeV}$
- $e + \text{jets}$ : 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  $60 < m_{ee} < 120 \text{ GeV}$
- $\gamma + \text{jets}$ : one medium photon  $\gamma_{\text{medium}}$  with  $p_T > 230 \text{ GeV}$

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  $t\bar{t}$ . The trigger requirements for the  $\mu + \text{jets}$  and  $\mu\mu + \text{jets}$  control regions are the same as for the signal region (Tab. 6.4) since the same primary dataset is used.

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. 6.5 and 6.6 elucidate how the trigger requirements are specified for each year. As before, each quantity and object is 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. 6.5 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. 6.5 and the passing of any of the photon triggers in Tab. 6.6 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. 6.5 and 6.6 may be satisfied. For Monte Carlo events, as they are not categorised by trigger, they may pass any of the triggers in either table for their respective year.<sup>23</sup>

For the  $\gamma + \text{jets}$  control region, we take data from CMS originating only from the dataset of  $\gamma$ -based triggers. Data and MC must satisfy any of the triggers from Tab. 6.6 for the

---

<sup>23</sup>See if I can describe the trigger requirements in a more elegant way. Might be easier give the descriptions and actual HLT path names a bit earlier, then just reference the HLT path names in the list.

## 6.7. SIGNAL REGION, CONTROL REGION, AND SIDEBAND DEFINITIONS

---

Year	$e$ SC $E_T$ 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 6.5: 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  $E_T$  of the supercluster (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.

respective year.

Year	$\gamma$ $E_T$ threshold (GeV)	$H/E$
2016	165	< 0.1
	175	—
2017	200	—
2018	200	—

Table 6.6: 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  $E_T$  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.

As explained in Chpt. 5.2.6, the  $\vec{p}_T^{\text{miss}}$  is recalculated for 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  $\vec{p}_T$ . The single lepton control regions estimate the semi-leptonic  $t\bar{t} + \text{jets}$  and  $W(\rightarrow \ell\nu) + \text{jets}$  backgrounds. In both cases, they may enter the signal region if the lepton is lost, becoming a source of  $p_T^{\text{miss}}$ . The dilepton and  $\gamma + \text{jets}$  regions consider  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ . The former parallels the decay as it is enriched in  $Z(\rightarrow \ell^+\ell^-) + \text{jets}$ —possessing very similar 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.

The following figures reveal the pre-fit  $p_T^{\text{miss}}$  distributions in each control region after the analysis-level selections with the full Run-2 dataset. The combined  $t\bar{t}H$  and  $VH$  subcategories for boosted and resolved topologies are used to demonstrate the shapes and data–simulation agreement. Background estimation is performed separately for each year due to the differing running conditions, detector configuration, and the different effects or problems seen in the

## CHAPTER 6. COMBINED SEARCH FOR INVISIBLY DECAYING HIGGS BOSONS IN HADRONIC CHANNELS

data.<sup>24</sup>

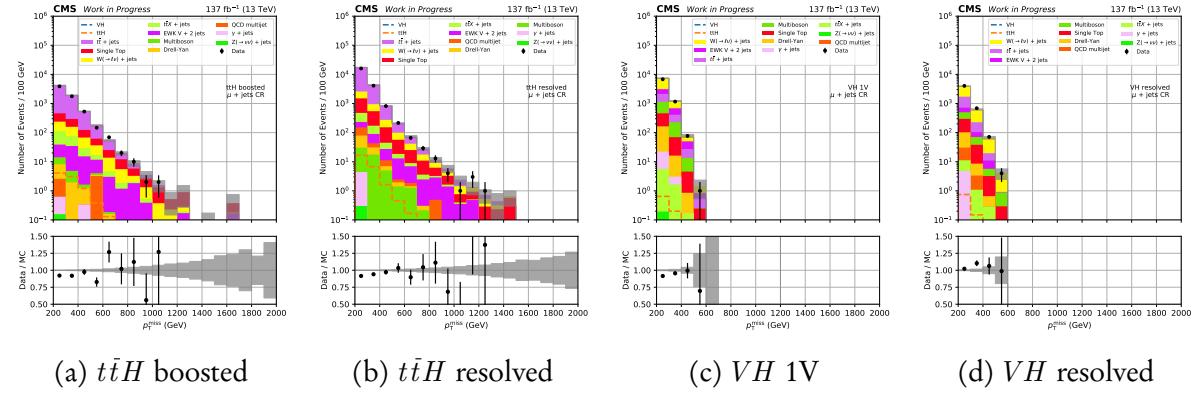


Figure 6.7: Pre-fit yields of the  $p_T^{\text{miss}}$  distribution in the  $\mu + \text{jets}$  control region for the combined boosted and resolved categories for the  $t\bar{t}H$  and  $VH$  processes, using the full Run-2 dataset for data and simulation.

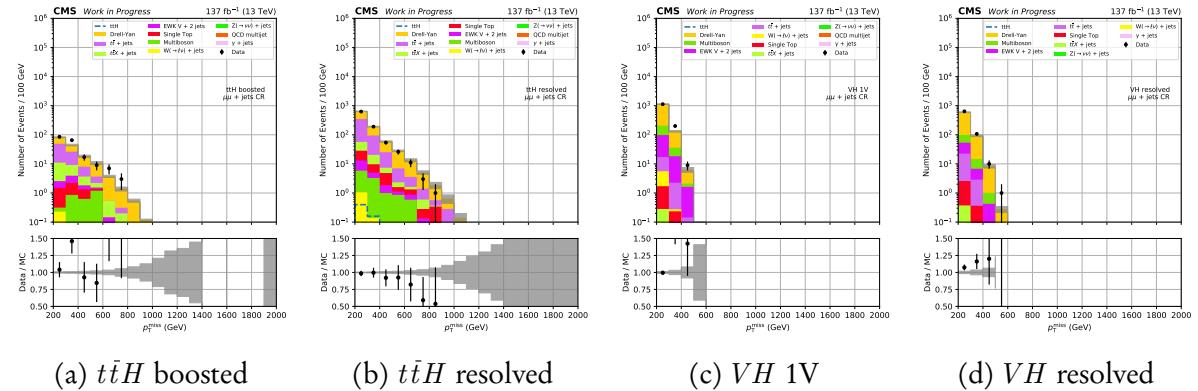


Figure 6.8: Pre-fit yields of the  $p_T^{\text{miss}}$  distribution in the  $\mu\mu + \text{jets}$  control region for the combined boosted and resolved categories for the  $t\bar{t}H$  and  $VH$  processes, using the full Run-2 dataset for data and simulation.

### 6.7.3 Sidebands to the signal region

As explained in Chpt. 6.3.3, it is difficult to accurately estimate the QCD multijet content in the signal region. Several kinematic cuts in the analysis are designed to reject QCD in the signal region, so by inverting one or two of them, we can construct multijet-enriched regions: *sidebands*.

<sup>24</sup>Might not need to show all of these plots. They're just here for now, and can be tidied up later.

## 6.7. SIGNAL REGION, CONTROL REGION, AND SIDEband DEFINITIONS

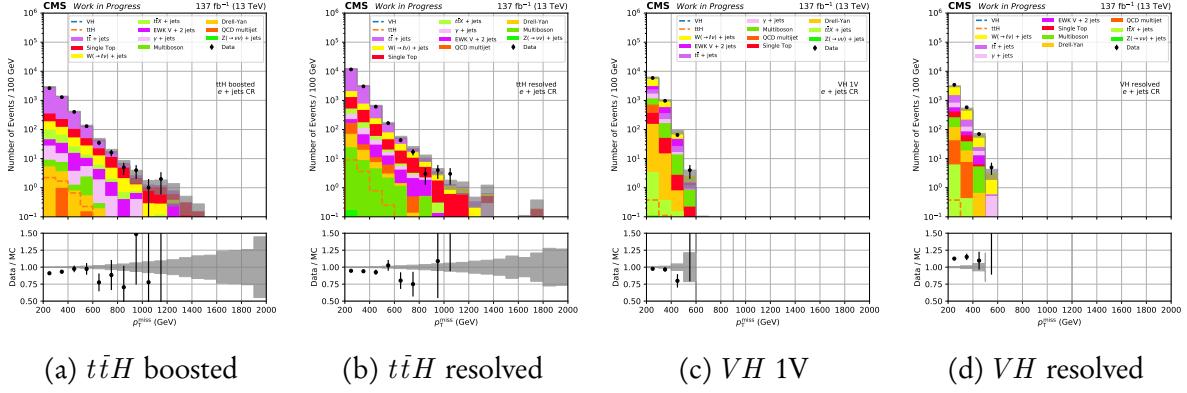


Figure 6.9: Pre-fit yields of the  $p_T^{\text{miss}}$  distribution in the  $e + \text{jets}$  control region for the combined boosted and resolved categories for the  $t\bar{t}H$  and  $VH$  processes, using the full Run-2 dataset for data and simulation.

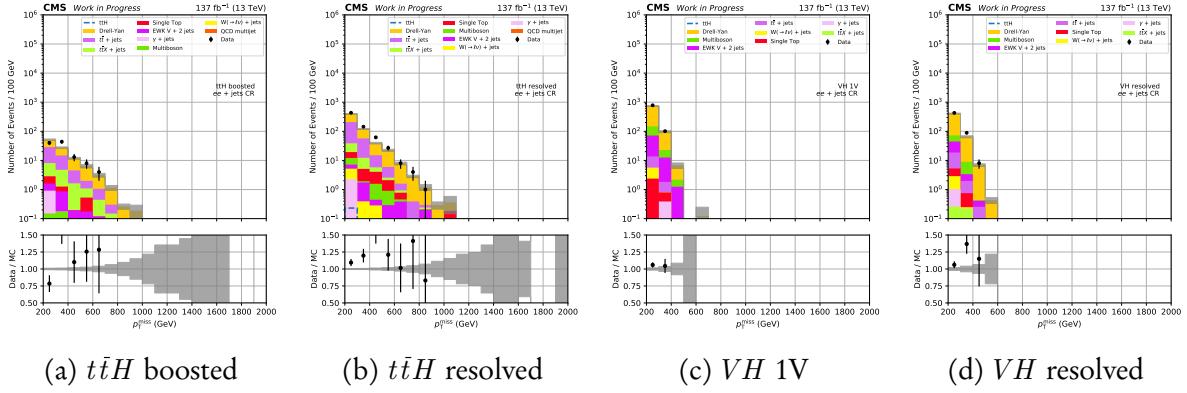


Figure 6.10: Pre-fit yields of the  $p_T^{\text{miss}}$  distribution in the  $ee + \text{jets}$  control region for the combined boosted and resolved categories for the  $t\bar{t}H$  and  $VH$  processes, using the full Run-2 dataset for data and simulation.

We invert the requirements on  $H_T^{\text{miss}}/p_T^{\text{miss}}$ —where a ceiling is placed at  $H_T^{\text{miss}}/p_T^{\text{miss}} = 4$  to reject severely-mismeasured jets—and the angular variable designed to optimise the categorisation of the production modes (see Chpt. 6.6.2). Several sidebands are constructed where each cut is inverted separately, and then when both are inverted. When the angular cut is inverted, the remaining phase space may be split into two sidebands, dubbed “loose” and “tight.” These are summarised for the different categories in Tabs. 6.7, 6.8, and 6.9.

The sidebands follow the same categorisation as for the other regions, i.e., Tab. 6.3, and are binned in  $p_T^{\text{miss}}$  with the same scheme as the other regions.

A sideband composed of the inversion of only one variable maintains similar kinematic properties to the signal region. When performing analysis while still blind to data in the signal region, inspecting one of these sidebands is a good indicator of how new cuts or corrections

## CHAPTER 6. COMBINED SEARCH FOR INVISIBLY DECAYING HIGGS BOSONS IN HADRONIC CHANNELS

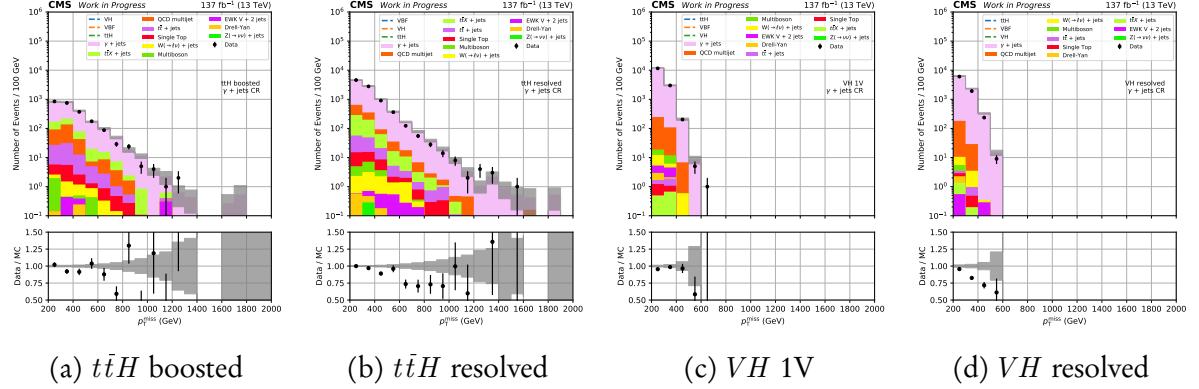


Figure 6.11: Pre-fit yields of the  $p_T^{\text{miss}}$  distribution in the  $\gamma + \text{jets}$  control region for the combined boosted and resolved categories for the  $t\bar{t}H$  and  $VH$  processes, using the full Run-2 dataset for data and simulation. The  $\gamma + \text{jets}$  is more useful in predicting  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$  in the  $VH$  subcategories, and may not be used for the  $t\bar{t}H$  subcategories.

	$\tilde{\omega}_{\min} < 0.2$	$0.2 \leq \tilde{\omega}_{\min} \leq 0.5$	$\tilde{\omega}_{\min} > 0.5$
$H_T^{\text{miss}}/p_T^{\text{miss}} < 1.2$	“Tight” $\tilde{\omega}_{\min}$ SB	“Loose” $\tilde{\omega}_{\min}$ SB	Signal region
$1.2 \leq H_T^{\text{miss}}/p_T^{\text{miss}} < 4$	“Tight” double SB	“Loose” double SB	$H_T^{\text{miss}}/p_T^{\text{miss}}$ SB

Table 6.7: Definitions of the sidebands (SBs) used to determine the QCD multijet background in the signal region for the  $t\bar{t}H$  and  $VH$  categories.

	$\tilde{\omega}_{\min} < 0.2$	$0.2 \leq \tilde{\omega}_{\min} \leq 0.7$	$\tilde{\omega}_{\min} > 0.7$
$H_T^{\text{miss}}/p_T^{\text{miss}} < 1.2$	“Tight” $\tilde{\omega}_{\min}$ SB	“Loose” $\tilde{\omega}_{\min}$ SB	Signal region
$1.2 \leq H_T^{\text{miss}}/p_T^{\text{miss}} < 4$	“Tight” double SB	“Loose” double SB	$H_T^{\text{miss}}/p_T^{\text{miss}}$ SB

Table 6.8: Definitions of the sidebands (SBs) used to determine the QCD multijet background in the signal region for the  $ggH$  categories.

	$0.1 \leq \Delta\phi_{\min}(j_{1,2,3,4}, p_T^{\text{miss}}) \leq 2.5$	$\Delta\phi_{\min}(j_{1,2,3,4}, p_T^{\text{miss}}) > 2.5$
$H_T^{\text{miss}}/p_T^{\text{miss}} < 1.2$	$\Delta\phi_{\min}(j_{1,2,3,4}, p_T^{\text{miss}})$ sideband	Signal region
$1.2 \leq H_T^{\text{miss}}/p_T^{\text{miss}} < 4$	Double sideband	$H_T^{\text{miss}}/p_T^{\text{miss}}$ sideband

Table 6.9: Definitions of the sidebands used to determine the QCD multijet background in the signal region for the monojet categories.

## 6.7. SIGNAL REGION, CONTROL REGION, AND SIDEband DEFINITIONS

will affect the signal region composition, and compatibility between data and MC. Like the control regions, sidebands serve a secondary purpose.

The following figures reveal the pre-fit  $p_T^{\text{miss}}$  distributions in each sideband after the analysis-level selections with the full Run-2 dataset. The combined  $t\bar{t}H$  and  $VH$  subcategories for boosted and resolved topologies are used to demonstrate the shapes and data–simulation (dis)agreement. Background estimation is performed separately for each year due to the differing running conditions, detector configuration, and the different effects or problems seen in the data.<sup>25</sup>

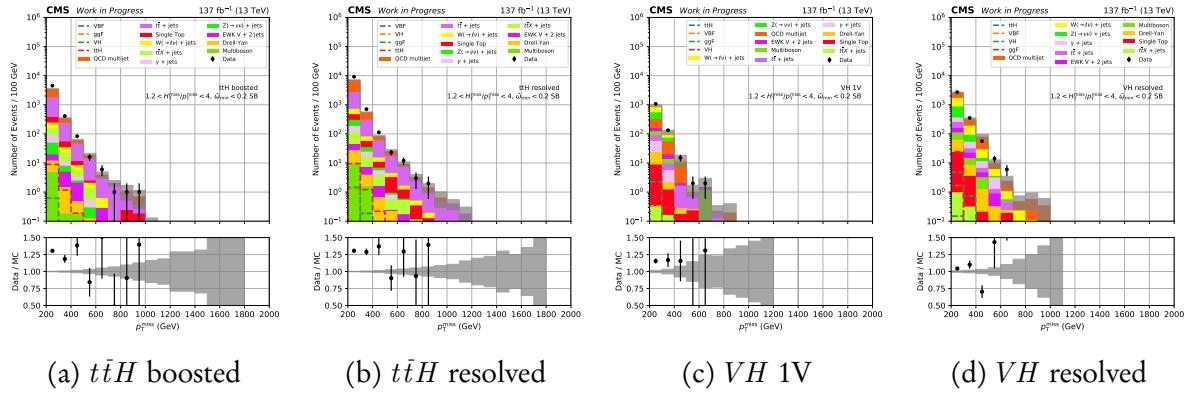


Figure 6.12: Pre-fit yields of the  $p_T^{\text{miss}}$  distribution in the tight double sideband for the combined boosted and resolved categories for the  $t\bar{t}H$  and  $VH$  processes, using the full Run-2 dataset for data and simulation.

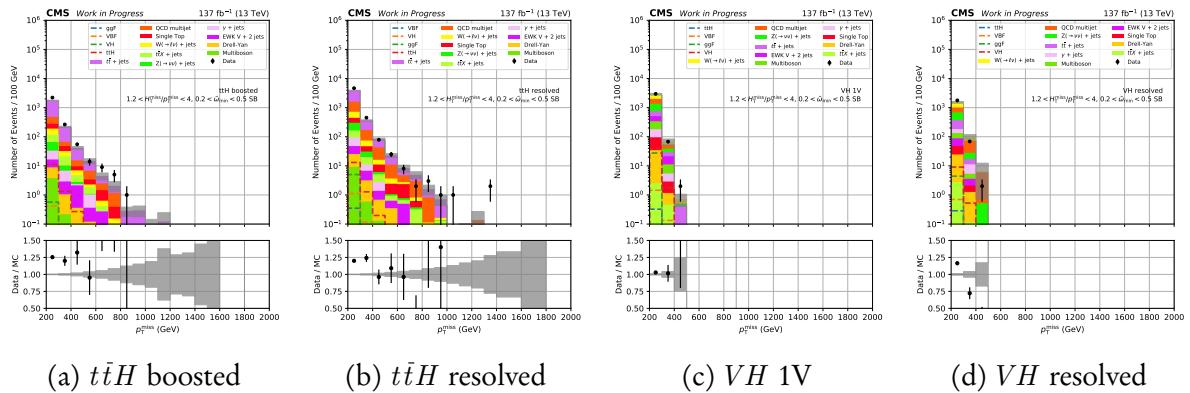


Figure 6.13: Pre-fit yields of the  $p_T^{\text{miss}}$  distribution in the loose double sideband for the combined boosted and resolved categories for the  $t\bar{t}H$  and  $VH$  processes, using the full Run-2 dataset for data and simulation.

<sup>25</sup>Might not need to show all of these plots. They’re just here for now, and can be tidied up later.

## CHAPTER 6. COMBINED SEARCH FOR INVISIBLY DECAYING HIGGS BOSONS IN HADRONIC CHANNELS

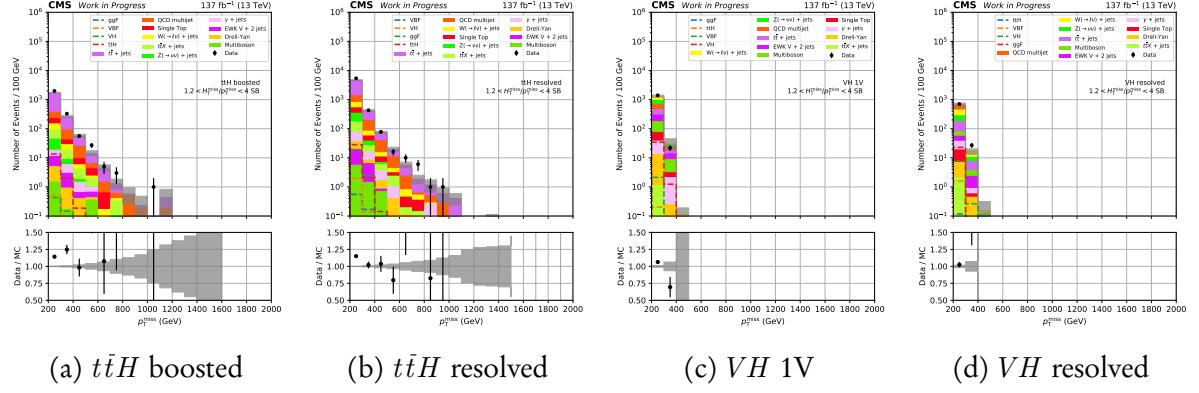


Figure 6.14: Pre-fit yields of the  $p_T^{\text{miss}}$  distribution in the  $H_T^{\text{miss}} / p_T^{\text{miss}}$  sideband for the combined boosted and resolved categories for the  $t\bar{t}H$  and  $VH$  processes, using the full Run-2 dataset for data and simulation.

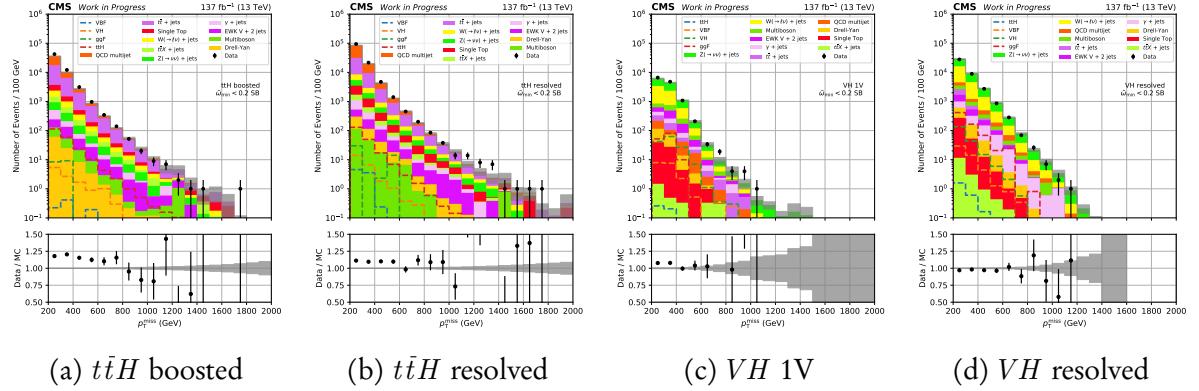


Figure 6.15: Pre-fit yields of the  $p_T^{\text{miss}}$  distribution in the tight  $\tilde{\omega}_{\min}$  sideband for the combined boosted and resolved categories for the  $t\bar{t}H$  and  $VH$  processes, using the full Run-2 dataset for data and simulation.

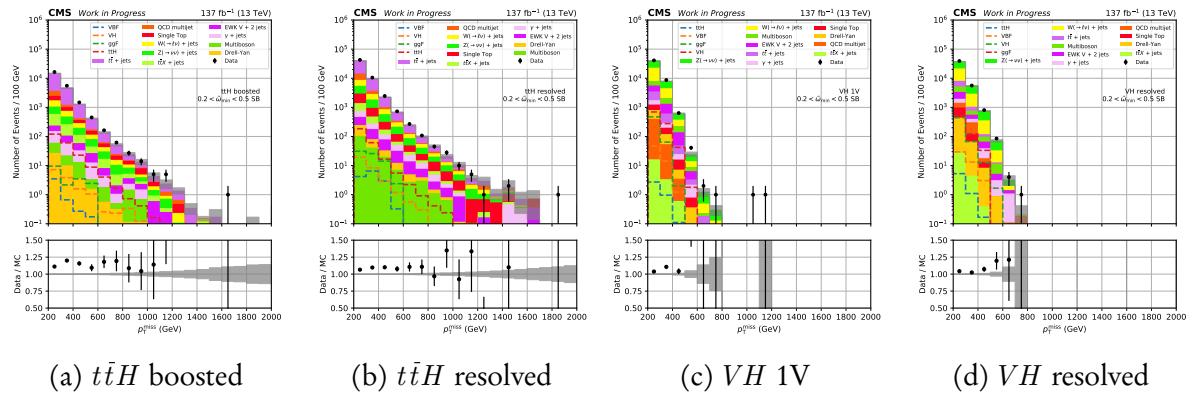


Figure 6.16: Pre-fit yields of the  $p_T^{\text{miss}}$  distribution in the loose  $\tilde{\omega}_{\min}$  sideband for the combined boosted and resolved categories for the  $t\bar{t}H$  and  $VH$  processes, using the full Run-2 dataset for data and simulation.

## 6.8 Background estimation

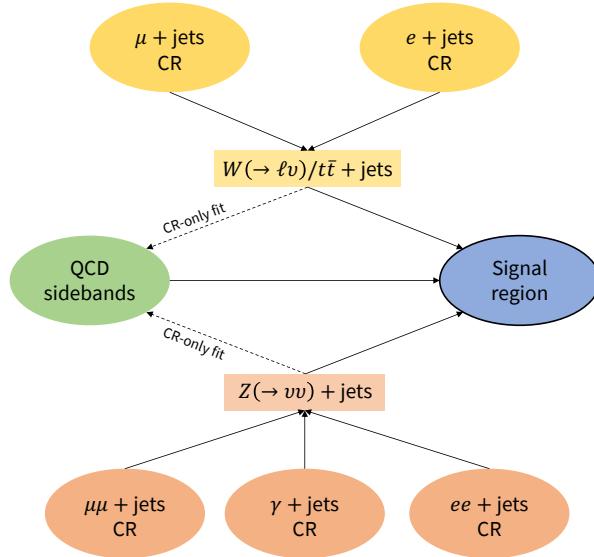


Figure 6.17: An infographic showcasing the role of each analysis region in the final fit. The control regions predict the lost lepton and  $Z \rightarrow \nu\bar{\nu}$  backgrounds, and a control region-only fit informs the QCD multijet prediction that contributes to the eventual background determination in the signal region.

### 6.8.1 Lost lepton ( $W$ and $t\bar{t} + \text{jets}$ )

### 6.8.2 $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$

### 6.8.3 QCD multijet

To reiterate what has been mentioned previously, estimating QCD multijet occupancy directly from simulation does not provide an adequate representation of the process. This can be mitigated by using a data-driven method to estimate it from the multijet-enriched sidebands described in Chpt. 6.7.3.

The effects of jet mismeasurements are, however, difficult to quantify. With a final state of several jets, low, or even no,  $p_T^{\text{miss}}$  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 process that suffers—jets from “cleaner” processes may also be affected—those with real  $p_T^{\text{miss}}$  in an event (e.g.,  $Z(\rightarrow \nu\bar{\nu}) + \text{jets}$ ) are unlikely to be significantly affected by one stray object. The enormous cross section of QCD multijet also 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.

## 6.9 Statistical model and fit

## 6.10 Analysis of the $t\bar{t}H$ mode

Fig. 6.18 showcases the median expected limit on  $\mathcal{B}(H \rightarrow \text{inv.})$  with  $1\sigma$  and  $2\sigma$  bounds for the  $t\bar{t}H$  category and its subcategories.<sup>26</sup>

## 6.11 Analysis of the $VH$ mode

Fig. 6.19 showcases the median expected limit on  $\mathcal{B}(H \rightarrow \text{inv.})$  with  $1\sigma$  and  $2\sigma$  bounds for the  $VH$  category and its subcategories.<sup>27</sup>

## 6.12 Analysis of the $ggH$ mode

Fig. 6.20 showcases the median expected limit on  $\mathcal{B}(H \rightarrow \text{inv.})$  with  $1\sigma$  and  $2\sigma$  bounds for the  $ggH$  category and its subcategories. The corresponding results from the monojet category are also given in Fig. 6.21.<sup>28</sup>

## 6.13 Combined results

Upper limits for  $\mathcal{B}(H \rightarrow \text{inv.})$  by combining all categories for a given year, and for the full Run-2 dataset are given in Figs. 6.22, 6.23, 6.24, and [REF].<sup>29</sup> Profile likelihood ratios as a function of  $\mathcal{B}(H \rightarrow \text{inv.})$  are also presented.

---

<sup>26</sup>Just placeholder plots for now for scenario 4. Plots made on 13th June.

<sup>27</sup>Just placeholder plots for now for scenario 4. Plots made on 13th June.

<sup>28</sup>Just placeholder plots for now for scenario 4. Plots made on 13th June.

<sup>29</sup>Expected limits only, so far, and really only placeholders. Probably don't need to show the likelihood scan for all subcategories.

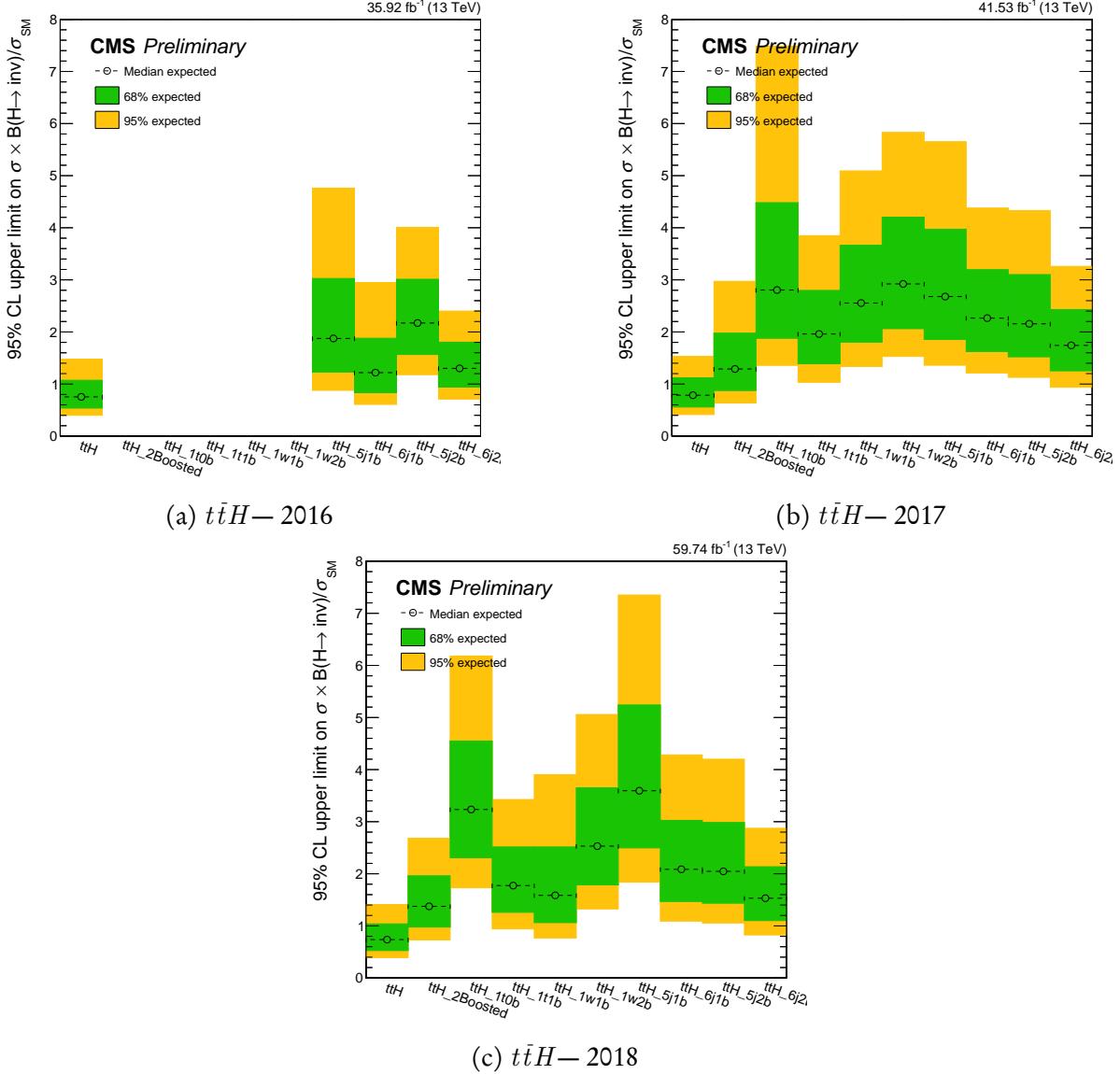


Figure 6.18: Expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the  $t\bar{t}H$  category, for both the individual subcategories, and the combination of them, for each data-taking year in Run-2.

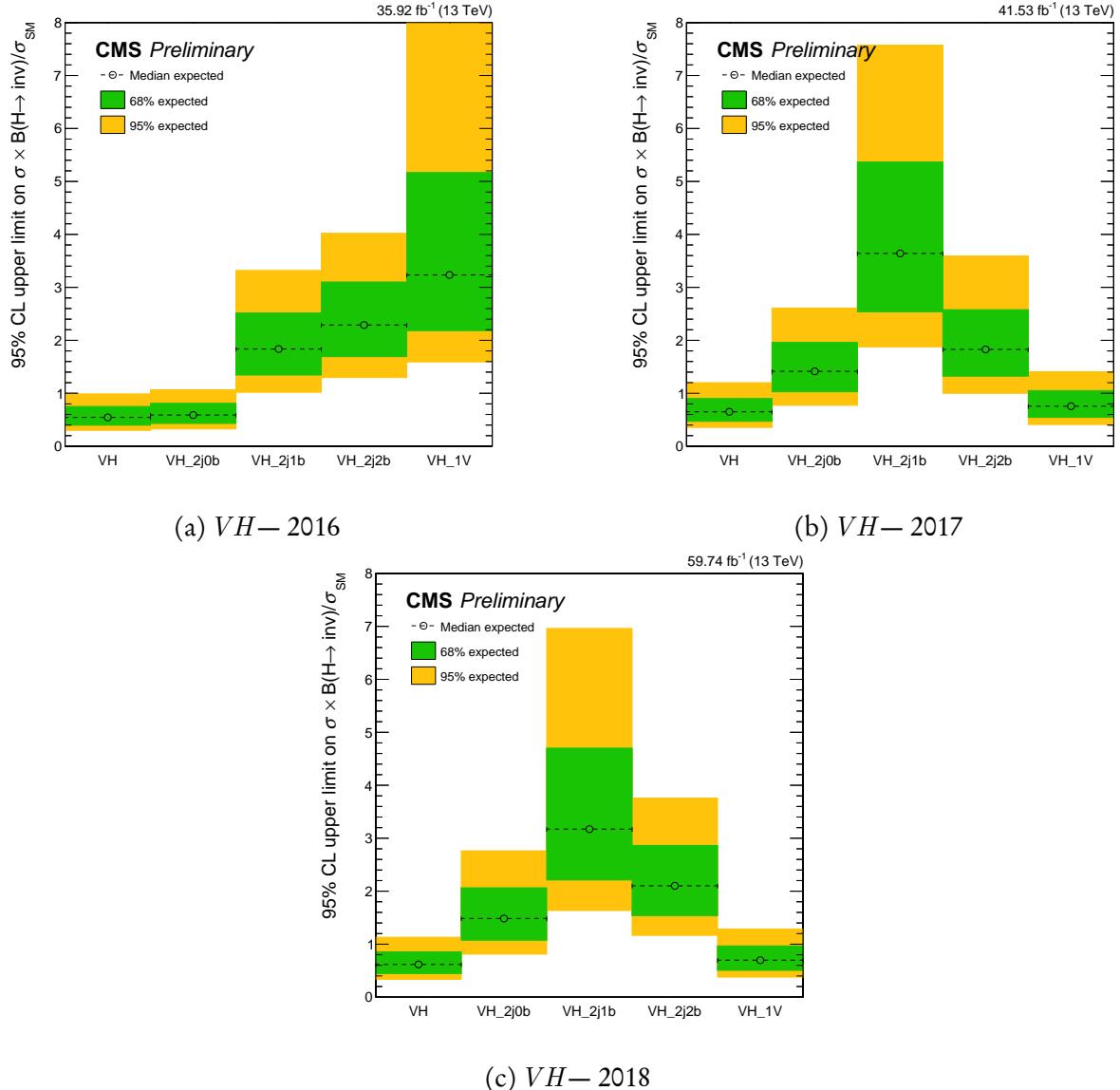


Figure 6.19: Expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the  $VH$  category, for both the individual subcategories, and the combination of them, for each data-taking year in Run-2.

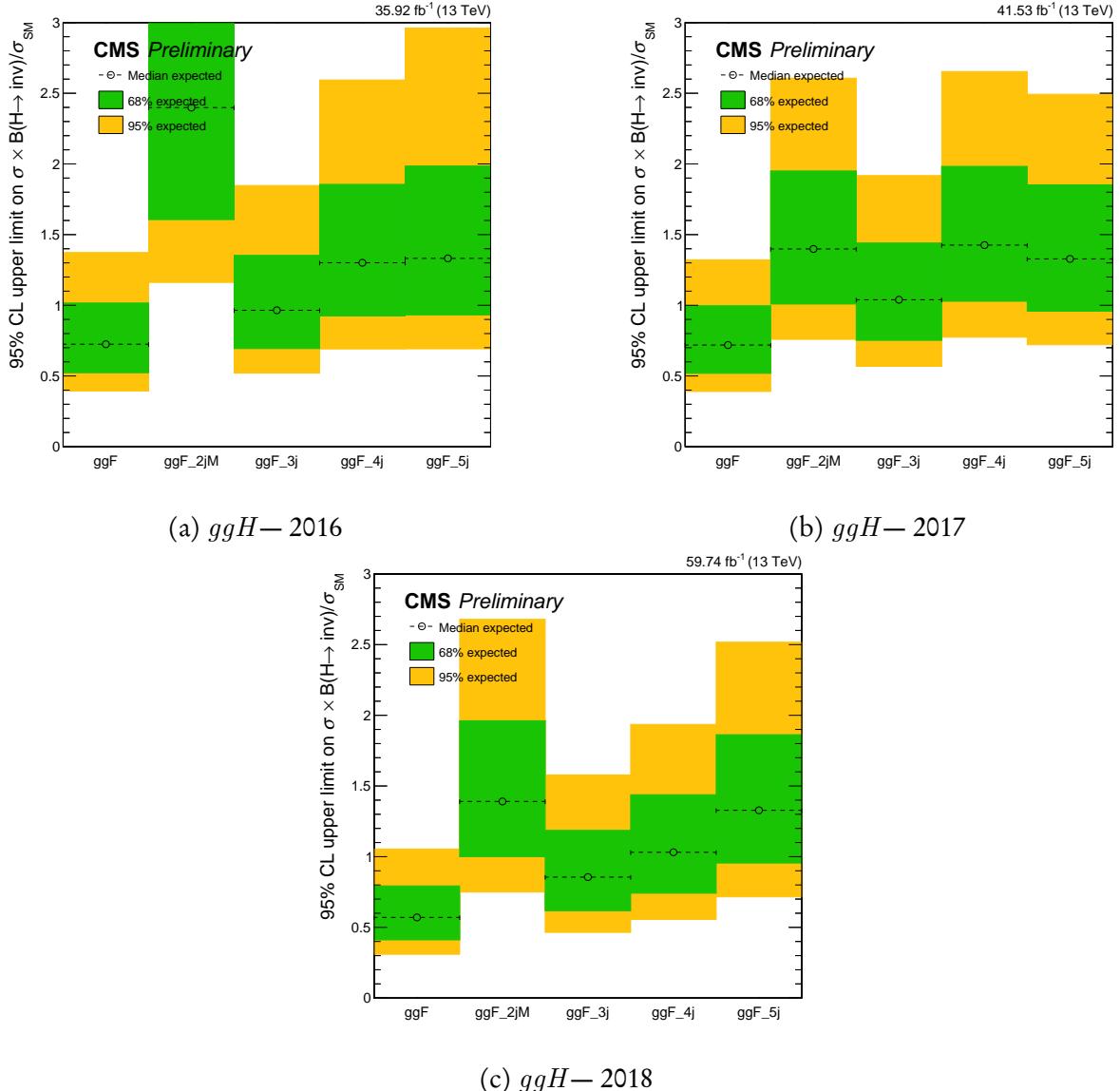


Figure 6.20: Expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the  $ggH$  category, for both the individual subcategories, and the combination of them, for each data-taking year in Run-2.

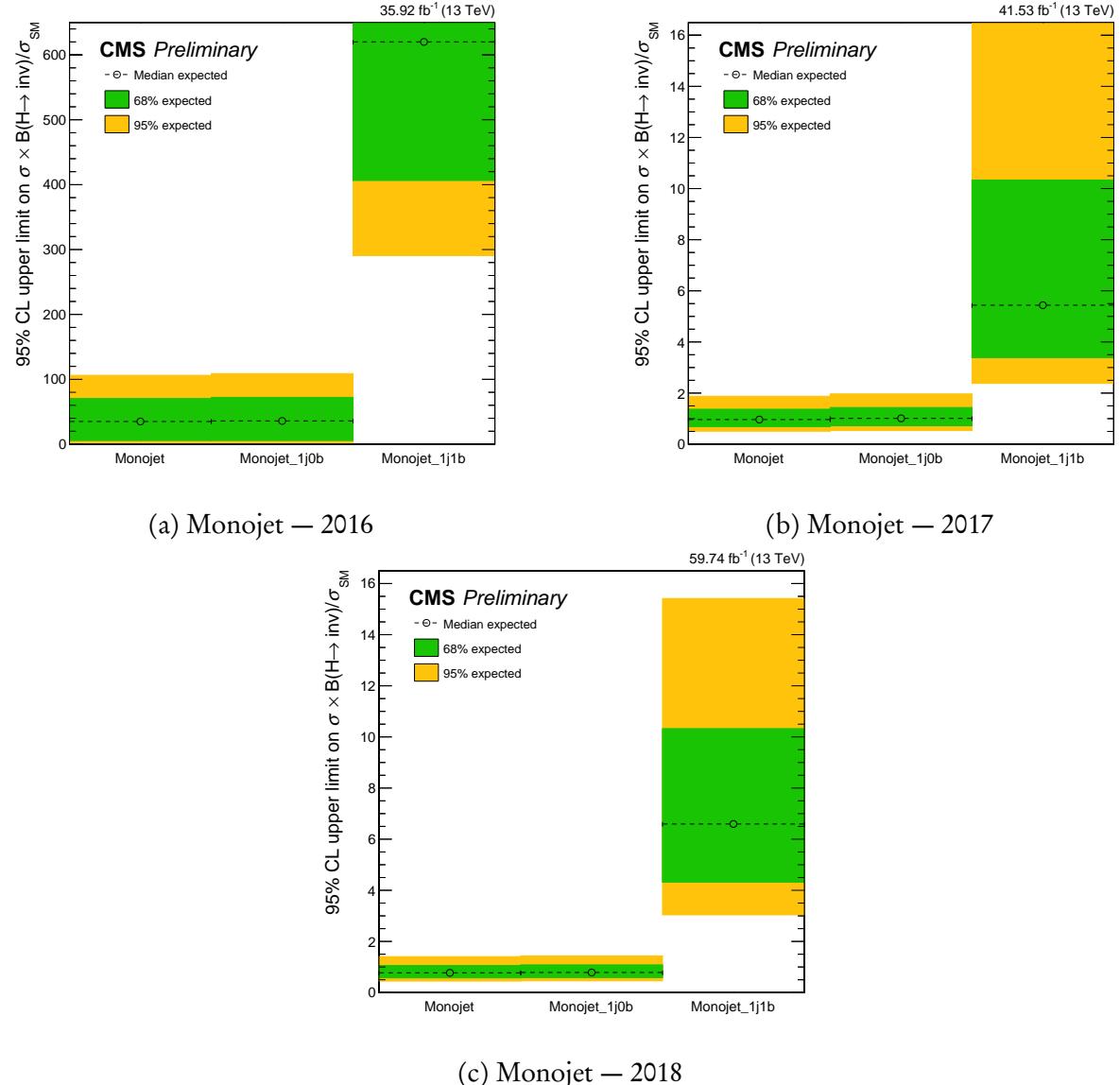


Figure 6.21: Expected 95 % CL upper limits on the Higgs boson to invisible state branching fraction in the monojet category, for both the individual subcategories, and the combination of them, for each data-taking year in Run-2.

## 6.13. COMBINED RESULTS

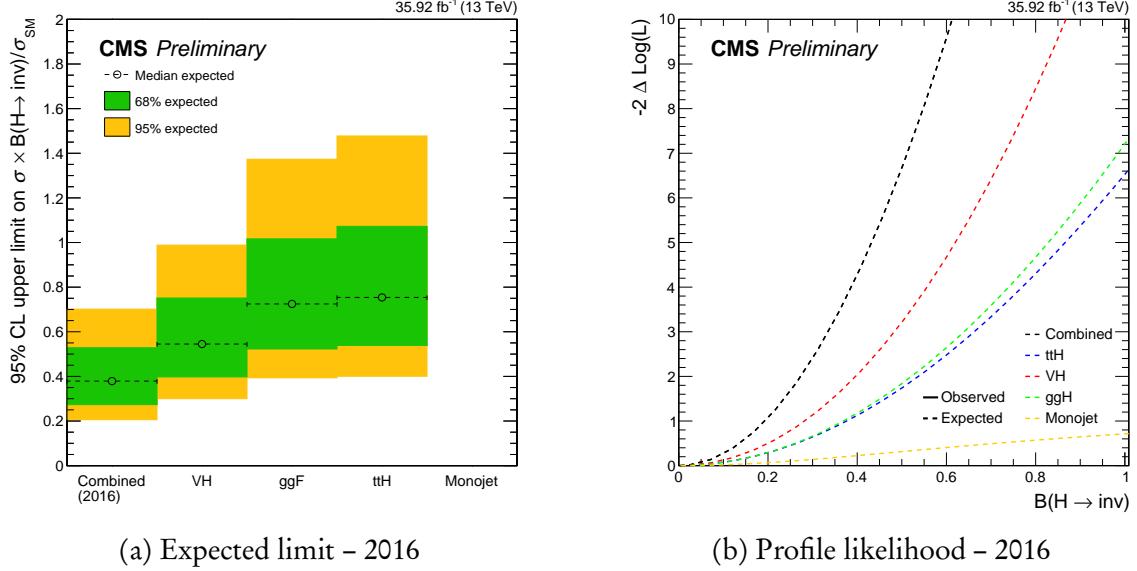


Figure 6.22: Expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction  $\mathcal{B}(H \rightarrow \text{inv.})$  (left) and the corresponding profile likelihood ratio as a function of it (right), for both the individual categories that target a specific production mode, as well as the combination of them, for the 2016 dataset. The standard model Higgs boson with its associated mass and production cross section are assumed.

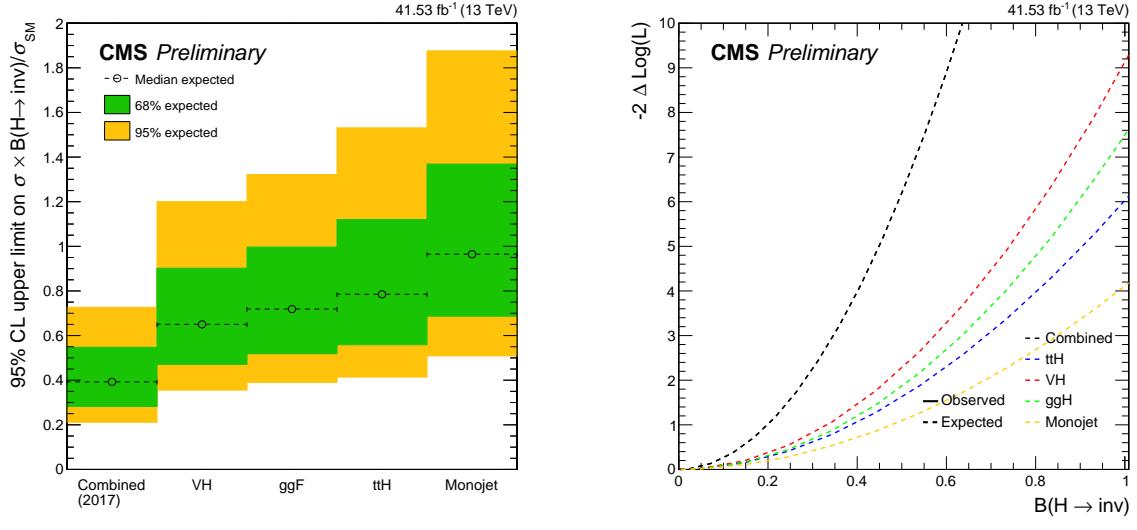


Figure 6.23: Expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction  $\mathcal{B}(H \rightarrow \text{inv.})$  (left) and the corresponding profile likelihood ratio as a function of it (right), for both the individual categories that target a specific production mode, as well as the combination of them, for the 2017 dataset. The standard model Higgs boson with its associated mass and production cross section are assumed.

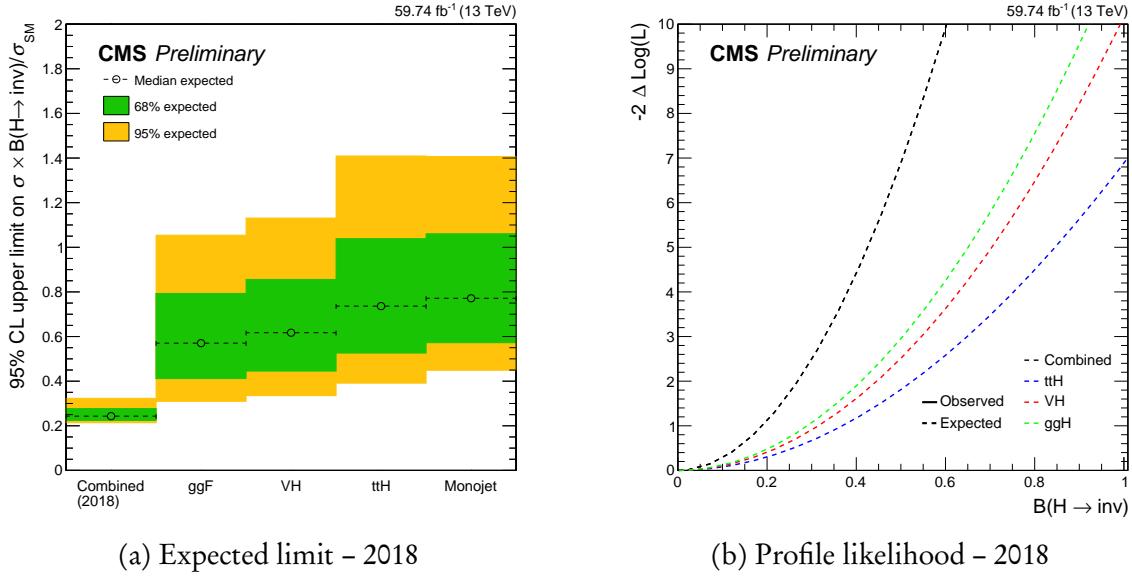


Figure 6.24: Expected 95 % CL upper limit on the Higgs boson to invisible state branching fraction  $\mathcal{B}(H \rightarrow \text{inv.})$  (left) and the corresponding profile likelihood ratio as a function of it (right), for both the individual categories that target a specific production mode, as well as the combination of them, for the 2018 dataset. The standard model Higgs boson with its associated mass and production cross section are assumed.

## 6.14 Interpretations in simplified dark matter scenarios

The results of the analysis fail to accurately constrain the branching ratio on the Higgs boson to invisible state decays, with a value XXX times that of the standard model and no observation of the process. However, limits on certain properties of dark matter may be set. Two BSM interpretations are presented: a Higgs portal model where the SM acts as the mediator, and the existence of an additional, invisibly decaying, SM-like Higgs boson.

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

One interpretation of our results 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. An effective field theory approach assumes that the invisible decays of the Higgs boson results in the pair production of dark matter particles  $\chi$ , with a frequency constrained by the upper limit on  $\mathcal{B}(H \rightarrow \text{inv.})$ . If  $m_\chi < m_H/2$ , on-shell production allows the translation of the  $H \rightarrow \text{inv.}$  width  $\Gamma_{\text{inv.}}$  into the spin-independent  $\chi$ -nucleon scattering cross section  $\sigma_{\text{SI}}$ , as in Ref. 73. 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 dark matter mass  $m_\chi$ . Comparisons can therefore be made between direct detection and collider experiments.

In this Higgs portal model, two cases are considered: the dark matter candidate is a scalar, or a Majorana fermion. The branching ratio is  $\mathcal{B}(H \rightarrow \text{inv.}) = \Gamma_{\text{inv.}} / (\Gamma_{\text{SM}} + \Gamma_{\text{inv.}})$ , where we assume  $\Gamma_{\text{SM}} = 4.07 \text{ MeV}$  [92]. From Ref. 73,  $\Gamma_{\text{inv.}}$  can be calculated for scalar  $\chi_S$  and fermion dark matter  $\chi_f$  as

$$(6.16) \quad \begin{aligned} \Gamma_{H \rightarrow \chi_S \chi_S}^{\text{inv.}} &= \frac{\lambda_{HSS}^2 v^2 \beta_S}{64\pi m_H}, \\ \Gamma_{H \rightarrow \chi_f \chi_f}^{\text{inv.}} &= \frac{\lambda_{Hff}^2 v^2 m_H \beta_f^3}{32\pi \Lambda^2} \end{aligned}$$

where  $\lambda$  is the coupling (scaled by  $\Lambda$  in the case of fermions)<sup>30</sup>,  $v$  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  $m_{q\text{dark}}$  are the physical masses after electroweak symmetry breaking with this new Higgs field. The spin-independent  $\chi$ -nucleon scattering cross sections are

$$(6.17) \quad \begin{aligned} \sigma_{\chi_S - N}^{\text{SI}} &= \frac{\lambda_{HSS}^2}{16\pi m_H^4} \frac{m_N^4 f_N^2}{(m_{\chi_S} + m_N)^2}, \\ \sigma_{\chi_f - N}^{\text{SI}} &= \frac{\lambda_{Hff}^2}{4\pi \Lambda^2 m_H^4} \frac{m_N^4 m_{\chi_f}^2 f_N^2}{(m_{\chi_f} + m_N)^2} \end{aligned}$$

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. 73. Limits on these cross sections as a function of dark matter mass are displayed in Fig. 6.25,<sup>31</sup> computed at 90% confidence level to compare with direct detection experiments, whose latest results are also presented.

For  $\mathcal{B}(H \rightarrow \text{inv.}) < XX$ , the Higgs portal models assuming scalar and fermion dark matter candidates set the lowest limits on  $\sigma_{\text{SI}}$  for  $m_\chi$  below Y and Z GeV, respectively.

## 6.14.2 Constraints on a standard model-like Higgs boson

If the SM Higgs boson does not decay into invisible particles beside the neutrino, one may still hypothesize a scenario in which the Higgs field is associated with dark matter. The next-to-minimal branch of supersymmetry and models with two Higgs doublets are examples

<sup>30</sup>Is  $\Lambda$  some sort of scale, similar to  $\Lambda_{\text{QCD}}$ ?

<sup>31</sup>The plot is a placeholder for now, since we don't have final non-VBF/combined limits.

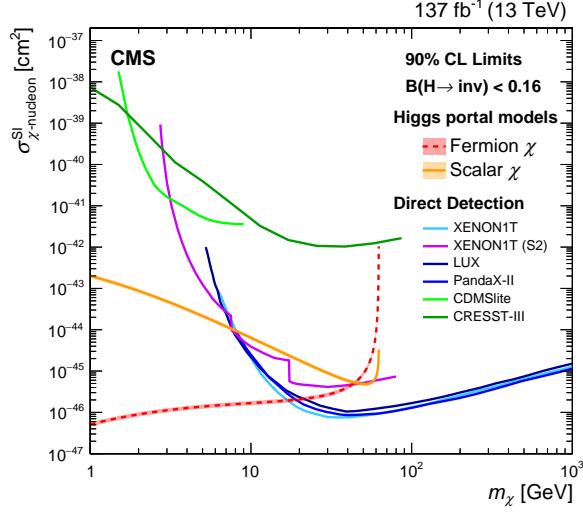


Figure 6.25: 90 % confidence level upper limits on the spin-independent dark matter-nucleon scattering cross section in Higgs portal models, where the standard model Higgs boson decays into a pair of scalar (solid orange) or fermion (dashed red) dark matter particles. Comparisons to direct detection experiments are also provided: XENON1T [20] (additionally with the S2-only analysis [21]), LUX [12], PandaX-II [69], CDMSlite [9], and CRESST-III [6].

that accommodate an additional Higgs boson with a different mass that does not couple to the SM variant. It is denoted as “standard model-like,” since it may be produced via the same mechanisms as the SM Higgs boson ( $t\bar{t}H$ ,  $VH$ ,  $ggH$ , and VBF) with the same relative contributions from each one. This extra Higgs is assumed to decay exclusively to invisible states.

We perform our analysis where we assume an additional Higgs boson, and show the limit on the branching ratio as a function of mass.

## CONCLUSIONS

Satisfaction lies in the effort, not in the attainment.

— Mahatma Gandhi

This is the conclusion.

Include a summary of thesis and work done over the course of my PhD with emphasis on the most important results/contributions.

- Mention the direction the semi-visible jet and Higgs to invisible analyses can take (sharing ideas/strategies I have, potential improvements with more LHC data and future prospects from potential future experiments).
  - SVJ has the  $t$ -channel and boosted searches underway, so those could yield interesting results. Long-lived searches are also an option. Higher energy collider like FCC gives larger range of phase space to explore.
  - For Higgs to invisible, precision collider like FCC-ee or ILC might be required to actually observe the process (assuming its  $\mathcal{B}$  is around SM value). Otherwise, we just need a lot of LHC data to narrow down the  $\mathcal{B}$  if it is enhanced by some BSM process.
  - Both searches can be improved with better ML techniques and stuff for categorisation/background rejection.





## APPENDIX A

B egins an appendix. Currently just a placeholder for any appendices that I plan to include.



## 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., *Combination of searches for invisible Higgs boson decays with the ATLAS experiment*, Phys. Rev. Lett. 122.23 (2019), p. 231801, doi: [10.1103/PhysRevLett.122.231801](https://doi.org/10.1103/PhysRevLett.122.231801), arXiv: [1904.05105 \[hep-ex\]](https://arxiv.org/abs/1904.05105).
- [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] J. ALLISON et al., *Recent developments in Geant4*, Nucl. Instrum. Methods Phys. Res. A 835 (2016), pp. 186–225, doi: [10.1016/j.nima.2016.06.125](https://doi.org/10.1016/j.nima.2016.06.125).
- [14] J. ALLISON et al., *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. 53.1 (Feb. 2006), pp. 270–278, doi: [10.1109/TNS.2006.869826](https://doi.org/10.1109/TNS.2006.869826).
- [15] 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).
- [16] 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).
- [17] 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).
- [18] 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).
- [19] 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).
- [20] 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).

- [21] 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).
- [22] 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).
- [23] G. ARNISON et al., *Experimental observation of isolated large transverse energy electrons with associated missing energy at  $\sqrt{s} = 540 \text{ GeV}$* , Phys. Lett. B 122.1 (1983), pp. 103–116, doi: [10.1016/0370-2693\(83\)91177-2](https://doi.org/10.1016/0370-2693(83)91177-2).
- [24] 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).
- [25] 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).
- [26] 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).
- [27] 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).
- [28] 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).
- [29] I. BALDES and K. PETRAKI, *Asymmetric thermal-relic dark matter: Sommerfeld-enhanced freeze-out, annihilation signals and unitarity bounds* (2017), arXiv: [1703.00478 \[hep-ph\]](https://arxiv.org/abs/1703.00478).
- [30] M. BANNER et al., *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider*, Phys. Lett. B 122.5 (1983), pp. 476–485, doi: [10.1016/0370-2693\(83\)91605-2](https://doi.org/10.1016/0370-2693(83)91605-2).
- [31] 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 (Aug. 2006), pp. 1055–1056, doi: [10.1111/j.1365-2966.2006.10688.x](https://doi.org/10.1111/j.1365-2966.2006.10688.x).

## BIBLIOGRAPHY

---

- [32] 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 (Dec. 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).
- [33] G. L. BAYATYAN et al., *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*, Technical Design Report CMS.
- [34] 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).
- [35] 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 (Oct. 2015), 101002. 19 p, doi: [10.1103/PhysRevSTAB.18.101002](https://doi.org/10.1103/PhysRevSTAB.18.101002).
- [36] P. BEVINGTON and D. ROBINSON, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill Higher Education, McGraw-Hill Education, 2003, ISBN: 9780072472271.
- [37] 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).
- [38] P. BLOCH et al., *Changes to CMS ECAL electronics: addendum to the Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 2002.
- [39] 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).
- [40] A. H. BROEILS, *The mass distribution of the dwarf spiral NGC 1560*, Astron. Astrophys. 256 (Mar. 1992), pp. 19–32.
- [41] 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).
- [42] 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).

- [43] 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).
- [44] 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).
- [45] 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).
- [46] 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).
- [47] 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).
- [48] S. CHATRCHYAN et al., *Search for invisible decays of Higgs bosons in the vector boson fusion and associated ZH production modes*, Eur. Phys. J. C 74 (2014), p. 2980, DOI: [10.1140/epjc/s10052-014-2980-6](https://doi.org/10.1140/epjc/s10052-014-2980-6), arXiv: [1404.1344 \[hep-ex\]](https://arxiv.org/abs/1404.1344).
- [49] 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.
- [50] 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.
- [51] CMS COLLABORATION, *CMS Luminosity Public Results*, URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults> (visited on 22/02/2020).
- [52] 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).
- [53] 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).
- [54] 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).

## BIBLIOGRAPHY

---

- [55] 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).
- [56] 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).
- [57] 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).
- [58] 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.
- [59] CMS COLLABORATION, *Reweighting recipe to emulate Level 1 ECAL prefireing*, access may be restricted, URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1ECALPrefiringWeightRecipe> (visited on 09/07/2020).
- [60] CMS COLLABORATION, *The CMS electromagnetic calorimeter project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.
- [61] CMS COLLABORATION, *The CMS hadron calorimeter project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.
- [62] CMS COLLABORATION, *The CMS magnet project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.
- [63] CMS COLLABORATION, *The CMS tracker: addendum to the Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 2000.
- [64] 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).
- [65] 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).
- [66] 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).
- [67] 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).

- [68] B. COX and J. FORSHAW, *Universal: A Journey Through the Cosmos*, Penguin Books Limited, 2016, ISBN: 9780141968346.
- [69] 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).
- [70] 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.
- [71] 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).
- [72] 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).
- [73] 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).
- [74] 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).
- [75] J. EINASTO, *Dark Matter*, ArXiv e-prints (Jan. 2009), arXiv: [0901.0632 \[astro-ph.CO\]](https://arxiv.org/abs/0901.0632).
- [76] 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.
- [77] 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](https://doi.org/10.1103/PhysRevLett.13.321).
- [78] V. FANTI et al., *A new measurement of direct CP violation in two pion decays of the neutral kaon*, Phys. Lett. B 465.1-4 (Oct. 1999), pp. 335–348, doi: [10.1016/s0370-2693\(99\)01030-8](https://doi.org/10.1016/s0370-2693(99)01030-8).
- [79] 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](https://doi.org/10.1103/PhysRevD.90.015014), arXiv: [1305.6068 \[hep-ph\]](https://arxiv.org/abs/1305.6068).
- [80] B. D. FIELDS et al., *Big-Bang Nucleosynthesis After Planck*, J. Cosmol. Astropart. Phys. 2003.03 (2020), p. 010, doi: [10.1088/1475-7516/2020/03/010](https://doi.org/10.1088/1475-7516/2020/03/010), arXiv: [1912.01132 \[astro-ph.CO\]](https://arxiv.org/abs/1912.01132).

## BIBLIOGRAPHY

---

- [81] C. FINANCE and A. P. DEPARTMENT, *2020 Annual Contributions to CERN budget*, URL: <https://fap-dep.web.cern.ch/rpc/2020-annual-contributions-cern-budget> (visited on 30/03/2020).
- [82] *First constraints on invisible Higgs boson decays using t̄tH production at  $\sqrt{s} = 13 \text{ TeV}$* , tech. rep. CMS-PAS-HIG-18-008, CERN, Geneva, 2019.
- [83] K. C. FREEMAN, *On the Disks of Spiral and S0 Galaxies*, *Astrophys. J.* 160 (June 1970), p. 811, DOI: [10.1086/150474](https://doi.org/10.1086/150474).
- [84] 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](https://doi.org/10.1088/1126-6708/2007/11/070), arXiv: [0709.2092 \[hep-ph\]](https://arxiv.org/abs/0709.2092).
- [85] P. GONDOLO and G. GELMINI, *Cosmic Abundances of Stable Particles: improved analysis*, *Nucl. Phys. B* 360 (Aug. 1991), pp. 145–179, DOI: [10.1016/0550-3213\(91\)90438-4](https://doi.org/10.1016/0550-3213(91)90438-4).
- [86] 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](https://doi.org/10.1103/PhysRevLett.13.585).
- [87] 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](https://doi.org/10.1007/JHEP03(2010)080), arXiv: [0911.1120 \[hep-ph\]](https://arxiv.org/abs/0911.1120).
- [88] 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](https://doi.org/10.1007/JHEP11(2013)008), arXiv: [1303.3040 \[hep-ph\]](https://arxiv.org/abs/1303.3040).
- [89] T. HAN, J. D. LYKKEN and R.-J. ZHANG, *On Kaluza-Klein states from large extra dimensions*, *Phys. Rev. D* 59 (1999), p. 105006, DOI: [10.1103/PhysRevD.59.105006](https://doi.org/10.1103/PhysRevD.59.105006), arXiv: [hep-ph/9811350 \[hep-ph\]](https://arxiv.org/abs/hep-ph/9811350).
- [90] F. HASERT et al., *Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment*, *Phys. Lett. B* 46.1 (1973), pp. 138–140, DOI: [10.1016/0370-2693\(73\)90499-1](https://doi.org/10.1016/0370-2693(73)90499-1).
- [91] F. HASERT et al., *Search for elastic muon-neutrino electron scattering*, *Phys. Lett. B* 46.1 (1973), pp. 121–124, DOI: [10.1016/0370-2693\(73\)90494-2](https://doi.org/10.1016/0370-2693(73)90494-2).
- [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 (Sept. 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] 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).
- [99] 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).
- [100] 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).
- [101] V. KARIMÄKI et al., *The CMS tracker system project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.
- [102] 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).
- [103] V. KHACHATRYAN et al., *Search for dark matter in proton-proton collisions at 8 TeV with missing transverse momentum and vector boson tagged jets*, JHEP 12 (2016), [Erratum: JHEP 08, 035 (2017)], p. 083, doi: [10.1007/JHEP12\(2016\)083](https://doi.org/10.1007/JHEP12(2016)083), arXiv: [1607.05764 \[hep-ex\]](https://arxiv.org/abs/1607.05764).
- [104] 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).

## BIBLIOGRAPHY

---

- [105] 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).
- [106] V. KHACHATRYAN et al., *Searches for invisible decays of the Higgs boson in  $pp$  collisions at  $\sqrt{s} = 7, 8,$  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).
- [107] 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).
- [108] G. KRNJACIĆ, *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).
- [109] 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).
- [110] J. G. LAYTER, *The CMS muon project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.
- [111] 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).
- [112] *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.
- [113] 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).
- [114] S. MERTENS, *Direct Neutrino Mass Experiments*, J. Phys. Conf. Ser. 718.2 (2016), p. 022013, doi: [10.1088/1742-6596/718/2/022013](https://doi.org/10.1088/1742-6596/718/2/022013), arXiv: [1605.01579 \[nucl-ex\]](https://arxiv.org/abs/1605.01579).
- [115] J. MOUSA et al., *Overview of the CMS Detector Performance at LHC Run 2*, Universe 5 (Jan. 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] 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).
- [118] G. RAUCO, *Search for New Physics in All-Hadronic FinalStates with the CMS Experiment*, PhD thesis, Universität Zürich, 2019.
- [119] T. SAKUMA, *Cutaway diagrams of CMS detector* (May 2019).
- [120] 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).
- [121] S. SCHRAMM, *Searching for Dark Matter with the ATLAS Detector*, Springer Theses, Springer International Publishing, 2016, ISBN: 9783319444536.
- [122] 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).
- [123] 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).
- [124] 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).
- [125] 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).
- [126] 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).

## BIBLIOGRAPHY

---

- [127] 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  $\sqrt{s} = 13$  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).
- [128] 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).
- [129] A. SIRUNYAN et al., *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13$  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).
- [130] A. SIRUNYAN et al., *Search for new physics in events with a leptonically decaying Z boson and a large transverse momentum imbalance in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, Eur. Phys. J. C 78.4 (2018), p. 291, doi: [10.1140/epjc/s10052-018-5740-1](https://doi.org/10.1140/epjc/s10052-018-5740-1), arXiv: [1711.00431 \[hep-ex\]](https://arxiv.org/abs/1711.00431).
- [131] A. M. SIRUNYAN et al., *Measurements of  $t\bar{t}$  differential cross sections in proton-proton collisions at  $\sqrt{s} = 13$  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).
- [132] A. M. SIRUNYAN et al., *Search for dijet resonances in proton-proton collisions at  $\sqrt{s} = 13$  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).
- [133] 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).
- [134] 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).
- [135] 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).

- [136] A. SOPCZAK, *Precision Measurements in the Higgs Sector at ATLAS and CMS*, PoS Proc. Sci. FFK2019 (Jan. 2020), p. 006, doi: [10.22323/1.353.0006](https://doi.org/10.22323/1.353.0006), arXiv: [2001.05927 \[hep-ex\]](https://arxiv.org/abs/2001.05927).
- [137] 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).
- [138] M. TANABASHI et al., *Review of Particle Physics*, Phys. Rev. D 98 (3 Aug. 2018), p. 030001, doi: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001).
- [139] L. G. TRENTADUE et al., *Neutrino counting*, CERN-TH-5528-89 (Sept. 1989), 42 p, doi: [10.5170/CERN-1989-008-V-1.129](https://doi.org/10.5170/CERN-1989-008-V-1.129).
- [140] 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).
- [141] S. WEINBERG, *Cosmology*, Cosmology, OUP Oxford, 2008, ISBN: 9780191523601.
- [142] 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).
- [143] 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).
- [144] A. ZABI et al., *The CMS Level-1 Calorimeter Trigger for the LHC Run II*, J. Instrum. 12.01 (Jan. 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 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  $E_{\mathrm{T}}^{\mathrm{miss}}$ , and also referred to in literature as “missing transverse momentum” ( $p_{\mathrm{T}}^{\mathrm{miss}}$ ).

## 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

**eV** electron volt.

**cmssw** CMS SoftWare.

**AEGIS** Antihydrogen Experiment: Gravity, Interferometry, Spectroscopy.

**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.

## ACRONYMS

---

**HF** hadron forward calorimeter.

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

**HLT** High-Level Trigger.

**ISOLDE** Isotope Separator On Line DEvice.

**JEC** jet energy corrections.

**JER** jet energy resolution.

**JES** jet energy scale.

**L1** Level-1.

**L1T** Level-1 Trigger.

**LEP** Large Electron-Positron Collider.

**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.

**MOND** modified Newtonian dynamics.

**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.

