

FROM TRIGGER TO DATA ANALYSIS: LOOKING FOR NEW PHYSICS AT THE
LHC USING DEEP LEARNING TECHNIQUES

By

Maria Mazza

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Physics — Doctor of Philosophy
Computational Mathematics, Science and Engineering — Dual Major

2024

ABSTRACT

The Standard Model (SM), crowned in 2012 with the discovery of the Higgs boson, exhibits remarkable predictive power. However, several phenomena remain unexplained and evidence for physics beyond the SM continues to emerge. The Higgs boson appears at the center of many of these pressing issues, making its study one of the top priorities at the Large Hadron Collider (LHC). To extend its discovery potential, the LHC will undergo a major upgrade that will bring a ten-fold increase in integrated luminosity and increase the center-of-mass energy to 14 TeV. Extracting relevant physics in these unprecedented extreme conditions will require an upgrade of the detector and trigger system, as well as innovative analysis techniques to enhance signal-to-background discrimination. The research presented in this work followed these new directions and challenges on two parallel fronts, with the shared goal of improving our understanding of the scalar sector, and with a common focus on the development of new machine learning methods.

On one front, this work contributed to a search for new heavy resonances decaying to two SM bosons (using the full Run 2 ATLAS dataset). Models that predict such particles are often interpreted in the context of two general frameworks – the Heavy Vector Triplet and the two-Higgs-doublet models – and address important open questions related to the Higgs sector: the naturalness problem and the possibility of an extended scalar sector. In particular, this work presents the development of a new multi-class deep neural network (DNN) jet tagger strategy to compete with traditional analysis techniques. The development of the tagger as a standalone tool, as well as the deployment within the analysis workflow to improve analysis sensitivity are presented.

On the other front, this work made several contributions to the High-Luminosity LHC upgrade of the ATLAS hardware-based trigger. These started from the development of the software simulation framework for trigger performance studies, and proceeded to focus on the development of new jet triggers, targeting in particular $HH \rightarrow b\bar{b}b\bar{b}$, an important signature for the measurement of the Higgs self-coupling. This work presents the development, benchmarking, and preliminary firmware simulation of a new jet reconstruction and triggering strategy, as well as the development and performance of a new DNN for pileup mitigation, with both algorithms designed for deployment on fast FPGA hardware.

To Kévin.

E quindi uscimmo a riveder le stelle.
Inferno, XXXIV

ACKNOWLEDGMENTS

I would like to begin by expressing my deepest gratitude to my Ph.D. advisor, Wade Fisher. Your invaluable teachings and your patience during the thorough discussions we shared have profoundly shaped my academic journey. I am also grateful for, and have sincerely valued, the trust you placed in me: thank you for providing me with opportunities to work on inspiring projects, while giving me the freedom to explore and define my own research path; for allowing me to go from student to teacher, by giving me the opportunity to prepare lectures for your machine learning class; and for trusting me with the freedom I needed to make important life choices. Your support really made a difference and I will always be deeply grateful.

I would like to sincerely thank my committee members Wolfgang Kerzendorf, Dean Lee, Saiprasad Ravishankar, and Andreas von Manteuffel, as well as Matt Hirn who left MSU before I graduated, for their support and guidance during these past few years. I would also like to extend a special thank you to Andreas and C.-P. Yuan for having been excellent professors of theoretical high energy physics. Finally, I must thank my committee member and ATLAS collaborator Daniel Hayden, for having always been available and supportive, going well beyond what was expected from him.

I would like to thank the MSU postdocs that I was lucky enough to work with. In particular, Garabed Halladjian, my first mentor in particle physics and now a dear friend, and Hector de la Torre, without whom all my work on the trigger would have not been possible. Thank you both for all your teachings, support, and friendship. I would also like to thank Garrit Reynolds, the undergraduate student I mentored, for the hard work.

I also want to thank my ATLAS collaborators, especially those I had the chance to work with on the analysis and trigger fronts. My work in the analysis would have not been possible without the contributions of the other analyzers and, in particular, of the analysis contact and MSU postdoc Robert Les. I would also like to thank the US ATLAS Award Committee for rewarding my thesis work with the US ATLAS Outstanding Graduate Student Award. A very nice and unexpected surprise at the end of this long journey, that I would have never thought possible when I started.

Vorrei ringraziare la mia famiglia per avermi sempre supportato nelle mie decisioni che mi hanno permesso di arrivare fino a qui. Mia madre, per i sacrifici e l'amore incondizionato, e i miei fratelli Antonio, Giulia, Elisabetta e Lorenzo, per essere sempre presenti e per darmi sempre la certezza di non essere mai sola, anche se lontana. Vorrei mandare un pensiero anche a mia nonna Marisa Freni, per essere stata un modello di donna e scienziata, e vorrei

ringraziare anche gli zii, cugini, e tutti i parenti per darmi il supporto che solo una grande famiglia sa dare.

J'aimerais aussi remercier Niels et Yulia pour m'avoir accueillie dans la famille et pour m'avoir apporté de l'innocence dans la vie.

Mon Kévin, thank you for going beyond yourself to help me these past few years and for never letting me down (even when I ask you to proofread two hundred pages of experimental physics). I will forever be grateful to you for having believed in me and for having given me a chance to be the person I was supposed to be. You inspire me to better myself every day. This thesis is dedicated to you and to the beginning of a new chapter of our lives, looking at the future with hope.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS	ix
Chapter 1 Introduction	1
Chapter 2 The Standard Model (SM)	5
2.1 A quantum theory of fields	9
2.2 The Lagrangian formulation	11
2.3 Symmetries and conservation laws	12
2.4 Deriving a gauge theory	13
2.5 The Higgs Mechanism	16
2.6 The Standard Model Lagrangian	20
2.7 The Higgs sector	24
2.8 Hints for physics beyond the Standard Model	30
Chapter 3 The Higgs boson as a portal to new physics	33
3.1 The Higgs self-coupling	34
3.2 Naturalness	35
3.3 The Heavy Vector Triplet model	36
3.4 The Two-Higgs-Doublet Model	39
Chapter 4 The LHC and the ATLAS experiment	43
4.1 The Large Hadron Collider (LHC)	43
4.1.1 Overview	43
4.1.2 The accelerator complex	44
4.1.3 LHC performance and operation	46
4.1.4 Brief timeline of LHC operation and upgrades	50
4.2 The ATLAS detector	51
4.2.1 The ATLAS coordinate system	52
4.2.2 The magnet system	53
4.2.3 The inner detector	54
4.2.4 The calorimeters	56
4.2.5 The muon spectrometer	63
4.2.6 The forward detectors	65
4.3 The ATLAS Trigger	65
4.3.1 The Level-1 trigger	66
4.3.2 The High-Level Trigger	69
4.3.3 Trigger operations	70
4.3.4 The Phase I trigger upgrade	71
4.3.5 The Phase II trigger upgrade	74
4.4 ATLAS Event reconstruction	78
4.4.1 Tracks and vertices	79
4.4.2 Electrons	79
4.4.3 Muons	80
4.4.4 Topological clustering	81
4.4.5 Missing transverse momentum	83

4.4.6	<i>b</i> -tagging	84
Chapter 5	Hadron collider physics	86
5.1	From QCD to jets	87
5.1.1	The strong coupling	87
5.1.2	The hard-scatter cross section	88
5.1.3	Showering and hadronization	91
5.1.4	Soft physics	91
5.1.5	Monte Carlo event generators	92
5.1.6	Jets	94
5.2	Jet reconstruction algorithms	95
5.2.1	Infrared-collinear safety	96
5.2.2	Cone algorithms	97
5.2.3	Sequential-recombination algorithms	98
5.3	Jets in ATLAS	100
5.3.1	Jet algorithm	100
5.3.2	Jet inputs and jet collections	101
5.4	Boosted jet tagging	103
5.5	Pileup suppression	106
5.5.1	Area-median subtraction	107
5.5.2	Grooming	108
5.5.3	Constituent-level	108
Chapter 6	Concepts of statistics and machine learning	111
6.1	Statistical inference	111
6.2	Neural networks	113
6.3	Hypothesis testing with profile likelihood ratio	115
Chapter 7	Search for new heavy resonances decaying to two SM bosons in semi-leptonic final states	120
7.1	The search for new heavy resonances	120
7.2	Analysis overview	122
7.2.1	Analysis strategy	125
7.2.2	Machine learning approach	126
7.2.3	The Multi-Class Tagger	127
7.3	Signal and background processes	129
7.4	Data taking and trigger selection	134
7.5	Object selection	135
7.6	Event selection	138
7.6.1	Jet requirements	139
7.6.2	0-lepton channel	140
7.6.3	1-lepton channel	141
7.6.4	2-lepton channel	142
7.7	Event categorization	143
7.8	Boosted jets Multi-Class Tagger (MCT)	146
7.8.1	Training	146

7.8.2	Testing performance	152
7.9	Resolved jets MCT	162
7.9.1	Training	162
7.9.2	Testing performance	164
7.10	MCT deployment in the analysis	176
7.10.1	Motivation	176
7.10.2	Studies overview	179
7.10.3	MCT strategy	180
7.10.4	Signal efficiency	181
7.10.5	Signal significance	182
7.10.6	Expected limit sensitivity	184
7.11	MCT Modeling	188
7.11.1	Derivation of background normalization scale factors	188
7.11.2	Modeling in pre-selection regions	189
7.11.3	Modeling in top-enriched control region	195
7.11.4	Sensitivity to systematic variations of MCT scores	196
Chapter 8 Firmware algorithm development for the HL-LHC Global Trigger upgrade		199
8.1	The Global Trigger (GT)	200
8.2	Trigger performance studies	203
8.2.1	The GT software simulation framework	205
8.2.2	Developing a jet trigger	207
8.3	A cone jet reconstruction algorithm	210
8.3.1	Development	212
8.3.2	Physics performance	219
8.3.3	Constituent-level pileup suppression	223
8.3.4	Preliminary firmware simulation	227
8.4	Pileup-jet rejection with neural networks	229
8.4.1	Pileup jet identification	230
8.4.2	Neural network development	234
8.4.3	Training and performance evaluation	236
8.4.4	Trigger performance	238
Chapter 9 Conclusion and outlook		245
BIBLIOGRAPHY		248
APPENDIX A. Analysis		261
APPENDIX B. Trigger		269

LIST OF ABBREVIATIONS

- 2HDM** Two-Higgs-Doublet Model. 1, 39–41, 120, 122, 123, 130, 132
- BC** Bunch Crossing. 46, 47, 200, 201, 203
- BR** Branching Ratio. 121, 122, 129, 131, 139
- BSM** Beyond Standard Model. 1, 8, 9, 31, 32, 34, 36, 83, 87, 103, 146
- CERN** European Organization for Nuclear Research. 8, 43, 51, 52
- CKM** Cabbibo-Kobayashi-Maskawa. 29, 33
- CoM** Center-of-Mass. 43, 44, 50, 51, 86, 89, 92
- CR** Control Region. 125, 126, 128, 145
- CTP** Central Trigger Processor. 66–69, 77, 201, 202
- DNN** Deep Neural Network. 2, 146, 147, 150, 163, 245, 247
- DY** Drell-Yan. 127, 129, 130, 143
- ECF** Energy Correlation Function. 105, 148
- EF** Event Filter. 66, 75, 77
- EL** Euler-Lagrange. 11, 12
- EMCal** Electromagnetic Calorimeter. 79
- EMEC** Electromagnetic End-cap Calorimeter. 59, 60, 62
- EOR** Energy Overlap Removal. 217–219, 238
- EWSB** Electroweak Symmetry Breaking. 34, 37
- FCal** Forward Calorimeter. 59, 61, 62, 68, 82
- FPGA** Field Programmable Gate Array. 3, 69, 72, 73, 77, 199–204, 210, 212, 213, 227–229
- GEP** Global Event Processor. 200–204
- ggF** Gluon-Gluon Fusion. 127, 129, 131, 143, 206
- GT** Global Trigger. viii, 3, 199–208, 210–212, 215, 223, 227–232, 235
- GWS** Glashow-Weinberg-Salam. 8

HCal Hadronic Calorimeter. 58, 60, 68

HL-LHC High-Luminosity Large Hadron Collider. 1, 2, 28, 33, 34, 51, 54, 63, 71, 74, 106, 109, 199, 200, 205, 206, 221, 229, 230, 246, 247

HLS High-Level Synthesis. 203, 227, 228

HLT High-Level Trigger. 66, 67, 69–71, 76

HVT Heavy Vector Triplet. 33, 36, 38, 120, 122, 129, 131, 132, 143, 145, 162, 176, 177, 179, 182, 184

IBL Insertable B-Layer. 54, 56

ID Inner Detector. 53–56, 58, 59, 62, 68, 69, 74, 78–81, 84, 102

IP Interaction Point. 43, 44, 48, 49, 51–53, 62, 65

IRC Infrared-Collinear. 91, 96, 97, 99, 105, 107, 210

ITk Inner Tracker. 54, 74, 77

jFEX Jet Feature EXtractor. 72–74, 269

LAr Liquid Argon. 58–62, 68, 71, 72, 74, 134–136, 202, 203

MC Monte Carlo. 92–94, 100, 103, 125, 126, 131–133, 135, 162, 205–208, 221

MCT Multi-Class Tagger. 128, 145, 146, 152, 162, 163, 179–182, 184, 186, 188–191, 195–197, 245, 246

MDTs Monitored Drift Tubes. 64, 65, 74, 77

MET Missing Energy Transverse. 83, 134, 135, 140, 142

MS Muon Spectrometer. 74, 80, 81

MSSM Minimal Supersymmetric Standard Model. 39, 122, 123

MUCTPI Muon Central Trigger Processor Interface. 67, 69, 77, 200–202

MUX Multiplexer Processor. 200–202

NN Neural Network. 127, 148, 149

PDF Parton Distribution Function. 89, 90, 92

QCD Quantum Chromodynamics. 20, 86–90, 94, 96, 103, 104, 108, 128, 132, 140, 141, 149, 162

QED Quantum Electrodynamics. 13–15, 22, 23

RNN Recurrent Neural Network. 127, 143

ROC Receiver Operating Characteristic. 153, 154, 156–161, 167, 171–175

RoI Region of Interest. 66, 68, 69

RPC Resistive Plate Chamber. 65, 67, 74, 77

RS Randall-Sundrum. 121, 122, 129–132

SCT SemiConductor Tracker. 54, 55, 79

SF Scale Factor. 188, 191–195, 261–268

SK Soft-Killer. 109, 110, 137, 206, 223

SLR Super Logic Region. 203, 205, 228

SR Signal Region. 125, 128, 145, 179

SSB Spontaneous Symmetry Breaking. 16, 18, 19, 24–28, 38–40

TCC Track-CaloCluster. 102, 136

TDAQ Trigger & Data Acquisition. 65–67, 71, 74, 75, 209, 211, 246

TGC Thin-Gap Chamber. 65, 67, 74

TileCal Tile Calorimeter. 59, 60, 68

TOB Trigger Object. 68, 69, 72, 77, 201, 202, 215, 269

TRT Transition Radiation Tracker. 54–56

UE Underlying Event. 91, 92, 94, 106, 108

UFO Unified Flow Object. 102, 136

VBF Vector Boson Fusion. 38, 127, 129–132, 143

vev Vacuum expectation value. 24, 30, 34, 40, 41

VR Variable Radius. 102, 137, 141, 143

WP Working Point. 79, 81, 85

Chapter 1

Introduction

The Standard Model (SM) of particle physics has proven to be a remarkably successful description of nature. However, several phenomena remain unexplained and evidence for physics beyond the SM continues to emerge. The Higgs boson appears at the center of many of these pressing issues, making its study one of the top priorities at the Large Hadron Collider (LHC). To extend its discovery potential, the accelerator will soon undergo a major upgrade that will raise the center-of-mass energy to $\sqrt{s} = 14 \text{ TeV}$ and bring the instantaneous luminosity up to $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. At the end of the High-Luminosity LHC (HL-LHC), ATLAS will have ten times the amount of data collected so far. This unprecedented opportunity will open up new search channels, previously inaccessible cross-sections, and more precise tests of SM observables. At the same time, the higher luminosity will generate unprecedented levels of radiation and pileup. Extracting relevant physics in these extreme conditions will require a substantial upgrade of the detector and trigger system, as well as innovative techniques to enhance signal-to-background discrimination, both in offline analyses and on real-time event selection. The research presented in this thesis followed these new directions and challenges on two parallel fronts, with the shared goal of improving our understanding of the scalar sector, and with a common focus on the development of new machine learning methods.

The observation of a light scalar with a mass of 125 GeV agrees with SM predictions, but necessarily leads to the naturalness problem - the Higgs mass is unstable under radiative corrections, making its observed value the result of an unnatural fine-tuning [1, 2]. This can be prevented if one postulates the existence of new heavy particles with masses around the TeV scale that couple to the Higgs boson. Several beyond-the-SM (BSM) models predict such resonances and are tested experimentally via a general Heavy Vector Triplet model [3], which assumes a simplified phenomenological Lagrangian where only the relevant couplings and mass parameters are retained. New heavy resonances at a similar mass scale are also predicted by Two-Higgs-Doublet Models (2HDMs) [4], which assume the simplest extension of the scalar sector by predicting the existence of two $\text{SU}(2)$ complex doublets. Most of these models predict sizable couplings of the new particles to the SM Higgs and weak gauge

bosons, making such final states rich landscapes where to look for new physics.

This thesis contributed to the ATLAS search for such new heavy resonances in final states with two SM bosons (WW , WZ , ZZ , ZH , or WH) decaying semi-leptonically. Due to the large multiplicity of the different final states considered simultaneously, standard analysis strategies using cut-based event selections had to be rethought to avoid complex overlapping of selection criteria. One of the critical tasks in the event selection of this type of search is the correct identification of the hadronically decaying jets: signal-like events are identified by the presence of jets originating from a Higgs, W , or Z boson, while jets originating from $t\bar{t}$ and V +jets processes characterize the primary SM backgrounds. A significant part of this work was the development of a new multi-class jet tagging algorithm for improved identification of the hadronic decay. Because the search probes mass resonances from 220 GeV to 5 TeV, the analysis is sensitive to a wide range of transverse momenta, requiring different jet reconstruction strategies: the jets are resolved as two small radius jets at low energies, while they are identified as a single large radius jet in the boosted regime. Therefore, two different 5-class deep neural networks (DNN) were trained, one for each reconstruction strategy, and with the output of each model giving the probability of the decay to be originating from a Higgs boson, a W boson, a Z boson, a top quark, or light quarks and gluons produced via the strong interaction. The work presented here covered the development of the models as standalone tools, as well as their deployment within the analysis workflow. Within the context of the analysis, their discrimination power and modelling was assessed, and a new strategy for the event categorization was designed for improved analysis sensitivity.

Electroweak baryogenesis, which predicts the Higgs boson to have developed a vacuum expectation value via a first-order phase transition in the early universe, provides a possible solution to the puzzle of the observed baryon asymmetry [5]. The nature of the transition can be accessed by the yet unmeasured Higgs trilinear self-coupling, as models that predict a first-order phase transition predict large deviations from the SM prediction [6]. Measurement of the Higgs self-coupling would also be a direct test of electroweak symmetry breaking and of the shape of the Higgs potential, the latter in turn connected to questions regarding the stability of the universe. The production of two Higgs bosons can provide a direct probe of the Higgs self-coupling, making the measurement of di-Higgs (HH) production one of the major goals of the LHC programme. Because of the low production cross section, the ATLAS and CMS experiments have so far only been able to set limits [7, 8]. However, the HL-LHC is expected to reach the ultimate sensitivity [9], making di-Higgs one of the flagship signatures for the HL-LHC and one of the main drivers of the HL-LHC trigger upgrade. In fact, for reasons that will be explained later, to retain sensitivity to λ_{HHH} it is pivotal to

retain the low m_{HH} events, an extremely challenging task for the trigger: in this kinematic region the decay products of the Higgs bosons are at low p_T , where signal efficiency competes with pileup rejection and is critically dependent on trigger thresholds.

Successful data collection has to start with the first step of the trigger chain, which in ATLAS is the Level-0 hardware-based trigger. The Global Trigger (GT) will be a new addition at Level-0 that will allow to deploy complex algorithms on fast FPGA hardware and bring the event rate from 40 MHz down to 1 MHz [10]. The GT is primarily a firmware project, with many algorithms under study. The contributions of this work to the GT upgrade included the development of the software simulation framework for the study of new firmware algorithms, the development of a new jet reconstruction and triggering strategy to make use of the new trigger capabilities, and the exploration of new machine learning algorithms for pileup mitigation targeting di-Higgs production.

The jet reconstruction algorithm was optimized by considering the trade-off between reducing algorithm complexity, required to meet FPGA resources and latency limitations, and maintaining high performance to preserve the physics goals of the collaboration. The algorithm was benchmarked against target signal simulations. In particular, the channel $HH \rightarrow b\bar{b}b\bar{b}$, with four low p_T b -quarks in the final state, was the prime target in the development of multi-jet triggers.

As the trigger thresholds are driven by the rate of pileup jets, a new method was proposed to mitigate the negative effect of pileup on trigger efficiencies and further increase the acceptance of $HH \rightarrow b\bar{b}b\bar{b}$ events at small m_{HH} values. Pileup-like radiation is uncorrelated from the hard scatter, resulting in a more diffuse energy pattern in pileup jets than in signal jets. This is another problem of pattern recognition well suited for deep learning applications. This thesis presents the development of a new neural network to identify pileup-like jets starting from topological cluster information, and the studies of its effect on the trigger performance.

The content of this thesis is structured as follows. The first part of the manuscript lays the relevant background information for this work. Chap. 2 reviews the SM of particle physics, with an emphasis on the Higgs sector. After highlighting the motivation for beyond the SM physics, Chap. 3 discusses how the scalar sector could be a portal to new physics, with a focus on the aspects relevant for this thesis. The LHC and the ATLAS experiment are described in Chap. 4, being the experimental setup necessary to perform this research. Chap. 5 covers concepts of hadron collider physics, with a focus on jets, reviewing all the technical information referenced in the following chapters. Chap. 6 provides a brief summary of the concepts of statistics and machine learning applied in this work. The latter part of the manuscript discusses the research contributions: Chap. 7 presents the contributions to

the search for new heavy resonances, and Chap. 8 details the contributions to the HL-LHC trigger upgrade. Chap. 9 summarizes the findings and their implications for future research.

Chapter 2

The Standard Model (SM)

Particle physics is the study of the fundamental particles of nature and their interactions. Already in 430 BC the philosopher Democritus theorized a Universe composed of fundamental building blocks that he named *atomos* – “indivisible” in Greek. What constitutes an elementary particle has, however, evolved over time. By the first half of the last century, it was well established that atoms — the elements of the periodic table — were, in fact, divisible, composed of a tightly bound nucleus made of protons and neutrons and a cloud of electrons around it. Three fundamentally different types of interaction were also known: the very feeble force of gravity, responsible for making Newton’s apple fall from the tree and for keeping the planets in orbit around the Sun; the electromagnetic force, which seemed to govern most of the physical processes in our daily life and was described by a mature theory developed in the previous century; and a strong force that prevented the positively charged protons from tearing the nucleus apart, but whose fundamental nature remained a mystery. In the course of the last century, serendipity coupled with technological advancements led experimental physicists to observe unexpected new particles and phenomena. It was shown that protons and neutrons were not elementary, but rather composed of a new type of particle called *quarks* coming in two *flavors* (*up* and *down*). Electrons and up and down quarks were observed to have sibling particles, which behaved similarly, but with heavier masses. A new form of interaction was also theorized to explain new observed phenomena, such as radioactive decays, which required the existence of new types of particles, the neutrinos. It was called the weak interaction, owing its name to being much feebler than the strong and electromagnetic forces. After a century of discoveries and a mix of failures and successes, a coherent description of what (for now) are known to be the fundamental building blocks of nature came into shape into what is called the Standard Model (SM).

The SM¹ is the mathematical framework of particle physics, describing the fundamental particles of nature and their electromagnetic, weak, and strong interactions (gravity is still not included, but since its strength is much weaker than any other force, its absence does not

¹A more complete introduction and in-depth explanation of the topics discussed in this chapter can be found in the textbooks and reviews this chapter is based on [12–18].

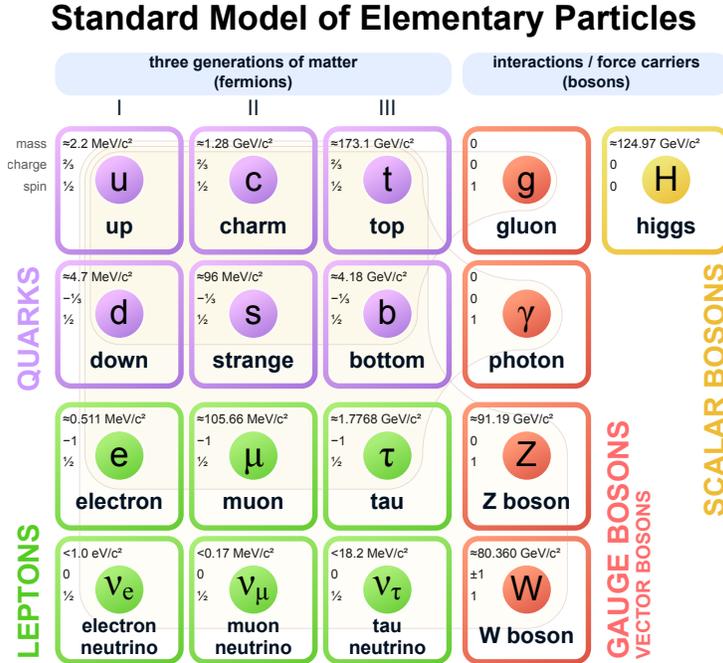


Figure 2.1: The Standard Model of elementary particles [11].

affect the predictive power of the model in most conditions). The particles and interactions that it describes are summarized in Fig. 2.1. All the fundamental particles that had been observed before 2012 fell into one of two categories determined by their spin quantum number: fermions with spin $1/2$ and gauge bosons with spin 1. Fermions make up all ordinary matter. They interact via the fundamental forces to form nuclei, heat up the Sun, and run the electric current in our computers. Gauge bosons are the mediators of these forces.

Fermions are of two types: the leptons and the quarks. The six quarks are organized in pairs of one *up-type* and one *down-type* quark, and the pairs are arranged in three *generations* of increasing mass and different flavor quantum number. The up (u), charm (c), and top (t) quarks have electric charge $Q = 2/3$, while the down (d), strange (s), and bottom (b) quarks have $Q = -1/3$. Similarly, leptons are arranged in pairs across three generations of increasing mass and different lepton quantum number. Each pair is composed of an electrically charged lepton with $Q = -1$ and its associated neutrino with no electric charge. These are the electron (e) and the electron-neutrino (ν_e), followed by the heavier muon (μ) and tau (τ) leptons and their respective neutrinos. Each fermion particle has a corresponding anti-particle, with equal mass but opposite quantum numbers. An interesting feature of the SM is that atoms, and hence all ordinary matter, are composed only of fermions from the first generation, while the heavier siblings are unstable and are only produced for short times before decaying. All fermions with non-zero electric charge participate in the electromagnetic

interaction. In addition to the electric charge, quarks and leptons carry an isospin charge and hence participate in the weak interaction. Quarks are the only fermions that carry another quantum number, called color charge, which allows them to interact via the strong force.

The fundamental forces are characterized by their strength, determined by their coupling constants, and by their range, determined by the mass of the gauge boson that mediates the interaction. The photon is the mediator of the electromagnetic interaction. Because the force is long-range, due to the photon being massless, it is the force that we interact with the most in our daily lives. The weak interaction is mediated by the W and Z bosons, which are some of the heaviest particles observed in nature, in the order of 100 GeV, making the weak interaction very short range. Nonetheless, the weak force is necessary to explain important phenomena, such as β -decay. Lastly, the strong interaction is mediated by the gluon. Like the photon, the gluon is massless, making the strong interaction technically long range. However, the coupling of the strong interaction has the peculiar feature of increasing at larger distances, which has the effect of preventing individual quarks to ever be observed alone. As a consequence, the strong force is effectively mediated by the exchange of massive particles called mesons, composed of a quark and an anti-quark. The mass of the lightest meson, the pion, gives nuclear forces an effective range of about 10^{-15} m, which controls the size of the atomic nucleus.

In 2012, a new type of particle, whose existence had been predicted decades earlier, was finally discovered [19]. The Higgs boson was the first fundamental particle to have been observed with zero spin. This fundamentally different nature allowed the Higgs boson to play a special role in shaping the Universe we live in, including being responsible for the mechanism that gives mass to all other particles. It is currently believed that at the time of the Big Bang the vacuum state of the Universe was symmetrical. At this time, all particles were massless and the four fundamental forces were unified into one single force. Then, shortly after the Big Bang, the potential of the Higgs field changed shape, the symmetrical position that used to be the lowest energy state became unstable, and the Universe decayed into a lower vacuum energy state that broke the symmetry. Upon the spontaneous symmetry breaking, the three weak gauge bosons and the fermions acquired mass and the original symmetry unifying the weak and electromagnetic interaction was hidden from view. The Universe we live in is currently in this broken phase.

Because elementary particles are, by definition, microscopic, and can easily reach velocities close to the speed of light, they have to be described by equations that obey both the laws of relativity and of quantum mechanics. Such a theory is a relativistic quantum field theory, where quantum mechanics is applied to dynamical systems of relativistic fields. Forces and fundamental particles are both described as fields that permeate the four-dimensional

space-time we are in. The particles that we detect are localized vibrations — or *quanta* — of the field and propagate through it like waves. As it turns out, the SM is a special type of quantum field theory, referred to as a *gauge theory*, where the fields are invariant under certain space-time-dependent phase transformations.

In the 1960s, Glashow proposed the unification of the electromagnetic and weak interactions using local gauge symmetry arguments [20]. However, his model predicted massless weak gauge bosons and fermions, in disagreement with the experimental observations. In the same years, it was discovered that a local gauge symmetry could be spontaneously broken by the addition of a massless complex scalar field, which would give rise to massive gauge bosons. This phenomenon, called the *Higgs mechanism*, was proposed in 1964 independently by Higgs [21], and Englert and Brout [22], opening the possibility of constructing an electroweak gauge theory with massive particles. The Higgs mechanism was applied to Glashow’s theory of the electroweak interaction by Weinberg [23] and Salam [24]. The prediction of electroweak symmetry breaking completed the last missing piece of the SM electroweak theory, also known as the Glashow-Weinberg-Salam (GWS) model. In the 1970s, a non-Abelian gauge theory of the strong interaction of quarks and gluons came also to maturity, and combined with the GWS model, forms what today is known as the SM of particle physics.

Glashow, Weinberg, and Salam were awarded the Nobel Prize in Physics in 1979 for “their contributions to the theory of the unified weak and electromagnetic interaction between elementary particles.” Experimental confirmation of their predictions soon followed with the discovery of the massive W and Z gauge bosons at CERN in 1983². Lastly, as mentioned earlier, the Higgs boson was finally discovered in 2012 at CERN by the ATLAS [25] and CMS [26] Collaborations, ultimately confirming the validity of the SM, and followed shortly after by the Nobel Prize in Physics awarded to Higgs and Englert.

The SM has proven to be a remarkably successful description of nature, whose structure was dictated by symmetries and guided by the experimental discoveries of the past century. However, it remains an empirical model, with several free parameters whose measured values bring to the surface a non-intuitive and unexplained structure. Several phenomena remain also unaccounted for, including gravity and evidence of dark matter. This leads physicists to regard the SM as an effective theory, valid only up to a certain energy scale. The belief that a more fundamental theory exists motivates the quest for beyond-the-SM (BSM) physics.

In Secs. 2.1 through 2.4, fundamental concepts for the development of a quantum field theory are introduced. The Higgs mechanism is discussed in Sec. 2.5. The SM Lagrangian is introduced in Sec. 2.6 and a more detailed discussion of the Higgs sector is presented in

²To which also followed the Nobel Prize in Physics in 1984 to Rubbia and Van der Meer for “their decisive contributions to the large project.”

Sec. 2.7. The motivation for looking for BSM physics will be briefly discussed in Sec. 2.8 and it will be the topic of the next chapter.

2.1 A quantum theory of fields

The concept of field was already introduced in Maxwell's classical formulation of electrodynamics as a way to prevent action at a distance [16]. Imagine a test charge placed in proximity of a source charge that will instantaneously feel the effect of an electric force produced by the source charge. Without an intermediary — a force carrier — this seems to violate locality. The problem of action-at-a-distance was solved by the introduction of the concept of field, where a field is a function that assigns a value to every point in space and time. An electromagnetic field permeates space, so that when a source charge is placed in the field, the field responds to it locally and then propagates the effect through the field at the speed of light. When a test charge is introduced at some distance away, it feels the influence of the modified field instantaneously. The classical theory of electrodynamics was well established by the beginning of the last century (and later found to be already consistent with special relativity). However, shortly after these successes, the new paradigm of quantum physics started to emerge, requiring a fundamental alteration of our understanding of nature.

A systematic quantum theory of fields started with Dirac's 1927 paper [27]. The solutions to Maxwell's equations in free space³, are transverse waves whose Fourier components behave like individual harmonic oscillator modes. Upon canonical quantization of the dynamical variables – the energy and the phase – describing each individual mode⁴, Dirac showed the equivalent interpretation of the number of quanta of energy as the number of particles moving at the speed of light and satisfying Bose-Einstein statistics, i.e. the number of photons. It follows that in quantum mechanics photons are excitations of the electromagnetic field and

³Maxwells' equations in free space can be written as

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \partial^2\right) A^\mu = 0. \quad (2.1)$$

Here, the field $A^\mu(\mathbf{x}, t) = (V, \mathbf{A})$ is the four-vector ($\mu = 0, 1, 2, 3$) electromagnetic field, introduced in place of the classical electric and magnetic fields, which can be obtained as $\mathbf{E}(\mathbf{x}, t) = -\nabla V - \frac{\partial \mathbf{A}}{\partial t}$ and $\mathbf{B}(\mathbf{x}, t) = \nabla \times \mathbf{A}$. Note that \mathbf{E} and \mathbf{B} are the observable physical fields. For a fixed choice of the fields \mathbf{E} and \mathbf{B} , \mathbf{A} and V are not unique under certain type of transformations called, as shown later, *gauge* transformations.

⁴Dirac defined the new operators a and a^\dagger as a linear combination of the position q and momentum p operators: $a = (1/\sqrt{2\omega})(\omega q + ip)$. He then showed that the Hamiltonian could be written as $H = \hbar\omega(a^\dagger a + \frac{1}{2})$, with eigenvalues $E_n = \hbar(n + \frac{1}{2})$, where $n = 0, 1, 2, \dots$ can be interpreted as the number of quanta of energy, and a and a^\dagger as annihilation and creation operators of the quanta.

can be created and annihilated as quanta of the field. However, electrons and the other particles still obey the Schrödinger equation. As we will see, this picture was not complete for a quantum theory of relativistic particles.

Quantum mechanics (QM) results from the quantization of a classical theory of particles described by their positions and momenta, but if one tries to write down a single particle relativistic wave equation, several issues arise. These are in part due to the fact that QM does not allow the number of particles in a system to change while, as it turns out, requiring the validity of both the laws of special relativity and of QM implies that the number of particles in a system is not conserved. An intuitive argument can be made as follows [15]. From relativity, one inherits Einstein's equation $E = mc^2$, according to which the mass of a particle is proportional to its energy. QM provides Heisenberg's uncertainty principle, which can be expressed as $\Delta E > \hbar c / \Delta x$, stating that the more accurately one knows the position of a particle, the less accurately one will be able to know its energy. It follows that when $\Delta E \geq 2mc^2$, enough energy is available to produce a particle-anti-particle pair. In other words, if a physical system is probed at a length scale $\Delta x \lesssim \hbar / (2mc)$, the concept of a single particle breaks down, as the uncertainty in the energy is now large enough to allow for a cloud of particle-anti-particle pairs to surround the particle. A new framework is needed to describe this phenomenon.

In quantum field theory (QFT), fields are introduced to describe not only the photon, but also the electron. Because any relativistic theory has to obey Einstein's first principle of relativity (i.e. has to be Lorentz invariant), the particles are more appropriately described by four-momentum vectors $p_\mu = (t, p_x, p_y, p_z)$ in Minkowski space, where space and time are treated on equal footing. Space is therefore declassified from being an operator $\hat{\mathbf{x}}(t)$, as in quantum mechanics, to being a label identifying a space-time coordinate $x^\mu = (t, \mathbf{x})$ of the field, while the operator is now the field $\phi(x^\mu)$, which acts at every point in space-time.

Of particular importance in these developments was Yukawa's paper in 1935 that demonstrated how the interaction between particles could be shown to proceed via the exchange of virtual quanta – or *mediators* – of the force field [16]. In QFT, the field is the object of the quantization and both particles and force carriers arise as excitations of the fields and can be created and annihilated, just like the photon in QM. For instance, the electron and its anti-particle, the positron, can be viewed as the quanta of the electron-positron field. The particle and field pictures are equivalent in describing the system, but it turns out that the fields are the natural way to describe mathematically what is happening at these small distances [15]. Many other fields associated to new particles and interactions had to be introduced to make this description complete.

2.2 The Lagrangian formulation

Similarly to the classical approach, the equations of motion for a relativistic field can be derived from the Lagrangian L by the principle of least action. The action is expressed as

$$S = \int L dt = \int_{\Omega} \mathcal{L}(\phi(x), \partial_{\mu}\phi(x)) d^4x, \quad (2.2)$$

where \mathcal{L} is the Lagrangian density, which from now on will be referred to simply as the Lagrangian. This substitution is useful, as the four-dimensional volume element is Lorentz invariant, making the action explicitly Lorentz invariant provided \mathcal{L} is a Lorentz scalar. Note that \mathcal{L} is considered to be a functional of the fields and their first order time and spatial derivatives only⁵. The principle of least action requires the variation of the action δS to be zero for small fluctuations of the fields $\phi(x) \rightarrow \phi(x) + \delta\phi(x)$. Imposing this requirement brings to the Euler-Lagrange (EL) equations of motion for a field [12],

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_{\mu} \left(\frac{\partial \mathcal{L}}{\partial (\partial_{\mu} \phi)} \right) = 0. \quad (2.3)$$

In a combined treatment of particles and fields, the Lagrangian has three terms: a free field Lagrangian, a free-particle Lagrangian, and an interaction Lagrangian which describes the interaction between particles and fields. According to which degrees of freedom are considered for the variation of the action integral, EL equations of motion of particles or fields can be derived. A similar formulation for the dynamics of the system could be obtained in terms of the Hamiltonian. However, the Lagrangian formulation is particularly well suited for QFT as the theory is manifestly relativistically covariant and its symmetry properties and associated conservation laws are directly identifiable from the Lagrangian.

One could argue that the goal of particle physics is to find a model, defined by a Lagrangian, that describes the fundamental laws of nature [28]. In practice, one needs to identify what the symmetries and the fields in the theory are, and how the fields transform under the symmetries. The symmetries, plus a few theoretical requirements such as locality⁶ and renormalizability⁷, allow one to identify all the terms allowed in the Lagrangian. Once the Lagrangian is defined, the predictive power of the model can be tested against experi-

⁵This requirement allows one to treat space and time on equal footing and is sufficient to describe the physics observed by experiment.

⁶In a local theory the Lagrangian can only contain products of fields evaluated at the same space-time location. This removes the possibility of action-at-a-distance [13].

⁷A theory is renormalizable if all its physical predictions remain finite and well-defined once all the cut-offs of the theory are removed [13].

ment. In the following section, it will be shown why symmetries play such a fundamental role in model building due to their connection to conservation laws.

2.3 Symmetries and conservation laws

The Lagrangian is said to be invariant under a transformation if, when expressed in the new transformed coordinates and fields, it preserves the same functional form as the original Lagrangian (up to a 4-divergence, as such a term does not affect the derivation of the EL equations of motion). Assume the Lagrangian is invariant under some continuous transformation of the field

$$\phi(x) \rightarrow \phi'(x) = \phi(x) + \epsilon \delta\phi(x) + \mathcal{O}(\epsilon^2), \quad (2.4)$$

where ϵ is an infinitesimal parameter and $\delta\phi(x)$ is some deformation of the field configuration. Then, $\mathcal{L}(\phi(x), \partial\phi(x)) = \mathcal{L}(\phi'(x), \partial\phi'(x))$ and one can show that

$$\delta\mathcal{L} = 0 = \left[\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \right) \right] \delta\phi + \partial_\mu \left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \delta\phi \right]. \quad (2.5)$$

From the EL equations, the first term vanishes. Therefore, the system has a conserved current $\partial^\mu J_\mu = 0$ and a corresponding conserved charge⁸ given by [13],

$$J_\mu = \frac{\partial\mathcal{L}}{\partial(\partial^\mu\phi)} \delta\phi \quad \text{and} \quad Q \equiv \int d^3x J^0. \quad (2.6)$$

This result can be easily generalized to the case of transformations involving also the space-time coordinates and is known as Noether's theorem. More formally, the theorem states that for every continuous symmetry that leaves the Lagrangian invariant there is a conserved current and a corresponding locally conserved charge. For example, the invariance of \mathcal{L} under translations in time and space implies conservation of energy and momentum. The transformations are required to be unitary, as this ensures observable predictions to be invariant.

An important class of symmetries are internal symmetries, which involve transformations of the fields themselves and act identically at every point in space-time [15]. As an example, consider the Lagrangian density describing a free Dirac fermion

$$\mathcal{L}_0 = \bar{\psi}(x)(i\gamma^\mu\partial_\mu - m)\psi(x). \quad (2.7)$$

⁸The condition $\partial^\mu J_\mu = 0$ guarantees that $dQ/dt = 0$.

This Lagrangian is invariant under continuous rotations of the phase of $\psi(x)$ as $\psi(x) \rightarrow e^{i\alpha}\psi(x)$. Such rotations belong to the one-dimensional unitary group of transformations $U(1)$, whose operators are one-dimensional unitary matrices, i.e. complex numbers of unit modulus. These transformations bring the system from one physical state to a different one with the same physical properties. According to Noether's theorem, this invariance determines the conservation of some quantity. In general, the number of conserved quantities is equal to the number of the generators of the group of transformations. In this case, there is one conserved current $j^\mu = \bar{\psi}(x)\gamma^\mu\psi(x)$. This type of transformation is said to be *global*, to differentiate it from what are known as *local* gauge transformations.

As an example of a local gauge transformation, consider the free field Lagrangian of quantum electrodynamics (QED)

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (2.8)$$

where the field strength tensor is given by $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. This Lagrangian is invariant under the symmetry $A_\mu(x) \rightarrow A_\mu + \partial_\mu f(x)$, for any function $f(x)$:

$$F_{\mu\nu}(x) \rightarrow F'_{\mu\nu} = \partial_\mu(A_\nu + \partial_\nu f(x)) - \partial_\nu(A_\mu + \partial_\mu f(x)) = F_{\mu\nu}. \quad (2.9)$$

According to Noether's theorem, this should produce an infinite number of conserved quantities. However, these are not true internal symmetries, but expressions of a redundancy of degrees of freedom in the description of the system. If one tries to apply Noether's theorem for any of these transformations, it results in the same conserved quantity as for the global transformation where $f(x) = \text{const}$. When this is the case, the system is more correctly described as a set of configurations related to each other by a group of transformations. This type of symmetry is called a *gauge symmetry* or *gauge invariance*, and the vector field A_μ is called a *gauge field*. As shown later, to remove the redundancy one can "fix the gauge" by imposing some extra condition on the vector potential.

2.4 Deriving a gauge theory

The free field Lagrangian in Eq. (2.8) describes the electromagnetic theory in the absence of sources, while Eq. (2.7) describes free fermions. If one wants to build an interacting theory of light and matter, a new term has to be included, which couples A_μ to the matter fields. How to add the interaction term?

The Maxwell Lagrangian $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - j^\mu A_\mu$, called in this way because its equations

of motion are Maxwells' equations, adds the interaction via the term $j^\mu A_\mu$, where j^μ is a conserved current dependent on the fermion fields. Recall that the free fermion Lagrangian is invariant under the global U(1) phase transformation. To this true internal symmetry of the theory corresponds the conserved current $j^\mu = \bar{\psi}(x)\gamma^\mu\psi(x)$, which can be shown to result in the conservation of the electric charge e . A good attempt at including the interaction between the matter and the field is then

$$\mathcal{L}^{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(x)(i\gamma^\mu\partial_\mu - m)\psi(x) - e\bar{\psi}(x)\gamma^\mu A_\mu\psi(x), \quad (2.10)$$

where e has been introduced as the coupling constant. This is referred to as *minimal interaction*. But while the original free field Lagrangian (Eq. (2.8)) was invariant under the local gauge transformation

$$A_\mu(x) \rightarrow A'_\mu = A_\mu + \partial_\mu f(x), \quad (2.11)$$

the new interaction term is not. The invariance can be restored if the transformation of the vector field A_μ is coupled to the local gauge transformation of the fermion field

$$\psi(x) \rightarrow \psi(x)' = e^{iqf(x)}\psi(x), \quad (2.12)$$

$$\bar{\psi}(x) \rightarrow \bar{\psi}'(x) = \bar{\psi}(x)e^{-iqf(x)}. \quad (2.13)$$

The Lagrangian in Eq. (2.10) is invariant under the coupled gauge transformations from Eq. (2.11) and (2.12) and is in fact the QED Lagrangian sufficient to describe the experimental observations.

This derivation was only possible because QED had a fully developed classical counterpart in Maxwell's equations to guide it. However, it provided a prescription to derive other gauge theories without starting from classical inputs. When this derivation was generalized for other types of interaction, the procedure was reversed.

Using again QED as an example, one starts from the free particle Lagrangian and identifies the global U(1) phase transformation. This invariance indicates that the phase of the field $\psi(x)$ has no physical meaning, as one can rotate $\psi(x)$ by an arbitrary real constant at all points in space-time and obtain the same dynamics. However, if one allows the phase to depend on the space-time coordinate x , i.e. if one applies the local gauge transformation from Eq. (2.12), the Lagrangian is no longer invariant, as now

$$\partial_\mu\psi(x) \rightarrow e^{iqf(x)}(\partial_\mu + iq\partial_\mu f(x))\psi(x). \quad (2.14)$$

Thus, while the global phase of the field depends only on the chosen convention, it has to be fixed for all space-time points. This type of restriction seems unnatural and brought to the ideation of the “gauge principle”, or the requirement of local gauge invariance. In order to restore local gauge invariance, one introduces a vector field $A_\mu(x)$ that transforms in such a way as to cancel the $\partial_\mu f(x)$ term:

$$A_\mu(x) \longrightarrow A'_\mu = A_\mu + \partial_\mu f(x) \quad (2.15)$$

Then one changes the derivative $\partial_\mu \psi(x)$ to the *covariant derivative*

$$D_\mu \psi(x) = [\partial_\mu + ieA_\mu] \psi(x) \quad (2.16)$$

which has the property of transforming like the field itself. Note that replacing the ordinary derivative $\partial_\mu \psi(x)$ with the covariant derivative $D_\mu \psi(x)$ is equivalent to introducing the interaction term. The gauge field A_μ appears as the mediator of the electromagnetic interaction that couples to the field ψ with coupling strength proportional to e . Note that a mass term for the gauge field $\frac{1}{2}m^2 A_\mu A^\mu$ is forbidden as it would break the gauge invariance of the Lagrangian. Therefore, QED predicts the photon to be massless. To allow the new vector field to propagate in space, a gauge invariant kinetic term is added, which corresponds to the free field Lagrangian in Eq. (2.8). The final QED Lagrangian

$$\mathcal{L}^{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(x)(i\not{D} - m)\psi(x) \quad (2.17)$$

is manifestly invariant under the coupled gauge transformations from Eqs. (2.11) and (2.12).

QED is the simplest example of a gauge theory, where gauge fields are included in the Lagrangian to ensure local gauge invariance. The gauge field is a dynamical variable that interacts with other particles, as well as with itself. Upon quantization, the quanta of the gauge fields are called the gauge bosons. The number of gauge fields needed to restore local gauge invariance under the given gauge symmetry group is equal to the number of generators of the group. When the symmetry group is non-commutative, the theory is a non-Abelian gauge theory.

The concept of a gauge theory was formalized by Yang and Mills in 1954 starting from the Abelian gauge theory of QED and extended to non-Abelian gauge theories. This formulation had very fruitful implications. The modern theories of the strong and electroweak interactions are both examples of non-Abelian gauge theories and form what today is called the SM, whose mathematical formulation can be derived by the requirement of local gauge invariance under the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry group and the addition of

a scalar particle to drive the Higgs mechanism.

2.5 The Higgs Mechanism

The Higgs mechanism occurs when spontaneous symmetry breaking (SSB) happens within a gauge theory. The phenomenon of SSB occurs when the ground state of a system is not symmetric under a symmetry of its Lagrangian. Consider a Lagrangian that possesses a given symmetry and whose ground state is degenerate, so that the ground state eigenstates transform among themselves under the symmetry of the Lagrangian. When the system settles in its ground state, one of the degenerate states is arbitrarily chosen. The ground state is then no longer invariant under the original symmetry, which is now hidden. An example of SSB is ferromagnetism. In a ferromagnet, the ground state of the system requires the spins to be aligned along some direction producing a non-zero magnetization \vec{M} . The ground-state magnetization can be oriented in any direction because the system is invariant under rotation, but once the ferromagnet cools down and a choice for the direction is made, the system is not invariant under rotation anymore. Therefore, the choice of a ground state spontaneously breaks the global rotational symmetry of the system.

In a field theory, the ground state is the vacuum, so SSB can only occur if the vacuum state is not unique. To preserve Lorentz and translation invariance of the vacuum state, any spinor or vector field vacuum expectation value must vanish $\langle 0|\psi(x)|0\rangle = \langle 0|V^\mu(x)|0\rangle = 0$, so that in order to break the symmetry a scalar field $\phi(x)$ has to be introduced.

In the following [13], the Goldstone model is presented as a simple example of SSB in a field theory to illustrate how SSB leads to the appearance of massless particles known as Goldstone bosons. When SSB is applied to a gauge theory, however, things are a bit different. In the context of a gauge theory, gauge fixing allows to convert the new non-physical degrees of freedom of the Goldstone bosons into mass terms for the gauge vector bosons. The original gauge symmetry is broken, but its effect remains visible in the way the interactions of the massive vector bosons are constrained. Via this mechanism, called the Higgs mechanism, the gauge bosons acquire mass and a new massive scalar field remains in the theory, the Higgs boson. This will be illustrated using the simplest example of a U(1) gauge theory. The mechanism was studied and generalized to the case of a non-Abelian gauge theory by Higgs, Kibble, Guralnik, Hagen, Brout, and Englert, and was subsequently applied to the gauge theory of electroweak interactions by Weinberg and Salam.

The Goldstone model

Consider a complex scalar field $\phi(x) = \frac{1}{\sqrt{2}}[\phi_1(x) + i\phi_2(x)]$ described by the Lagrangian

$$\mathcal{L}(x) = \partial^\mu \phi^*(x) \partial_\mu \phi(x) - \mu^2 |\phi(x)|^2 - \lambda |\phi(x)|^4, \quad (2.18)$$

invariant under the global U(1) phase transformation

$$\phi(x) \rightarrow \phi'(x) = e^{i\alpha} \phi(x), \quad \phi(x)^* \rightarrow \phi'^*(x) = e^{-i\alpha} \phi^*(x) \quad (2.19)$$

The potential of this Lagrangian is

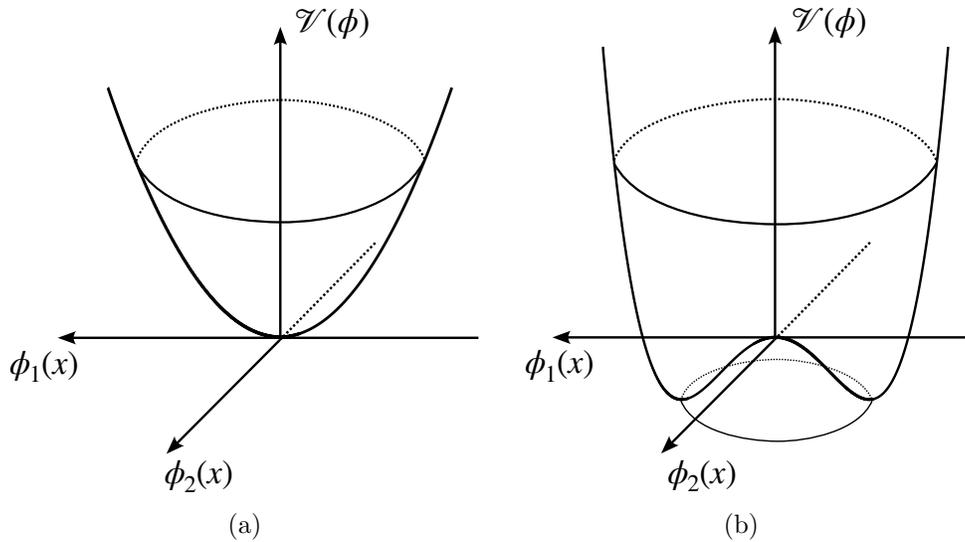


Figure 2.2: Potential of Eq. (2.20) with $\lambda > 0$, for (a) $\mu^2 > 0$ and (b) $\mu^2 < 0$. Adapted from Ref. [13].

$$V(\phi) = \mu^2 |\phi(x)|^2 + \lambda |\phi(x)|^4, \quad (2.20)$$

with $\lambda > 0$ for it to be bounded from below. The parameter μ can take on two possible values, as shown in Fig. 2.2. For $\mu^2 > 0$, the potential has a unique absolute minimum at $\phi(x) = 0$, while for $\mu^2 < 0$, the potential has a circle of absolute minima at $\phi(x) = \phi_0 = \left(\frac{-\mu^2}{2\lambda}\right)^{1/2} e^{i\theta}$ for $0 \leq \theta < 2\pi$. The ground state ϕ_0 is degenerate, as the angle θ determines an arbitrary direction in the complex plane. The choice of one particular ground state breaks the rotational U(1) symmetry of the theory. Without loss of generality, one can choose the

ground state ϕ_0 to be at $\theta = 0$, so that ϕ_0 is on the real axis,

$$\phi_0 = \left(\frac{-\mu^2}{2\lambda} \right)^{1/2} = \frac{1}{\sqrt{2}}v \quad (2.21)$$

One can then redefine the field $\phi(x)$ in terms of deviations from the equilibrium ground state

$$\phi(x) = \frac{1}{\sqrt{2}}[\phi_1(x) + i\phi_2(x)] \longrightarrow \phi(x) = \frac{1}{\sqrt{2}}[(v + \sigma(x)) + i(\eta(x))], \quad (2.22)$$

where $\sigma(x)$ and $\eta(x)$ are two real fields. Rewriting the Lagrangian with this substitution

$$\mathcal{L}(x) = \frac{1}{2}\partial^\mu\sigma(x)\partial_\mu\sigma(x) - \frac{1}{2}(2\lambda v^2)\sigma^2(x) + \frac{1}{2}\partial^\mu\eta(x)\partial_\mu\eta(x) \quad (2.23)$$

$$- \lambda v\sigma(x)[\sigma(x)^2 + \eta^2(x)] - \frac{1}{4}\lambda[\sigma^2(x) + \eta^2(x)]^2, \quad (2.24)$$

the first line can be interpreted as the free field Lagrangian, while the second line contains the interaction terms between the fields $\sigma(x)$ and $\eta(x)$. From the first line one can infer that $\sigma(x)$ and $\eta(x)$ are real Klein-Gordon fields, which, upon quantization, lead to a spin-0 $\sigma(x)$ boson with mass $\sqrt{2\lambda v^2}$ and a massless spin-0 $\eta(x)$ boson. Note that the massive $\sigma(x)$ field describes oscillations of $\phi(x)$ along the radial direction of the potential, where $V(\phi)$ has a non-vanishing second derivative, while the massless $\eta(x)$ field is associated to displacements in the tangential direction of constant $V(\phi)$. The $\eta(x)$ boson is an example of a Goldstone boson, a massless particle that appears in a field theory as a consequence of the degeneracy of the ground state. This is formalized by the Goldstone Theorem, which states that for every spontaneously broken continuous symmetry, the theory contains massless scalar particles whose number is equal to the number of broken symmetries.

The Higgs mechanism

To consider the simplest example of SSB in a gauge theory, one can generalize the Goldstone model by requiring invariance under a local U(1) phase transformation of the same Lagrangian. Following the prescription to derive a gauge theory, the covariant derivative $D_\mu\phi(x) = [\partial_\mu + iqA_\mu(x)]\phi(x)$ is introduced, with the resulting Lagrangian given by

$$\mathcal{L}(x) = [D^\mu\phi(x)]^*[D_\mu\phi(x)] - \mu^2|\phi(x)|^2 - \lambda|\phi(x)|^4 - \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) \quad (2.25)$$

The Lagrangian has the same potential and is invariant under the coupled U(1) gauge transformations, similarly to Eqs. (2.11) and (2.12). Performing the same substitution into \mathcal{L} ,

one obtains

$$\mathcal{L}(x) = \frac{1}{2}\partial^\mu\sigma(x)\partial_\mu\sigma(x) - \frac{1}{2}(2\lambda v^2)\sigma^2(x) \quad (2.26)$$

$$- \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + \frac{1}{2}(qv)^2 A_\mu(x)A^\mu(x) \quad (2.27)$$

$$+ \frac{1}{2}\partial^\mu\eta(x)\partial_\mu\eta(x) \quad (2.28)$$

$$+ qvA^\mu(x)\partial_\mu\eta(x) + \text{interaction terms} \quad (2.29)$$

While the result looks similar to what was obtained with the Goldstone model, the interpretation of the second line as a massive vector field and of the third line as a massless boson fails because of the term in the last line, which mixes derivatives of A_μ and $\eta(x)$, making the two fields not independent. However, upon more careful look, one can notice that the new Lagrangian contains an extra degree of freedom. This can be removed by an appropriate choice of gauge. Specifically, in the unitary gauge, a $U(1)$ rotation is used to transform $\phi(x)$ into a real field $\phi(x) = \frac{1}{\sqrt{2}}[v + \sigma(x)]$. Upon the transformation, the $\eta(x)$ field disappears and the Lagrangian becomes,

$$\mathcal{L}(x) = \frac{1}{2}\partial^\mu\sigma(x)\partial_\mu\sigma(x) - \frac{1}{2}(2\lambda v^2)\sigma^2(x) \quad (2.30)$$

$$- \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + \frac{1}{2}(qv)^2 A_\mu(x)A^\mu(x) \quad (2.31)$$

$$- \lambda v\sigma^3(x) - \frac{1}{4}\lambda\sigma^4(x) + \frac{1}{2}q^2 A_\mu(x)A^\mu(x)[2v\sigma(x) + \sigma^2(x)]. \quad (2.32)$$

The first two lines can now be interpreted as the free field Lagrangian of a scalar boson of mass $\sqrt{2\lambda v^2}$ and a vector boson with mass $|qv|$, respectively. Via SSB, the original Lagrangian with a complex scalar field and a massless real vector field turned into a Lagrangian of a real scalar field and a massive real vector field. The number of degrees of freedom remained constant (fixed to 4), but one of the two degrees of freedom of the complex scalar field $\phi(x)$ was taken up by the vector field that has become massive. This was done via the Goldstone boson $\eta(x)$, which appeared because of SSB, but was unphysical and could be eliminated by fixing the gauge. This phenomenon by which a Goldstone boson, produced as a consequence of SSB, gets “eaten” by a gauge boson that subsequently acquires mass is known as the *Higgs mechanism* and the massive spin-0 boson $\sigma(x)$ that survives is called a *Higgs boson*.

2.6 The Standard Model Lagrangian

The SM Lagrangian before spontaneous symmetry breaking describes the electromagnetic and weak interaction between quarks and leptons and the strong interaction between quarks. It contains two types of fields, the matter fields describing the spin 1/2 fermions and the spin 1 gauge bosons, which mediate the interactions and are introduced in the theory via the requirement of gauge invariance. This Lagrangian is formed from combining the $SU(2)_L \times U(1)_Y$ invariant electroweak theory and the $SU(3)_C$ gauge theory of quantum chromodynamics (QCD), resulting in an $SU(2)_L \times U(1)_Y \times SU(3)_C$ gauge invariant theory. In this Lagrangian the gauge bosons and the fermions are assumed massless, as introduction of mass terms breaks the gauge invariance of the theory. The addition of the Higgs scalar field is necessary to provide a mechanism for the fermions and gauge bosons to acquire masses, while preserving gauge invariance via the process of spontaneous symmetry breaking.

The SM Lagrangian can be summarized by four terms

$$\mathcal{L} = \mathcal{L}_{\text{Fermion}} + \mathcal{L}_{\text{Gauge bosons}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}. \quad (2.33)$$

The first two terms describe the fermion and gauge fields and their interactions, while the last two terms appear after the introduction of the Higgs doublet and represent the Higgs sector, discussed in the next section.

Fermions

The fermion fields are the quarks and leptons, whose free field Lagrangian is given by

$$\mathcal{L}_0^f = i\bar{\psi}_f(x)\gamma^\mu\partial_\mu\psi_f(x), \quad (2.34)$$

where f runs over each fermion type. It is useful to separate each spinor into its left-handed and right-handed components, according to how they transform under the helicity projection operators. For massless particles, or for massive particles in the high energy limit, the left-handed and right-handed charged lepton fields are defined as,

$$\psi_l^L(x) \equiv P_L\psi_l(x) = \frac{1}{2}(1 - \gamma_5)\psi_l(x), \quad (2.35)$$

$$\psi_l^R(x) \equiv P_R\psi_l(x) = \frac{1}{2}(1 + \gamma_5)\psi_l(x). \quad (2.36)$$

This is a useful distinction, because only left-handed fermions experience the weak force. In fact, the left-handed fields transform under the $SU(2)$ symmetry group of the weak interac-

tion as isospin doublets, while the right-handed components transform as singlets. In the SM neutrinos are assumed to be massless, so that only left-handed neutrinos and right-handed anti-neutrinos couple to SM interactions. The matter particles can then be summarized as:

$$\begin{array}{ccc}
L_1 = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, & l_{R1} = e_R^- & Q_1 = \begin{pmatrix} u \\ d \end{pmatrix}_L, & u_{R1} = u_R, & d_{R1} = d_R \\
L_2 = \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, & l_{R2} = \mu_R^- & Q_2 = \begin{pmatrix} c \\ s \end{pmatrix}_L, & u_{R2} = c_R, & d_{R2} = s_R \\
L_3 = \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}, & l_{R3} = \tau_R^- & Q_3 = \begin{pmatrix} t \\ b \end{pmatrix}_L, & u_{R3} = t_R, & d_{R3} = b_R
\end{array} \tag{2.37}$$

Leptons
Quarks

The leptons and quarks are grouped in the three generations and the left-handed components are combined into SU(2) doublets. In the SU(2)_L × U(1)_Y electroweak theory, gauge invariance determines the conservation of the weak hypercharge Y and of the weak isospin I , while spontaneous symmetry breaking of SU(2)_L × U(1)_Y → U(1)_{EM} brings to the conservation of the electric charge Q . The three are related by the relationship

$$Y = 2Q - 2I^3, \tag{2.38}$$

where I^3 is the third component of the weak isospin and Q is in units of the proton charge e . All isospin singlets have $I_i^R = 0$, indicating that they do not partake in the weak interaction. The isospin doublets have $I_i^L = \frac{1}{2}\tau_i$, where τ_i are the Pauli spin matrices, so that all upper (lower) components of an isospin doublet have $I_3^L = 1/2$ ($-1/2$). This leads to a hypercharge of $Y_l^L = -1$ for left-handed leptons, and $Y_l^R = -2$ for the right-handed singlets. Similarly, $Y_q^L = \frac{1}{3}$, $Y_u^R = \frac{4}{3}$ and $Y_d^R = -\frac{2}{3}$ for quarks.

In terms of left and right-handed fields, Eq. (2.34) can be rewritten as

$$\mathcal{L}_0^f = \bar{L}_i i D_\mu \gamma^\mu L_i + \bar{e}_{Ri} i D_\mu \gamma^\mu e_{Ri} + \bar{Q}_i i D_\mu \gamma^\mu Q_i + \bar{u}_{Ri} i D_\mu \gamma^\mu u_{Ri} + \bar{d}_{Ri} i D_\mu \gamma^\mu d_{Ri}. \tag{2.39}$$

The fact that left-handed and right-handed components transform differently under SU(2) prevents fermion mass terms to be added explicitly in the Lagrangian, as mass terms mix left and right-handed components, which violates SU(2) gauge invariance:

$$-m_e \bar{\psi}_e \psi_e = -m_e \bar{\psi}_e \left(\frac{1}{2}(1 - \gamma_5) + \frac{1}{2}(1 + \gamma_5) \right) \psi_e = -m_e (\bar{\psi}_e^R \psi_e^L + \bar{\psi}_e^L \psi_e^R). \tag{2.40}$$

In addition to participating in the electroweak interaction, the quark fields Q_i , u_i^R , and d_i^R ($i = 1, 2, 3$) carry a color charge associated to the SU(3) symmetry of the strong interaction. Any quark can exist in one of three different color states, denoted as *red*, *green*, and *blue*, and transform from one color state to another under the SU(3)_C group as triplets,

$$q_i = \begin{pmatrix} q_i^r \\ q_i^g \\ q_i^b \end{pmatrix}, \quad (2.41)$$

with $C = r, g, b$ representing the color charge. Only color singlet combinations are observed in nature, as baryons and mesons,

$$B = \frac{1}{\sqrt{6}} \epsilon^{\alpha\beta\gamma} |q_\alpha q_\beta q_\gamma\rangle, \quad M = \frac{1}{\sqrt{3}} \delta^{\alpha\beta} |q_\alpha q_\beta\rangle. \quad (2.42)$$

This is known as *color confinement*: quarks are confined within color-singlet bound states.

Gauge fields

The introduction of force mediators in the theory can be obtained via gauge symmetry arguments. The free-fermion Lagrangian of Eq. (2.39) is invariant under global U(1)_Y, SU(2)_L, and SU(3)_C transformations, which determines the conservation of hypercharge, weak isospin, and color charge, respectively.

Analogously to QED, the operators of the U(1) group are complex numbers of unitary module. The elements of the SU(2) group are 2×2 unitary matrices with determinant one. The generators of the group are $T^a = \frac{1}{2} \tau^a$, where τ^a ($a = 1, 2, 3$) are the Pauli spin matrices. The group is non-Abelian as the operators do not commute: $[T^a, T^b] = i \epsilon^{abc} T_c$, where ϵ^{abc} is the anti-symmetric tensor. Similarly, the elements of the fundamental representation of the SU(3)_C group are the set of unitary 3×3 matrices with determinant one, and the generators of the algebra are the matrices $T^a = \frac{1}{2} \lambda_a$ ($a = 1, 2, \dots, 8$), where λ_a are the Gell-Mann matrices. The matrices T^a satisfy the commutation relations $[T^a, T^b] = i f^{abc} T_c$, where f^{abc} are the SU(3) structure constants, which are real and totally antisymmetric.

In the context of electroweak theory, under the local SU(2)_L \times U(1)_Y gauge transformations, the fermion fields transform as

$$L(x) \rightarrow L'(x) = e^{i\alpha_a(x)T^a + i\beta(x)Y} L(x), \quad (2.43)$$

$$R(x) \rightarrow R'(x) = e^{i\beta(x)Y} R(x). \quad (2.44)$$

Similarly, in color space the fermion fields transform as

$$\psi_q(x) \rightarrow \psi'_q(x) = e^{i\alpha_a(x)\frac{\lambda^a}{2}}. \quad (2.45)$$

Following the prescription to derive a gauge theory, invariance under the local U(1) transformation requires the addition of one field, denoted as B_μ , similarly to the A_μ field in QED. Similarly, three gauge fields W_μ^i are included to preserve SU(2) gauge invariance, corresponding to the three generators of SU(2). Lastly, to the eight generators of the SU(3) group corresponds the octet of gluon fields, G_μ^i . The new fields, with their appropriate coupled gauge transformations, are introduced via the covariant derivative,

$$D_\mu\psi = \left(\partial_\mu - ig_s T_a G_\mu^a - ig_2 T_a W_\mu^a - ig_1 \frac{Y_q}{2} B_\mu \right) \psi. \quad (2.46)$$

with g_s , g_2 , and g_1 the coupling constants of SU(3)_C, SU(2)_L and U(1)_Y, respectively. Gauge invariant terms describing the free fields in the absence of fermions need to also be included. The final Lagrangian describing the free fermion and gauge boson fields, together with their interactions, is given by

$$\mathcal{L} = \sum_j \bar{\Psi}_L^j i\gamma^\mu D_\mu^L \Psi_L^j + \sum_{j,\sigma} \bar{\psi}_{R\sigma}^j i\gamma^\mu D_\mu^R \psi_{R\sigma}^j \quad (2.47)$$

$$- \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \quad (2.48)$$

where the gauge invariant field strength tensors are given by,

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c, \quad (2.49)$$

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g_2 \epsilon^{abc} W_\mu^b W_\nu^c, \quad (2.50)$$

$$G_{\mu\nu}^a = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (2.51)$$

Note that the richer structure of $G_{\mu\nu}$ and $W_{\mu\nu}$ is due to the non-Abelian nature of the corresponding groups. These terms are responsible for self-interaction vertices of the weak gauge bosons and of the gluons. Note also that the theory presented so far predicts massless gauge bosons. While the photon is indeed massless, the W and Z bosons are known not to be. Adding *ad hoc* mass terms for the W and Z bosons, such as $m_W^2 W_\mu^\dagger(x) W^\mu(x) + \frac{1}{2} m_Z^2 Z_\mu(x) Z^\mu(x)$, breaks gauge invariance. To include mass terms and preserve gauge invariance a new mechanism is necessary. After the discovery of the Higgs boson, this was confirmed to be the Higgs mechanism.

2.7 The Higgs sector

The electroweak SM Lagrangian is invariant under the $SU(2)_L \times U(1)_Y$ gauge symmetry group, with three generators associated to the $SU(2)$ symmetry and one to $U(1)$, for a total of four generators. Via the Higgs mechanism, one would like three of the vector bosons to acquire mass and one of them, the photon, to remain massless. In order to have SSB a scalar field with a non-vanishing vacuum expectation value invariant under some symmetry of the Lagrangian has to be introduced. To break the symmetries associated to three generators, at least three degrees of freedom are needed for the scalar field. The simplest choice (providing four degrees of freedom) is to add a complex scalar field $\phi(x)$ that is an isospin doublet of $SU(2)$ with hypercharge $Y_\phi = +1$,

$$\Phi = \begin{pmatrix} \phi_+(x) \\ \phi_0(x) \end{pmatrix} \quad (2.52)$$

where $\phi_+(x)$ and $\phi_-(x)$ are scalars under Lorentz transformations. Note that, according to Eq. (2.38), this choice of hypercharge makes the upper (lower) component of the doublet have $Q = 1$ ($Q = 0$). The simplest way of including the new field in the electroweak Lagrangian \mathcal{L}_{EW} is by letting $\mathcal{L} = \mathcal{L}_{EW} + \mathcal{L}_\phi$. The \mathcal{L}_{EW} term is already $SU(2) \times U(1)$ invariant, while the Lagrangian \mathcal{L}_ϕ describing the scalar field can be made invariant by introducing the EW covariant derivative from Eq. (2.46) and is given by,

$$\mathcal{L}_\Phi = (D^\mu \Phi(x))^\dagger [D_\mu \Phi(x)] - \mu^2 \Phi(x)^\dagger \Phi(x) - \lambda (\Phi^\dagger(x) \Phi(x))^2. \quad (2.53)$$

The scalar field has a very similar potential as in the Higgs model. For $\mu^2 < 0$, the vacuum state Φ_0 , which occurs at the minimum of the potential, is degenerate and occurs whenever $\Phi_0^\dagger \Phi_0 = \frac{-\mu^2}{2\lambda}$. Upon the choice of a particular vacuum expectation value (vev) for the ground state $\langle \Phi \rangle_0$, the system is not invariant under $SU(2) \times U(1)$ transformations anymore, so the symmetry is spontaneously broken. Without loss of generality, the value of the vev is chosen so that Φ develops a vev only in the lower component of the doublet,

$$\Phi_0 = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \quad \text{with } v = \left(-\frac{\mu^2}{\lambda} \right)^{1/2}. \quad (2.54)$$

One can parametrize the scalar field in terms of its deviations from the vacuum state Φ_0

and move to the unitary gauge,

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \eta_1(x) + i\eta_2(x) \\ v + \sigma(x) + i\eta_3(x) \end{pmatrix} \longrightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \sigma(x) \end{pmatrix} \quad (2.55)$$

As observed with the Higgs model, the extra degrees of freedom of the massless $\eta(x)$ bosons are unphysical. These extra degrees of freedom are “rotated away” by moving to the unitary gauge, through a gauge transformation that combines first an $SU(2)$ rotation which converts the isospinor into a down-isospinor, followed by a $U(1)$ transformation which makes the down isospinor real. All other fields in the Lagrangian transform accordingly, but being the SM Lagrangian $SU(2) \times U(1)$ invariant, this does not affect the equations of motion. In the following the fields are assumed to have been rotated and the same notation for the fields is kept.

Note that, because SSB occurs in the component of the isospinor that is electrically neutral, electric charge is conserved in the vacuum state, meaning that one symmetry survives. More precisely, under a global $SU(2) \times U(1)$ transformation the Higgs doublet transforms as $\Phi(x) \rightarrow \Phi'(x) = \exp[i(\alpha_i \tau_i / 2 + \beta Y)] \Psi(x)$. For the choice of $\alpha_1 = \alpha_2 = 0$ and $\alpha_3 = 2\beta$, one finds a gauge transformation that leaves the vacuum field invariant:

$$\phi_0 \rightarrow \phi_0 = \exp[i(2I_3^W + Y)\beta] \phi_0 = 1 \cdot \phi_0. \quad (2.56)$$

One can identify this new $U(1)$ gauge transformation as the electromagnetic interaction and the conserved charge as the electric charge $Q = I_3^W + \frac{1}{2}Y$.

Before SSB the SM is an $SU(3) \times SU(2)_L \times U(1)_Y$ gauge invariant theory. After SSB the $SU(2)_L \times U(1)_Y$ symmetry is broken, but a new $U(1)_{EM}$ symmetry appears. The $SU(3)$ symmetry stays unbroken because the Higgs field is not charged under $SU(3)$. In the following the scalar sector Lagrangian will be analyzed after SSB and rotation to the unitary gauge. From the covariant derivative term, it will be shown that the degrees of freedom of the Goldstone bosons have been eaten by three gauge bosons, which acquired a mass. These will be identified as being the W^+ , W^- , and Z bosons. The photon, on the other hand, will remain massless thanks to the surviving $U(1)_{EM}$ symmetry. Analysis of the term related to the Higgs potential will show that a new massive scalar boson appeared, the Higgs boson $\sigma(x)$. From now on, $\sigma(x)$ will be referred to as H . Lastly, it will be shown that the addition of a scalar field in the theory allows to introduce new gauge invariant terms to the SM Lagrangian to provide fermion masses.

The covariant term

Writing out the covariant term after SSB in the unitary gauge one obtains,

$$|D_\mu \Phi|^2 = \left| \left(\partial_\mu - ig_2 \frac{\tau_a}{2} W_\mu^a - ig_1 \frac{1}{2} B_\mu \right) \Phi \right|^2 \quad (2.57)$$

$$= \frac{1}{2} \left| \begin{pmatrix} \partial_\mu - \frac{i}{2}(g_2 W_\mu^3 + g_1 B_\mu) & -\frac{1}{2}g_2(W_\mu^1 - iW_\mu^2) \\ -\frac{1}{2}g_2(W_\mu^1 + iW_\mu^2) & \partial_\mu + \frac{i}{2}(g_2 W_\mu^3 - g_1 B_\mu) \end{pmatrix} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \right|^2 \quad (2.58)$$

$$= \frac{1}{2}(\partial_\mu H)^2 + \frac{1}{8}g_2^2(v + H)^2|W_\mu^1 + iW_\mu^2|^2 + \frac{1}{8}(v + H)^2|g_2 W_\mu^3 - g_1 B_\mu|^2. \quad (2.59)$$

The first term in the last line is the kinetic term for the Higgs field. The other terms can be divided into two groups, according to whether they pick out the v^2 term from the factor $(v + H)^2$ or not. In the following, it will be shown that the former group produces bilinear terms in the gauge fields that can be interpreted as mass terms for the gauge bosons. The second group provides interaction terms between the gauge bosons and the Higgs boson.

In order to obtain explicit mass terms the fields can be redefined as:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \quad Z_\mu = \frac{g_2 W_\mu^3 - g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}}, \quad A_\mu = \frac{g_2 W_\mu^3 + g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}}, \quad (2.60)$$

where two vector bosons W_μ^\pm with electric charge ± 1 are obtained from linear combinations of the first two components of the W_μ field, and two electrically neutral fields Z_μ and A_μ are written as linear combinations of the field B_μ and the third field component W_μ^3 . The fields Z_μ and A_μ are orthogonal to each other and are related to the original fields via the rotation matrix,

$$B_\mu(x) = -\sin \theta_W Z_\mu(x) + \cos \theta_W A_\mu(x) \quad (2.61)$$

$$W_\mu^3(x) = \cos \theta_W Z_\mu(x) + \sin \theta_W A_\mu(x). \quad (2.62)$$

The Lagrangian with the terms containing v^2 can then be rewritten as

$$\mathcal{L}_\Phi \supset \frac{1}{8}g_2^2 v^2 |W_\mu^1 + iW_\mu^2|^2 + \frac{1}{8}v^2 |g_2 W_\mu^3 - g_1 B_\mu|^2 \quad (2.63)$$

$$\rightarrow \left(\frac{1}{2}vg_2\right)^2 W_\mu^+ W^{-\mu} + \frac{1}{2}\left(\frac{1}{2}v\sqrt{g_2^2 + g_1^2}\right)^2 Z_\mu Z^\mu. \quad (2.64)$$

The mass terms can then be read off directly as,

$$M_W = \frac{1}{2}vg_2 \quad \text{and} \quad M_Z = \frac{1}{2}v\sqrt{g_2^2 + g_1^2}, \quad (2.65)$$

while the A_μ boson remains massless. One recognizes the W^+ , W^- , and Z bosons as the massive weak vector bosons, and the A_μ boson as the massless photon. The angle θ_W is the *weak mixing angle* (or Weinberg angle), which specifies the mixture of the electromagnetic and weak interaction.

The Higgs potential

The remaining part of the Higgs Lagrangian involves the potential $V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2$. Plugging in the Higgs doublet after SSB one gets,

$$V = \frac{\mu^2}{2}(v + H)^2 + \frac{\lambda}{4}(v + H)^4. \quad (2.66)$$

Including the kinetic term obtained from the covariant derivative and with the substitution $\mu^2 = -\lambda v^2$ the Higgs Lagrangian is given by

$$\mathcal{L}_H = \frac{1}{2}(\partial_\mu H)^2 - \lambda v^2 H^2 - \lambda v H^3 - \frac{\lambda}{4}H^4. \quad (2.67)$$

As expected, a new massive scalar boson, the Higgs boson, appeared in the theory with mass,

$$M_H = \sqrt{2\lambda v^2}. \quad (2.68)$$

The remaining terms represent the Higgs self-interactions with couplings⁹

$$g_{HHH} = (3!)i\lambda v = 3i\frac{M_H^2}{v}, \quad (2.69)$$

$$g_{HHHH} = (4!)i\frac{\lambda}{4} = 3i\frac{M_H^2}{v^2}. \quad (2.70)$$

The vacuum expectation value is fixed by the value of the W boson mass (measured via muon decay) through the relation in Eq. 2.65 and was measured to be $v = 246.22$ GeV. The Higgs boson mass is a free parameter in the SM, dependent only on the unknown parameter λ . The most precise measurement to date of the Higgs boson mass is $m_H = 125.11 \pm 0.11$ GeV [29]. The Higgs self-coupling is instead still unmeasured, although increasingly stringent limits are being set on its allowed value. Measurement of the Higgs self-coupling is one of the most pressing experimental goals and one of the physics motivation for the High-Luminosity LHC

⁹According to the Feynman rules, the couplings are given by the coupling term from the Lagrangian multiplied by a factor of $-i$ and by a factor $n!$, where n is the number of identical particles interacting at the vertex [17].

(HL-LHC) upgrade (see Sec. 3.1).

The Yukawa interactions

Fermion mass terms cannot be included *ad hoc* in the Lagrangian, as they would violate gauge invariance, and neither do they appear via the Higgs mechanism, like the gauge boson masses do. The introduction of a scalar field in the theory turns out to again be useful, as it provides a new way to add mass terms via new couplings. The fermions and Higgs fields are coupled through gauge invariant interactions, called *Yukawa interactions*. These occur with terms of the form $\bar{\psi}(x)\phi(x)\psi(x)$. The SM Lagrangian is augmented with the Yukawa Lagrangian given by

$$\mathcal{L}_{Yukawa} = -(Y_l)^{ij} \bar{L}_L^i \Phi l_R^j - (Y_d)^{ij} \bar{Q}_L^i \Phi d_R^j - (Y_u)^{ij} \bar{Q}_L^i \tilde{\Phi} u_R^j + h.c. \quad (2.71)$$

where $\tilde{\Phi} = i\tau_2 \Phi^*$ is the isodoublet with hypercharge $Y = -1$, the indices i and j run over each quark or lepton generation, and the matrices $Y_f (f = u, d, l)$ are general complex-valued matrices introduced to realize the couplings between the scalar and the fermion fields.

In the following, the new interactions are analyzed for the quarks, while generalization to the lepton case is straightforward. The following notation is used:

$$Q_L^i = \begin{bmatrix} u_L^i \\ d_L^i \end{bmatrix} = \left(\begin{bmatrix} u_L \\ d_L \end{bmatrix}, \begin{bmatrix} c_L \\ s_L \end{bmatrix}, \begin{bmatrix} t_L \\ b_L \end{bmatrix} \right) \quad (2.72)$$

For a fixed choice of quark flavor i and j , after SSB and rotation to the unitary gauge, the following terms appear:

$$\mathcal{L}_{Yukawa,q} = y_d^{ij} \bar{d}_L^i d_R^j (v + H) + y_u^{ij} \bar{u}_L^i u_R^j (v + H). \quad (2.73)$$

These look like candidates for fermion mass terms and fermion coupling terms to the Higgs field. However, the matrices Y are not diagonal, as there is no symmetry principle that requires them to be. This means that there are non-zero terms with $i \neq j$ that can mix fermion generations. Hence, the Yukawa interactions break the flavor symmetry of the Lagrangian. In order to obtain the physical masses and couplings observed in the laboratory, the matrices have to be diagonalized. This can be obtained via bi-unitary transformations of the form,

$$M_{\text{diag}}^q = V_L^{q\dagger} M^q V_R^q, \quad \text{where} \quad m_{ij} = y_{ij} \frac{v}{\sqrt{2}}. \quad (2.74)$$

Upon diagonalization only the terms \hat{y}_{ij} with $i = j$ survive. Looking at the case for $i = j = 3$,

corresponding to the up-type top quark and the down-type bottom quark,

$$\mathcal{L}_{\text{Yukawa, tb}} = -\frac{1}{\sqrt{2}}(\hat{y}_b \bar{b}_L b_R + \hat{y}_t \bar{t}_L t_R)(v + H) \quad (2.75)$$

$$= \underbrace{-\hat{y}_b \frac{v}{\sqrt{2}} \bar{b}_L b_R - \hat{y}_t \frac{v}{\sqrt{2}} \bar{t}_L t_R}_{\text{Mass terms}} - \underbrace{\hat{y}_b \frac{1}{\sqrt{2}} \bar{b}_L b_R - \hat{y}_t \frac{1}{\sqrt{2}} \bar{t}_L t_R}_{\text{Higgs couplings}}, \quad (2.76)$$

mass terms and new couplings between the fermions and the Higgs boson appear of the form,

$$m_f = \hat{y}_f \frac{v}{\sqrt{2}} \quad \text{and} \quad g_{Hff} = i \frac{m_f}{v}. \quad (2.77)$$

However, because the weak interaction mixes up- and down-fermions, the fermion couplings to the W boson, arising from the EW covariant derivative term, now contain off-diagonal elements:

$$\bar{\Psi} \not{D} \Psi \supset \frac{g_2}{\sqrt{2}} \bar{u}_L i \gamma_\mu d_L W^\mu \quad \longrightarrow \quad \frac{g_2}{\sqrt{2}} \bar{u}_L i \gamma_\mu (V_L^u V_L^{d\dagger}) d_L W^\mu \quad (2.78)$$

The three matrices M^u , M^d , and $V_{CKM} \equiv V_{uL} V_{dL}^\dagger$ cannot be diagonalized simultaneously, resulting in quark flavor violating interactions. The matrix V_{CKM} is the CKM matrix, named after Cabibbo [30], and Kobayashi and Maskawa [31]. It is a unitary 3×3 matrix parametrized by four parameters: three mixing angles θ_i and one phase δ . The phase δ is responsible for all CP-violating phenomena in the SM. Experimentally, the magnitude of all CKM has been found to be [32]

$$|V_{CKM}| = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97 & 0.22 & 0.004 \\ 0.22 & 0.97 & 0.04 \\ 0.009 & 0.04 & 0.999 \end{pmatrix} \quad (2.79)$$

consistent with the unitary assumption of the SM and with the interesting feature of being almost diagonal.

The Higgs couplings

The couplings of the Higgs boson with the fermions and vector bosons are obtained from the

interaction terms in the Lagrangian and are given by,

$$g_{Hff} = i\frac{m_f}{v}, \tag{2.80}$$

$$g_{HVV} = -2i\frac{M_V^2}{v}g_{\mu\nu}, \tag{2.81}$$

$$g_{HHVV} = -2i\frac{M_V^2}{v^2}g_{\mu\nu} \tag{2.82}$$

The tree level couplings of the Higgs boson to fermion mass eigenstates are flavor diagonal, CP conserving, and proportional to the mass of the fermion, making the coupling to the top-quark by far the largest. The couplings to the vector bosons are instead proportional to the square of the vector boson masses. Note that the Higgs boson does not couple at tree-level to the massless photon, nor to the gluon, as it does not carry color charge. However, these couplings appear at higher orders via loop corrections, where top-quark-induced loops provide the largest cross sections because of the larger g_{Htt} coupling.

All tree level couplings of the Higgs boson to SM particles are functions of only two parameters, either λ and μ , or v and m_H . Measurement of the couplings is therefore a direct test of the mechanism of spontaneous symmetry breaking. As shown in Fig. 2.3, the measurements of the fermions and gauge couplings have so far agreed extremely well with the predictions of the SM.

Although all fermion masses appear via the same mechanism of Yukawa interactions and their physics scale is set by the vev value of $v = 246 \text{ GeV}$, the observed masses span six orders of magnitude, from the top quark mass of $m_t \approx 175 \text{ GeV}$ down to the up-quark mass of $m_u \approx 2 - 5 \text{ MeV}$ and to the electron mass of $m_e \approx 0.5 \text{ MeV}$. Given that the structure of fermion masses originates solely from the Yukawa couplings, which are added as free parameters in the SM, the origin of their hierarchical structure is one of the most fundamental questions today.

2.8 Hints for physics beyond the Standard Model

The SM has proven to be a remarkably successful description of nature, whose structure was dictated by symmetries and guided by the experimental discoveries of the past century. However, symmetry arguments alone are not sufficient to explain the complex structure that experiments have brought to light, such as the non-general hierarchical structure in the Yukawa couplings and the naturalness problem, which will be discussed in the next section.

Important phenomena are also not accounted for, the foremost example being gravity, one of the four fundamental forces of nature. This makes physicists regard the SM as an

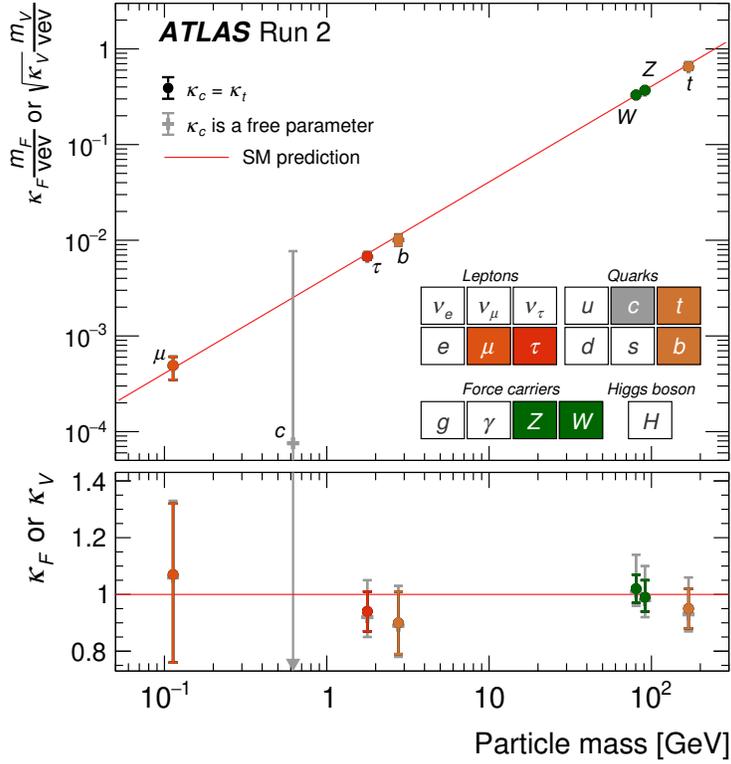


Figure 2.3: Reduced Higgs boson coupling strength modifiers and their uncertainties [33].

effective low energy theory, valid only up to a certain energy scale. This scale is generally expected to be smaller than the Planck energy scale $M_P \sim 10^{19}$ GeV, where the strength of the gravitational interaction is predicted to become comparable to the other forces.

Astrophysical observations, including galactic rotation speeds [34] and gravitational lensing (the curvature of space-time near gravitating mass) [35], indicate the existence of massive matter that seems not to interact electromagnetically with the SM particles. Because it cannot be detected directly, the presence of this *dark matter* is inferred from its gravitational pull on ordinary matter. Dark matter is estimated to represent $\sim 80\%$ of all matter in the Universe, but its origin and nature remain unknown.

While the SM assumes massless neutrinos, the observation of neutrino oscillations [36, 37] has proven that neutrinos must have mass, albeit a very small one. While being six orders of magnitude lighter than the electron and 10^{12} lighter than the top quark, the masses of the three neutrino flavors display themselves a significant hierarchy and their origin is still unknown. Whether neutrino mixing arises from a different mechanism, whether they are Majorana or Dirac particles, or whether they couple to some non SM interaction, their mass points to some BSM physics.

Lastly, some tensions with the SM predictions have started to arise, including the most recent B-physics anomalies [38] and the measurement of the anomalous magnetic moment of the muon (Muon $g - 2$) [39].

All these questions call for BSM physics. Thanks to its special role in the theory, the Higgs boson is at the center of many of these questions. In the next section, the importance of the Higgs sector for BSM physics scenarios will be discussed, with particular focus for the topics relevant for this thesis.

Chapter 3

The Higgs boson as a portal to new physics

The fundamental role of the Higgs boson in the SM model makes many consider it the key to explain several open questions in particle physics.

Most of the observed unexplained structure brought to light by experimental observations is in fact connected to the Higgs sector. While “a SM” has the structure described in the previous section, “the SM” is an empirical model with 19 free parameters¹, whose values are set by experimental measurements [28]. Of these, four parameters – three gauge couplings and the weak mixing angle – arise from the gauge sector of the theory, while the remaining fifteen parameters – six and three from the Yukawa couplings of quarks and charged leptons, four from the CKM matrix, and two from the Higgs potential – arise from the Higgs sector.

The scalar sector of the SM remains greatly unexplored experimentally, as the Higgs self-coupling and the shape of the Higgs potential have still not been measured. These have important implications for some fundamental questions in cosmology. Sensitivity to di-Higgs production with the HL-LHC might help shed light on these open questions, as discussed further in Chap. 8 in relation to the HL-LHC trigger upgrade.

Additionally, there is much that is not well understood about the Higgs boson itself. One issue related to the Higgs boson is the naturalness problem, which sees unnatural mathematical cancellations arise in the theory due to its scalar nature. In order to remove this naturalness, some models predict the existence of new gauge vector bosons. These models are often studied via the generalized Heavy Vector Triplet (HVT) framework. The question remains also of whether there is only one Higgs boson or if there might be an extended scalar sector. An important theoretical framework used to study these questions is the Two-Higgs-Doublet Model. These models are relevant for the analysis discussed in Chap. 7.

¹Assuming three generations and massless neutrinos

3.1 The Higgs self-coupling

The Higgs potential is fundamentally connected to the origin of electroweak symmetry breaking (EWSB), but while the vev and the Higgs boson mass have been measured with high precision, the Higgs self-coupling λ remains unmeasured. Measurement of the Higgs self-coupling would help shed light on the shape of the potential, which makes it relevant for several open questions in cosmology, including the stability of the Universe and the observed baryon asymmetry.

Quantum corrections are observed to affect the shape of the Higgs potential [40]. The measured values of the Higgs and top quark masses indicate that, when running the Higgs self-coupling to high renormalization scales, λ turns negative at a scale $\Lambda \sim 10^{10}$ GeV [41]. This indicates that the vacuum state of our Universe is not the absolute minimum and that a non-zero probability exists for quantum fluctuations to cause the decay of the Universe into a lower energy state. While the lifetime of the Universe is orders of magnitude greater than its current age, making its metastability not an issue for the survival of humanity, it is nonetheless puzzling. Is the puzzlement only due to our anthropocentric view or is there some BSM physics missing in the theory that would stabilize the vacuum?

The Higgs self-coupling is related to another important question concerning the origin of the matter-antimatter asymmetry in the Universe. In order for the Big Bang nucleosynthesis to have occurred the matter-antimatter composition of the Universe had to be already asymmetric to prevent annihilation between nucleons and antinucleons. Models of electroweak baryogenesis [5] provide a mechanism for the observed baryon asymmetry that would have occurred during the electroweak phase transition, which is the process by which the Higgs field acquired a vev. The Universe after the Big Bang is thought to have started in the unbroken phase, where the $SU(2)_L \times U(1)_Y$ gauge invariance was manifest. As the temperature cooled down below $T \lesssim 100$ GeV, the Higgs field settled into one of the absolute minima of the potential, spontaneously breaking the original symmetry. Electroweak baryogenesis is predicted to have taken place during this phase transition. However, for baryon creation to take place successfully, the transition has to be first order, where the departure from thermal equilibrium is violent, while the SM predicts the electroweak phase transition to be of second order, with a smooth crossover between the two phases as the temperature decreases. Any model of electroweak baryogenesis requires therefore physics beyond the SM to make the transition first order.

Probing the Higgs self-coupling would help shed light on these fundamental questions. The only direct probe at colliders is the measurement of di-Higgs production (indirect constraints can be obtained from single-Higgs production). The HL-LHC is expected to provide

sufficient sensitivity for the ATLAS and CMS Collaborations to measure SM di-Higgs production and the Higgs self-coupling. The experimental challenges related to this measurement will be discussed in more detail in Chap. 8.

3.2 Naturalness

The inability to include a gauge theory of gravity makes physicists regard the SM as an effective low energy theory valid up to the Planck scale. However, some theoretical reasons exist to believe that the SM might break down at much lower energies, related to the presence in the theory of a fundamental scalar particle.

In a quantum field theory, any scalar particle inevitably leads to ultraviolet divergences in the radiative corrections to its mass. The Feynman diagrams contributing to the one loop corrections to the Higgs boson mass are shown in Fig. 3.1. The divergent integrals can be

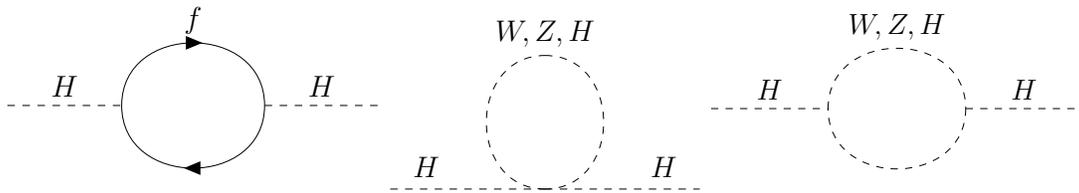


Figure 3.1: Feynman diagrams contributing to the one-loop corrections to the Higgs boson mass in the SM.

regularized by cutting off the loop integral momenta at a scale Λ . The theory can then be renormalized by expressing the mass of the physical particle in terms of the mass of the bare particle, so that the infinities only appear in the relation between the physical and bare mass, but the physical observable remains finite. Keeping only the dominant contributions, the resulting physical Higgs boson mass in the renormalized theory is given by:

$$m_H^2 = (m_H^0)^2 + \frac{3\Lambda^2}{3\pi^2 v^2} [M_H^2 + 2M_W^2 + M_Z^2 - 4m_t^2], \quad (3.1)$$

where m_H^0 is the bare mass from the unrenormalized Lagrangian. The quadratic, rather than logarithmic, divergence as a function of Λ in the counter-term is unique in the SM and it is due to the Higgs boson being a scalar field. If the theory is considered valid up to the Planck scale $\Lambda \sim 10^{19}$ GeV, a finely-tuned cancellation of 34 digits between the m_H^0 term and the counter-term proportional to Λ^2 would be necessary to obtain the observed renormalized mass square m_H^2 of $(\sim 10^2 \text{ GeV})^2$ [41]. This type of cancellation is considered unnatural and is referred to as the *naturalness* problem. This raises the question of whether there is

some larger symmetry or some new dynamics at work to protect the Higgs from these large radiative corrections. One way in which the fine-tuning would be removed or reduced is if new particles existed with masses around the TeV scale and coupling to the Higgs boson. Several BSM models, partly motivated by naturalness arguments, predict the existence of such new heavy resonances and are often studied within the framework of the HVT model.

3.3 The Heavy Vector Triplet model

New vector bosons are a common element of BSM models with an extended gauge symmetry group, where they appear as the gauge bosons of the new broken symmetries. Requirement of gauge invariance under the SM $SU(3)_C \times SU_L(2) \times U(1)_Y$ in the non-broken phase strongly constraints the quantum numbers and allowed interactions of the new vector bosons. Isospin triplets are particularly interesting as, experimentally, they can give rise to sizable resonant signals [42, 43] and, from a theoretical point of view, they appear in well-known extensions of the SM, including Little Higgs [44] models and composite Higgs models [45, 46].

While these models are theoretically consistent, it is hard to pin down specific observable predictions that would differentiate one model from another. Within a given model framework, different assumptions can also determine different phenomenologies. Tailoring a search for each model is unfeasible. However, resonant searches are generally not sensitive to all the free parameters of a model, but only to the mass and couplings of the predicted new particles, which determine the available decay channels, and the strength and location of the mass resonance. The HVT [3, 42] model provides a simplified framework with which one can test only the relevant phenomenological parameters: the experimental search determines the likelihood between the data and the general model; the phenomenological parameters can then be expressed analytically in terms of the parameters of the explicit theory. Note that the model assumes *on-shell* resonance production and decay.

The HVT model is based on a simplified Lagrangian, which, in addition to the SM fields, includes a new real vector $V_\mu^a (a = 1, 2, 3)$ charged under $SU(2)_L$ and with zero hypercharge, with the charge eigenstates

$$V_\mu^\pm = \frac{V_\mu^1 \mp iV_\mu^2}{\sqrt{2}}, \quad V_\mu^0 = V_\mu^3. \quad (3.2)$$

The Lagrangian describing the new fields and their interactions with SM particles is,

$$\mathcal{L}_V = -\frac{1}{4}D_{[\mu}V_{\nu]}^a D^{[\mu}V^{\nu]a} + \frac{m_V^2}{2}V_\mu^1 V^{\mu a} \quad (3.3)$$

$$+ ig_V c_H V_\mu^a H^\dagger \tau^a \overleftrightarrow{D}^\mu H + \frac{g^2}{g_V} c_F V_\mu^a J_F^{\mu a} \quad (3.4)$$

$$+ \frac{g_V}{2} c_{VVV} \epsilon_{abc} V_\mu^a V_\nu^b D^{[\mu}V^{\nu]c} + g_V^2 C_{VVHH} V_\mu^a V^{\mu a} H^\dagger H - \frac{g}{2} c_{VW} \epsilon_{abc} W^{\mu\nu a} V_\mu^b V_\nu^c. \quad (3.5)$$

where ϵ_{abc} is the Levi-Civita symbol. The first line contains the kinetic and mass terms of the new V bosons, plus trilinear and quadrilinear interactions with the SM vector bosons arising from the covariant derivatives,

$$D_{[\mu}V_{\nu]}^a = D_\mu V_\nu^a - D_{[\nu}V_{\mu]}^a, \quad D_\mu V_\nu^a = \partial_\mu V_\nu^a + g\epsilon^{abc}W_\mu^b W_\nu^c \quad (3.6)$$

where g is the $SU(2)_L$ gauge coupling. The second line of the equation contains the interactions of V with the Higgs boson and with the SM left-handed fermions,

$$iH^\dagger \tau^a \overleftrightarrow{D}^\mu H = iH^\dagger \tau^a D^\mu H - iD^\mu H^\dagger \tau^a H, \quad J_F^{\mu, a} = \sum_f \bar{f}_L \gamma^\mu \tau^a f_L, \quad (3.7)$$

where $\tau^a = \sigma^a/2$ and σ^a are the Pauli matrices. The last line contains vertices representing bosonic interactions. However, to first approximation, these interactions do not contribute to LHC phenomenology [3], so can be disregarded. All couplings are weighted by a new parameter g_V , which represents the typical strength of V interactions. The c coefficients are dimensionless parameters parametrizing the departure from the typical size.

Upon EWSB, the components of the new vector triplet mix with the SM gauge bosons. After diagonalization of the mass matrices [3, 42], expressions for the physical masses of the SM W and Z bosons and the new charged and neutral vector bosons, referred to as W' and Z' , can be obtained. In order to preserve custodial symmetry and the SM tree-level value of $\rho = 1$, the W' and Z' bosons are quasi-degenerate and their masses are assumed to be above ≈ 1 TeV. Thanks to the resulting mass hierarchy between the SM and the new gauge bosons, the mixing angles are naturally small and the SM couplings of the W and Z bosons are automatically close to the SM expectation. In general, the W' and Z' bosons are assumed to be degenerate and the data is interpreted in terms of one effective resonance with mass M_V .

The small mixing angles simplify the couplings to fermions, which are determined by the parameter combination $g_F = \frac{g^2}{g_V} c_F$. The parameter c_F controls therefore Drell-Yan

production and the fermionic decays of the new bosons. Here the coupling to fermions is assumed universal, but it could in principle be split into different couplings for leptons, and light and heavy quarks.

The coupling to the SM bosons is more subtle. Because of the small mixing angles, the couplings involving transversely polarized SM vectors are suppressed. However, via a different choice of gauge it can be shown [3] that direct couplings to the longitudinal components of the gauge bosons exist. After the change of basis, the couplings are given by,

$$\mathcal{L}_\pi \supset \frac{g_V c_H}{2} V_\mu^a (\partial^\mu h \pi^a - h \partial^\mu \pi^a + \epsilon^{abc} \pi^b \partial^\mu \pi^c) \quad (3.8)$$

where π_\pm and π_0 are the Goldstone bosons that reappear in this basis and that, by the *Equivalence Theorem*, correspond to the longitudinal W^\pm and Z bosons. Note that all the couplings are controlled by the same parameter combination $g_V c_H$. Therefore, c_H controls both the interaction with the Higgs boson and with the SM weak bosons and, in particular, the resonance production via vector boson fusion (VBF) and the decay into bosonic channels.

To a good approximation, the HVT phenomenology is completely described by the coupling to fermions $g_F = g^2/g_V c_F$, the coupling to bosons $g_H = g_V c_H$, and the mass of the resonance M_V . In order to test the broad phenomenological phase space, two benchmark scenarios are often studied for which the values of c_H and c_F are fixed, while scanning different “benchmark points” in the phase space traced by the parameters M_V and g_V . The model described in Ref. [47] is taken as representative of a weakly coupled model, where the new triplet appears upon SSB of an extended gauge symmetry, and will be referred to as *Model A*. For this type of model, only relatively small values $g_V \lesssim 3$ are considered, predicting comparable branching ratios into bosons and fermions. A generic Composite Higgs Model [48] is taken as representative of a strongly coupled model, referred to as *Model B*, where larger values $g_V \gtrsim 3$ are studied. For large g_V values, the coupling to fermions g_F is suppressed by g^2/g_V . The coupling to bosons g_H scales instead as g_V . Strongly coupled models predict therefore dominant branching ratios into diboson final states, while fermionic channels are suppressed. For Model B, the total width increases with increasing g_V and for values $g_V \geq 8$ the resonance becomes very broad $\Gamma/M \gg 0.1$. These values are therefore not considered, as the model is only valid for narrow resonances. For both models A and B the dominant production mechanism is Drell-Yan production. VBF production can be enhanced by suppressing the coupling of the HVT bosons to fermions. This is done in *Model C*, where $g_F = 0$ and $g_H = 1$.

3.4 The Two-Higgs-Doublet Model

One of the most stringent constraints on the SM are electroweak precision measurements, but while the value of the ρ parameter places stringent requirements on the scalar sector, it would in principle accommodate any number of scalar singlets and doublets in the theory [4]. Since the SM assumes the simplest possible scalar structure by introducing only one Higgs doublet, the question arises of whether the Higgs boson is not alone. Several examples of models with extended scalar sectors exist, including the Minimal Supersymmetric Standard Model (MSSM) [49], axions models [50], and baryogenesis models [51]. In particular, one important class of models, called the Two-Higgs-Doublet Models (2HDMs) [4] studies the addition of a new scalar doublet ².

Flavor Changing Neutral Currents

The introduction of two Higgs doublets in the Yukawa Lagrangian allows for flavor changing neutral currents (FCNCs) at tree level. Considering the quark terms only, $\mathcal{L}_{\text{Yukawa}}$ is now:

$$\mathcal{L}_{\text{Yukawa}} \supset - \sum_{k=1,2} (Y_d)^{ij,k} \bar{Q}_L^i \Phi_k d_R^j - (Y_u)^{ij,k} \bar{Q}_L^i \tilde{\Phi}_k u_R^j + h.c., \quad (3.9)$$

where the i and j quarks couple to a linear combination of the two scalar fields $\Phi_k (k = 1, 2)$. Consider the case of the down-type quarks. Upon SSB the mass terms appear as:

$$\mathcal{L}_{\text{Yukawa}} \supset - \sum_{k=1,2} \bar{d}_L^i \underbrace{(Y_d)^{ij,k} \frac{v^k}{\sqrt{2}}}_{M_{ij}^d} d_R^j, \quad M_{ij}^d = y_d^{ij,1} \frac{v_1}{\sqrt{2}} + y_d^{ij,2} \frac{v_2}{\sqrt{2}}. \quad (3.10)$$

Without further restrictions, the coupling matrices Y_d^1 and Y_d^2 and the mass matrix M_{ij}^d are not simultaneously diagonalizable, making the Yukawa couplings not flavor diagonal.

FCNCs are highly constrained by experiment and, if they exist, would have to be extremely small. For this reason, in the study of 2HDMs, a discrete Z_2 symmetry is generally introduced *ad hoc* to suppress FCNCs. As formalized by the Paschos-Glashow-Weinberg theorem [52, 53], the condition that all fermions with the same quantum number (the only ones that can mix) couple to the same Higgs doublet is necessary and sufficient for the absence of FCNCs at tree level. In the SM, this can be obtained in different ways. In the *Type I* 2HDM, all fermions couple only to one of the doublets, conventionally chosen to be Φ_2 .

²The MSSM is a special case of the Type II 2HDM described below, but the description given here is for the most general 2HDM following Ref. [4]

For instance, this can be enforced by requiring the discrete symmetry $\Phi_1 \rightarrow -\Phi_1$. In the *Type II* 2HDM, all up-type quarks couple to one Higgs doublet and all down-type quarks and charged leptons couple to the other one. In the *lepton specific* model, the couplings to quarks are the same as in the Type I model, while the couplings to charged leptons are as in the Type II. The *flipped* model has the same couplings to quarks as in the Type II model, and to charged leptons as in the Type I model [54].

The potential

With two scalar doublets, the most general scalar potential becomes quite complex, determined by 14 parameters and with various minima with different charge and CP conservation properties. For this reason, several simplifying assumptions are usually made for phenomenological studies. The most general CP-conserving potential with a softly broken Z_2 symmetry of two Higgs doublet fields Φ_1 and Φ_2 with hypercharge +1 is given by

$$V = m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 (\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1) + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \frac{\lambda_5}{2} \left[(\Phi_1^\dagger \Phi_2)^2 + (\Phi_2^\dagger \Phi_1)^2 \right], \quad (3.11)$$

where all the parameters are real. Each doublet has four degrees of freedom, for a total of eight fields:

$$\Phi_a = \begin{pmatrix} \phi_a^+ \\ (v_a + \rho_a + i\eta_a)/\sqrt{2} \end{pmatrix}, \quad a = 1, 2. \quad (3.12)$$

Upon SSB, the neutral components of the two doublets acquire vevs, v_1 and v_2 :

$$\langle \Phi_1 \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \quad \langle \Phi_2 \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_2 \end{pmatrix}. \quad (3.13)$$

where the observed vev value requires $v^2 = v_1^2 + v_2^2 = (246 \text{ GeV})^2$. The contributions of each doublet to the observed SM vev is parametrized by the angle β , as $\tan \beta = \frac{v_2}{v_1}$. The physical

mass eigenstates are obtained by the rotation matrices,

$$\begin{pmatrix} G^0 \\ A \end{pmatrix} = \begin{pmatrix} \cos \beta & \sin \beta \\ \sin \beta & -\cos \beta \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}, \quad (3.14)$$

$$\begin{pmatrix} h \\ H \end{pmatrix} = \begin{pmatrix} \sin \alpha & -\cos \alpha \\ -\cos \alpha & -\sin \alpha \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}, \quad (3.15)$$

$$\begin{pmatrix} G^\pm \\ H^\pm \end{pmatrix} = \begin{pmatrix} \cos \beta & \sin \beta \\ \sin \beta & -\sin \beta \end{pmatrix} \begin{pmatrix} \phi^\pm \\ \phi^\pm \end{pmatrix}, \quad (3.16)$$

where the angle β reappears, together with the parameter α , as the mixing angles of the mass matrices. The Higgs mechanism proceeds then as in the SM. Three Goldstone bosons (G^\pm and G^0) are “eaten” by the W^\pm and Z bosons, which subsequently acquire mass. The remaining five physical degrees of freedom correspond to five massive scalar fields: two are charged (H^\pm), two are neutral and CP-even (h and H , with $m_h < m_H$), and one is neutral and CP-odd (A). With the simplifying assumptions mentioned before, the 2HDM is fully determined by seven parameters: the masses of the five Higgs bosons and the two angles α and β .

The Standard Model Higgs and the alignment limit

The observed Higgs boson with a mass of 125 GeV and its measured SM couplings put stringent constraints on the phenomenology of the neutral scalars, requiring the mass eigenbasis of the two neutral scalars to lie very close to the Higgs basis. The Higgs basis is the basis where one of the two doublets is entirely responsible for the SM vev and is obtained with the following field redefinition:

$$\begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \begin{pmatrix} \cos & \sin \\ -\sin & \cos \end{pmatrix} \begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix}. \quad (3.17)$$

From the relation $\tan \beta = \frac{v_2}{v_1}$, one gets $\langle H_1 \rangle_0 = (v_1 \cos \beta + v_2 \sin \beta)/\sqrt{2} = v$ and $\langle H_2 \rangle_0 = 0$. In this basis, the physical neutral CP-even states are given by

$$H = (\sqrt{2} \text{Re} H_1^0 - v) \cos(\beta - \alpha) - \sqrt{2} \text{Re} H_2^0 \sin(\beta - \alpha), \quad (3.18)$$

$$h = (\sqrt{2} \text{Re} H_1^0 - v) \sin(\beta - \alpha) - \sqrt{2} \text{Re} H_2^0 \cos(\beta - \alpha). \quad (3.19)$$

The angle $(\beta - \alpha)$ characterizes the mixing of the neutral scalars. A SM-like Higgs boson exists if $(\sqrt{2} \text{Re} H_1^0 - v)$ is an approximate mass eigenstate. This occurs if there is negligible mixing between H_1^0 and H_2^0 , which is the case when $\cos(\beta - \alpha) = 0$ ($H^{SM} = h$) or when

$\sin(\beta - \alpha) = 0$ ($H^{SM} = H$). This is called the *alignment limit*. Alignment of the Higgs basis with the mass basis can also be obtained in the *decoupling limit*, when the new Higgs fields are assumed to be all much heavier than the SM-like neutral scalar field h . As shown in Ref. [55], once the heavier particles are integrated out, the low energy effective field theory is equivalent to the SM Higgs sector with one scalar doublet. The decoupling limit implies the condition $|\cos(\beta - \alpha)| \ll 1$. However, the latter condition is more general than the decoupling limit and can be obtained even for all Higgs boson masses $\leq \mathcal{O}(v)$. Therefore, alignment occurs automatically in the decoupling limit, but it is also possible without decoupling. In the decoupling limit, the light neutral scalar h is indistinguishable from the SM Higgs boson. On the other hand, in the alignment limit, even if the tree-level couplings of h are SM-like, deviations can appear when higher order corrections are included[55].

Chapter 4

The LHC and the ATLAS experiment

4.1 The Large Hadron Collider (LHC)

Accelerator development has historically been driven by particle physics research. The birth of collider physics can be traced back to 1932 when Lawrence's cyclotron, based on the principle of resonance acceleration, was able to produce 1.25 MeV protons to disintegrate the atom [56]. Since then, the need to look at increasingly smaller distances has meant continuously finding new ways to produce higher energy beams. By the 1980s, it was shared expectation that new physics discoveries (including the observation of at least one Higgs particle, of the top-quark, and of possible new physics phenomena such as supersymmetry or new gauge bosons [57]) should appear at substantially higher energies than ever tested before, in the range up to 1 TeV, and several options were considered [58]. The first serious investigation of the possibility of a hadron collider was produced in 1987 by the Long-Range Planning Committee, setup at the European Organization for Nuclear Research (CERN) in 1985 and chaired by Carlo Rubbia, which recommended that a proton-proton collider with a center-of-mass (CoM) energy of 13 – 15 TeV should be the next major project at CERN [57, 58]. Thus, the Large Electron Positron collider (LEP) [59] tunnel at CERN was designed from the start to provide enough space to later install superconducting magnets for a large proton collider [58]. The tunnel was constructed between 1984 and 1989. The LEP project started in 1989 and in 2000 it was terminated and installation of the Large Hadron Collider (LHC) began [60]. Today, the LHC is the largest and most powerful particle accelerator in the world, where the same principle of resonant acceleration used in 1932 is used to accelerate protons to 7 TeV.

4.1.1 Overview

The LHC [60, 61] is a 27 km proton-proton (pp) circular collider located at CERN, between 45 and 170 m under the surface on the border between France and Switzerland just outside of Geneva. Two counter-rotating beams collide at four interaction points (IP) where the four main experiments are located: ATLAS [62] and CMS [63], both multi-purpose experiments

designed to be sensitive to a wide range of SM processes and new physics searches; LHCb [64], a B-physics experiment; and ALICE [65], a heavy ion dedicated experiment for Pb-Pb operation, which will not be discussed in this work. The ring consists of eight 2.9 km long arcs with superconducting dipole magnets, alternating to eight straight sections of 210 m on either side of eight potential collision points where RF cavities are located¹. Being a particle-particle collider, the beams require opposite magnetic dipole fields. For this reason, the accelerator consists of two rings with separate magnetic fields and vacuum chambers in the main arcs, and common sections in the intersection regions (IR) around the IPs, where the beams share a common beam pipe approximately 130 m long. Because of the limited space in the tunnel, not enough room was available for two separate rings of magnets. For this reason, the twin-bore magnet design proposed by Blewett in 1971 [66] was adopted, which consists of two sets of coils and beam pipes within the same mechanical structure and cryostat, making the two magnets magnetically and mechanically coupled.

4.1.2 The accelerator complex

The accelerator chain

Before entering the LHC rings, the energy of the protons is increased by a series of smaller accelerators, each boosting the particles to the maximum allowed speed before injecting them into the next machine in the chain. The LHC accelerator chain [67] is shown in Fig. 4.1. The chain starts with a container of negative hydrogen ions (H^- , a hydrogen atom with two electrons). The linear accelerator Linac 4, which replaced Linac 2 in 2020, boosts the ions to 160 MeV and strips them of their two electrons during injection into the Proton Synchrotron Booster (PSB). The protons are then accelerated to 2 GeV before entering the Proton Synchrotron (PS), which brings the beam to 26 GeV and injects it into the Super Proton Synchrotron (SPS). Here, the protons are accelerated up to 450 GeV before finally entering the LHC rings.

The protons enter the LHC in bunches separated by 25 ns, taking about 4 minutes to fill the entire LHC ring. The protons are accelerated via a total of eight Radio Frequency (RF) cavities [69] per beam, each delivering 2 MV longitudinal voltage at 400 MHz. Twenty minutes later, after passing through the RF cavities 10 million times, the protons reach their maximum energy of 6.5 TeV, resulting in a CoM energy at collision point of $\sqrt{s} = 13$ TeV. After collision, the beams are reconstituted, and the process continues for approximately 10 hours until the beam is depleted of protons and is ready to be dumped. At this point,

¹The number of cavities was originally designed to compensate for electron synchrotron radiation losses at LEP, 10^{13} times that of a proton.

The CERN accelerator complex Complexe des accélérateurs du CERN

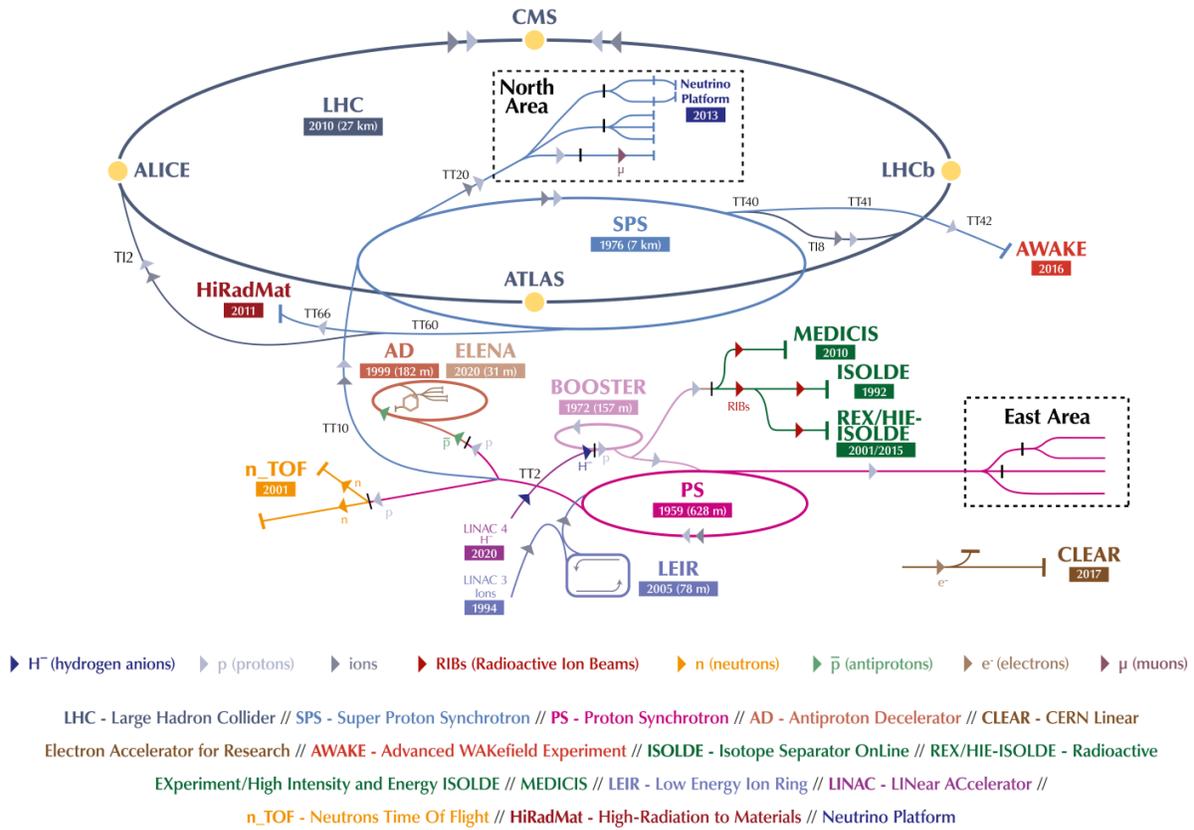


Figure 4.1: The LHC accelerator complex layout as of 2022 [68].

protons exit the LHC rings and travel along a straight line until collision with a block of concrete and graphite.

The RF system

The RF cavities are straight metallic chambers containing a longitudinal electromagnetic field and housed in cryomodules to operate at 4.5 K. In order to accelerate particles along a closed path, an oscillating voltage is necessary, as a DC voltage would cancel its accelerating effect over a full turn. For a particle to always see an accelerating voltage at the gap, the frequency of the voltage oscillation f_{RF} has to always be an integer multiple of the particle revolution frequency f_{rev} . Once the beam has reached the required energy, the ideally timed particle whose revolution frequency is identical to the RF frequency — the synchronous particle — will see a zero accelerating voltage every time it passes through the cavity. Any other particle with slightly different momenta will oscillate around the synchronous particle

along the longitudinal plane in what is called the *synchrotron motion* [70]. The result is that the particles get grouped into *bunches* around the synchronous particle in the bunch. The boundary of the bunch is called the *RF bucket*. Provided the energy deviations are not too large, the particles remain trapped in the bucket, which essentially acts as a potential well. The bunch structure is formed as soon as the RF system is on and the number of possible bunch crossings (BC) is fixed to f_{RF}/f_{rev} . In fact, the PS and SPS are also synchrotrons, and it is the PS that first determines the 25 ns bunch spacing.

The magnet system

Along the LHC rings, 10,000 superconducting magnets of about 50 different types are used to send the protons along the circular path. A nominal magnetic field of 8.3 T in the 1232 main dipoles is necessary to bend the path of the charged particles traveling close to the speed of light. This field is much higher than any other superconducting accelerator ever built before, requiring superfluid Helium at 1.9 K [71].

Particles in a bunch will occupy different positions on the transverse plane perpendicular to their trajectory. Displacements along the horizontal direction will simply cause the particles to follow different closed paths along the LHC circumference. The vertical plane, however, is unstable and the trajectory of particles with different initial conditions can end up spiraling towards the center. Combinations of 392 focusing and de-focusing quadrupole magnets are used to keep the beam stable along both horizontal and vertical axes. The resulting transverse oscillations on the horizontal and vertical planes are called *betatron oscillations* and the envelope function within which the particles oscillate is called the β -function [70]. The quadrupoles are also used to squeeze the beam and increase the beam luminosity. In particular, eight sets of low- β quadrupoles, called the *inner triplets*, are used at the intersection regions of the four experiments to make the beams narrower before collision, going from 0.2 mm across down to 16 μm .

4.1.3 LHC performance and operation

The bunch structure

The proton bunches have an elongated shape of about 7.48 cm along the longitudinal direction due to the synchrotron motion, and of $16 \times 16 \mu\text{m}$ in the transverse plane due to the betatron oscillations. The bunches are separated by 25 ns, or 7.5 m, giving a collision rate of 40 MHz and a maximum number of bunches in the ring of 3564. However, not all bunches are filled with protons, as empty bunches are necessary for, e.g., new bunch insertion when depleted bunches are dumped or for the abort gap needed to turn on the magnets to divert and dump the beam. This brings the effective number down to 2808. Each possible BC is

assigned a Bunch Crossing Identifier (BCID) from 0 to 3563. According to the LHC *filling scheme* set at the beginning of an LHC fill, each BC can have either two bunches colliding, one bunch, or be empty of protons.

Cross section

The probability of a given collision event to occur is expressed by the cross section σ , which is measured in units of squared-area. Because particle physicists are generally interested in rare events, σ is often more conveniently expressed in barns, where $1b = 10^{-24}\text{cm}^2$. The cross section for pp collisions at 7 TeV is approximately 110 mb, of which 60 mb are due to inelastic processes. Other contributions come from diffractive and elastic scattering events, which do not reach high enough energies in the transverse plane to be seen by the detectors. For a given process with cross section σ_{process} , the event rate in an LHC collision is,

$$\frac{dN_{\text{process}}}{dt} = L_I \sigma_{\text{process}}, \quad (4.1)$$

where L_I is the instantaneous luminosity provided by the machine, which is $L_I \approx 10^{34}\text{cm}^{-2}\text{s}^{-1}$ at the LHC. The cross section for di-Higgs production at $\sqrt{s} = 13$ TeV via gluon-gluon fusion, which is by far the largest production mode, is 31.05 fb [9]. This means that there are 31.05×10^{-5} events/second where two Higgs bosons are being produced or, in other words, a pair of Higgs bosons is produced every 53 minutes. This type of events is extremely rare when compared to the 11-14 orders of magnitude larger number of events that will have occurred during the same amount of time, as shown in Fig. 4.2. In order to draw statistically meaningful conclusions, it is crucial to produce enough of them. For this reason, as Eq. (4.1) shows, the most important parameter of an accelerator is the luminosity.

Luminosity

The luminosity is the number of collisions produced in a detector per squared-centimeter and per second, and it is dependent only on the beam parameters. Assuming two identical Gaussian beams colliding, the instantaneous luminosity is given by [60],

$$\mathcal{L}_I = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \sigma_x \sigma_y} \mathcal{F}, \quad (4.2)$$

where: N_b is the number of particles per bunch; n_b is the number of bunches per beam; γ_r is the relativistic gamma factor; f_{rev} is the LHC revolution frequency; σ_x and σ_y are the RMS cross-sectional size of the bunch in the x and y directions, which, in terms of the β -function and emittance ϵ , are given by $\sigma_i \approx \sqrt{\beta_i \epsilon}$; \mathcal{F} is the geometric luminosity reduction factor due

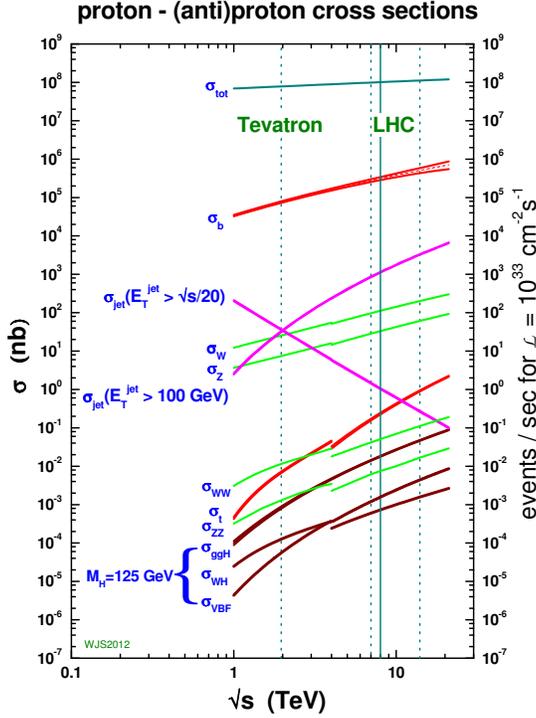


Figure 4.2: Standard Model cross sections as a function of collider energy [72].

to the crossing angle at the IP².

The discovery reach of the LHC ultimately depends on the total integrated luminosity, $L = \int L_I dt$, related to the total number of events of a given process as $N_{\text{process}} = L\sigma_{\text{process}}$. The total integrated luminosity has units of $[\text{cm}^{-2}]$. A precise knowledge of the luminosity is necessary to extract the visible cross section in any detector. The luminosity is measured by the experiments with specific detector sub-systems that are calibrated during special runs called van-der-Meer beam-separation scans [73].

To increase the number of rare events produced at the LHC requires therefore increasing the beam energy and intensity [60]. The maximum beam energy that can be attained is limited by the dipole magnetic field and the collider length. The nominal field of 8.33 T corresponds to a maximum beam energy of 7 TeV. The collision rate can be maximized by optimizing other parameters in Eq. (4.2). The beam intensity depends on the number of

²The geometric luminosity reduction factor is given by

$$\mathcal{F} = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*} \right)^2 \right)^{-1/2},$$

with θ_c the crossing angle, σ_z the RMS bunch length, and σ^* the transverse RMS beam size at the IP.

particles per bunch and on the size of the beam. The former is limited by several factors, such as beam-beam effects and collective beam instabilities, caused by the interaction of the protons with each other and with the vacuum chamber. The nominal and ultimate values of the number of protons per bunch are 1.15×10^{11} and 1.70×10^{11} , respectively [67]. The transverse beam dimensions can be optimized by improving the beam quality in terms of the *emittance* ϵ and the amplitude function β . The emittance is a measure of the spread of the beam, and the lower the emittance the closer the particles are together in distance and momentum. The emittance depends solely on the initial conditions set by the injection chain. The β -function can be adjusted during a run via the quadrupole magnets. In particular, the β^* parameter determines the transverse beam size at the IP, and the smaller it is, the larger the luminosity.

Several LHC upgrades have already been performed and others are planned in the incoming decade to bring the accelerator parameters to the ultimate design goals and beyond. However, with a higher collision rate come other challenges, including a higher probability of multiple simultaneous pp inelastic collisions, a phenomenon called *pileup*.

Pileup

The term *pileup* refers to the simultaneous pp inelastic collisions that accompany the hard scatter of interest. There are two types of pileup: *in-time pileup* refers to additional collisions occurring between protons in the same bunch crossing as the one of interest; *out-of-time pileup* refers to collisions occurring in bunch crossings just before or just after the one of interest, which, when the electronics integrate over more than 25 ns, can affect the signal of the collision of interest. These secondary collisions tend to be soft, but they can add hundreds to thousands of soft hadrons to the final state of the hard collision of interest, biasing and smearing the quantities reconstructed from the detector [74], as well as stressing the trigger and data acquisition systems, and increasing the radiation levels that detectors and front-end electronics have to withstand.

For in-time pileup, one usually reports the average *pileup multiplicity* μ , which follows a Poisson distribution. As the accelerator complex continues to be upgraded, the number of pileup interactions has consistently been increasing. The average pileup multiplicity was $\langle \mu \rangle = 20$ in Run 1 and will reach $\langle \mu \rangle = 200$ at the High-Luminosity LHC (see below), largely above the original design value. In order to provide a higher luminosity and at the same time cope with the increased levels of pileup, a series of concomitant upgrades was planned for the collider and for the experimental detectors.

4.1.4 Brief timeline of LHC operation and upgrades

Since its ideation, the LHC was planned to be built in stages, partly to spread the costs, and partly to await technological developments. The construction ended in 2008 and the collider was expected to start running at $\sqrt{s} = 11$ TeV in 2009. However, on September 19th, 2008 an incident occurred, in which a defective joint between superconducting cables caused several magnets to quench with severe collateral damage, including the loss of six tonnes of helium and pollution of the beam vacuum tubes [71]. After the incident, it was decided to operate at $\sqrt{s} = 8$ TeV until the first Long Shutdown (LS) planned for 2013. The LHC



Figure 4.3: Snapshot of LHC schedule showing collision energy (upper line) and luminosity (bottom line) as of 2022 [75].

plan is shown in Fig. 4.3. Years-long periods of consecutive data-taking, called *Runs*, are separated by major upgrades necessary to bring the collider to its ultimate performance [76]:

- Run 1 (2009-2013): The LHC was operated with 50 ns bunch spacing and $\sqrt{s} = 7 - 8$ TeV.
- LS1 (2013-2015): The LHC machine was consolidated to allow the increase of the CoM energy and luminosity to the design value.
- Run 2 (2015-2018): The LHC was operated with 25 ns bunch spacing and $\sqrt{s} = 13$ TeV. The luminosity was progressively increased until attaining the nominal value of

$1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ in June 2016. In 2018 the peak luminosity reached the ultimate design value of $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ thanks to small emittances and a smaller than design β^* value, while still keeping the nominal bunch population.

- LS2 (2019-2022): Significant upgrades were carried out in the injector chain: the new Linac-4 accelerator (160 MeV) replaced the Linac-2 (50 MeV) as injector to the PSB, while the PSB was also upgraded resulting in a lower emittance and higher intensity beam. Other improvements included consolidation of the dipole magnets and cryogenics upgrade. The Phase I of the detector upgrades was installed and commissioned to adapt to the new conditions and in preparation for the High-Luminosity LHC.
- Run 3 (2022-2025): The LHC CoM energy was increased to 13.6 TeV. With the current machine, the peak luminosity is limited at $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ by the luminosity-induced heating of the inner triplet magnets at the IPs. However, a 60% increase in beam intensity, combined with luminosity leveling, which will allow to operate near peak luminosity for a longer fraction of the running time, will result in a year integrated luminosity above 80 fb^{-1} . This run is ongoing.
- LS3 (2026-2028): The LHC will undergo the most extensive upgrade of its components, including low- β quadrupole triplets and new crab cavities [77] at the intersection regions. The Phase II upgrades of the detectors will be installed and commissioned.
- High-Luminosity LHC (2029-2040's): The LHC is expected to run at $\sqrt{s} = 14 \text{ TeV}$ and deliver a levelled instantaneous luminosity of $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ and an annual integrated luminosity of 250 fb^{-1} .

To further extend the physics potential of the LHC, CERN started in 2010 the High-Luminosity LHC (HL-LHC) project, aiming at a peak luminosity of $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, resulting in a total of 3000 fb^{-1} of data collected after 12 years of operation, a ten-fold increase with respect to the data that will have been collected at the end of Run 3 and well beyond the original design values of the collider.

4.2 The ATLAS detector

The ATLAS (A Toroidal LHC Apparatus) [62] experiment is one of the two high-luminosity general purpose experiments at CERN, together with CMS, built for the precise measurements of SM parameters and to search for a wide range of possible new physics phenomena.

The design of the ATLAS detector was driven by the vast physics program of the experiment and by the experimental difficulties posed by two major challenges: the unprecedented

levels of pileup and the large background of QCD jet production due to the nature of pp collisions. Several requirements had to be satisfied to provide a wide physics reach [62]: fast and radiation-hard electronics and sensors; high detector granularity to handle the large particle fluxes and possible simultaneous hard collisions; large pseudorapidity acceptance to allow detection of forward particles, and almost full azimuthal angle coverage to allow complete event reconstruction on the transverse plane; good electromagnetic calorimetry for identification of electrons and photons, and hadronic calorimetry with full coverage for reconstruction of jets and missing transverse energy; a tracking detector to provide good charged particle momentum resolution, especially complementing at low transverse momentum the poorer calorimetry energy resolution, and good reconstruction efficiency of secondary vertices for τ -leptons and b -jets identification; good muon identification and momentum resolution, and ability to determine the charge of high p_T muons; highly efficient triggering, especially on low transverse-momentum objects, to reduce the event rate, while keeping high efficiency for rare processes. These requirements set new standards for the design of particle detectors and the final result was only possible thanks to the work of several thousands of physicists, engineers, and technicians over fifteen years.

The ATLAS detector is located in the experimental cavern at Point 1 at CERN. With its 25 m in height and 44 m in length, it is the largest detector at the LHC, weighting approximately 7000 tonnes. Coaxial layers of sub-detectors, each sensitive to different types of particles, surround the interaction point (IP). The detector has a cylindrical shape and was designed to be forward-backward symmetric and to provide an almost full azimuthal coverage. The detector layout is subdivided in two parts: a main cylinder coaxial to the beam line, called the *barrel*, and two *end-cap* regions closing the cylinder on both sides. The full detector system is immersed in a magnetic field for the bending of charged-particle trajectories necessary for charge-momentum measurement.

This section presents a description of the detector as it was during Run 2, the period in which the data set used in this thesis was collected.

4.2.1 The ATLAS coordinate system

ATLAS uses a right-handed coordinate system, as sketched in 4.4. The IP is the center of the coordinate system. The z -axis is placed along the beam direction, while the x - y plane, referred to as the *transverse plane*, is perpendicular to the beam trajectory. The positive x -axis points towards the center of the LHC, while the positive y -axis points upwards. Cylindrical coordinates are used, with the azimuthal angle ϕ measured on the transverse plane around the beam axis, and the polar angle θ measured on the z - y plane from the

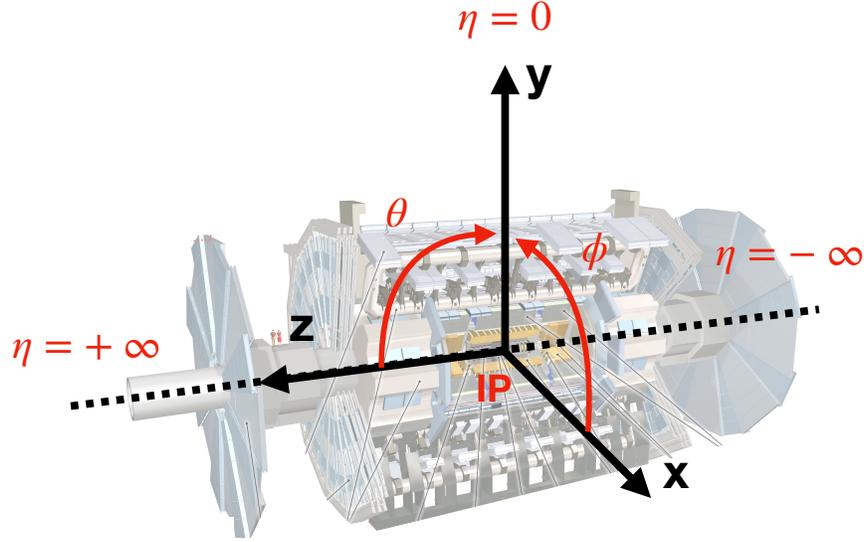


Figure 4.4: ATLAS coordinate system. Background taken from Ref. [62].

positive z -axis. The rapidity of a particle with energy E and momentum along the z direction p_z is defined as $y = 1/2 \ln [(E + p_z)/(E - p_z)]$. Differences in rapidity are Lorentz invariant under boosts along the z (beam) direction. The ΔR distance between two objects i and j in the rapidity-azimuthal angle space is also Lorentz invariant and is defined as $\Delta R_{i,j} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$. Lorentz invariance under boosts in the z -direction is important in a pp collider where, because of the complex QCD structure of protons, the four-momentum is not conserved along the z component. Related to θ is the pseudorapidity $\eta = 1/2 \ln [(p + p_z)/(p - p_z)] = -\ln \tan(\theta/2)$, with $\eta = 0$ along the y -axis, and $\eta = \pm\infty$ along the z -axis. For a massless particle, rapidity and pseudorapidity are equal.

4.2.2 The magnet system

The detector is fully immersed in a magnetic field provided by a system of four large superconducting magnets [62], as shown in Fig. 4.5. The field covers approximately 12,000 m² (22 m in diameter and 26 m in length) and has a total stored energy of 1.6 GJ. A thin superconducting solenoid [78] aligned with the beam axis surrounds the ID cavity and provides an axial magnetic field of 2 T at the IP. The solenoid assembly was optimized to minimize the radiative thickness between the ID and the EM calorimeter, resulting in only ≈ 0.66 radiation lengths at normal incidence. Three superconducting toroids, one surrounding the barrel and two at the end-caps, produce toroidal magnetic fields of approximately 0.5 T and 1 T, respectively. The entire system is under vacuum and cooled down by the ATLAS cryogenics system. A precise description of the magnetic field in the detector volume is necessary

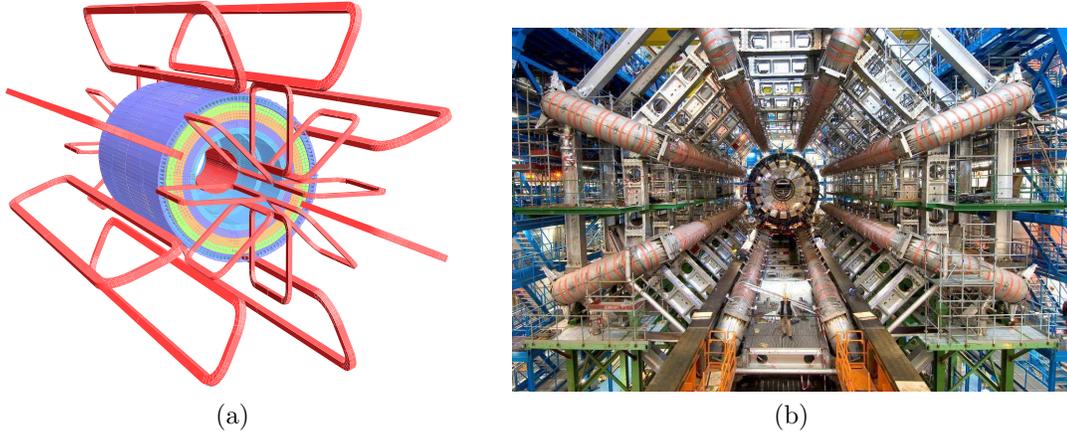


Figure 4.5: a) Sketch of the ATLAS magnet system, with the barrel and end-cap toroids, and the solenoid placed inside the Tile calorimeter volume [62]. b) Picture of the barrel toroid as installed in the underground cavern, with the barrel calorimeter and embedded solenoid visible on the other side, awaiting to be put in position. The person standing in front of the structure provides the scale [62].

for a high momentum measurement resolution. For this purpose, the ID is provided with four NMR probes located at $z = 0$ and equally spaced in azimuth distance, which provide a $|B|$ measurement with an accuracy of 0.01 mT, while the rest of the solenoid and the muon chambers are equipped with 3D Hall cards to measure each field component.

4.2.3 The inner detector

The Inner Detector (ID) [79, 80] is the closest sub-detector to the beam pipe and has to sustain the largest radiation dose and flux of particles. Despite the harsh environment, the ID has to provide precise momentum and vertex measurements. This is achieved through three complementary subcomponents, shown schematically in Fig. 4.6. On the inner part, the Pixel and SemiConductor Tracker (SCT) silicon detectors provide high-resolution tracking and vertex reconstruction; on the outer part, the Transition Radiation Tracker (TRT) provides straw-tube tracking with transition radiation detection capability for electron identification. The Insertable-B Layer (IBL) was added in front of the Pixel detectors before Run 2 to react to the harsher conditions. The different components are arranged as concentric cylinders in the barrel regions and as stacked disks perpendicular to the beam axis in the end-caps. The entire ID system is 2.3 m in diameter and 7 m in length.

The current ID was designed for 10 years of operation at nominal LHC parameters. Albeit minor upgrades, the ID performance remained adequate even once the LHC exceeded the design values. However, in order to sustain the HL-LHC conditions, the ID will have to be fully replaced by a new tracking system, the Inner Tracker (ITk) [81, 82], which will be

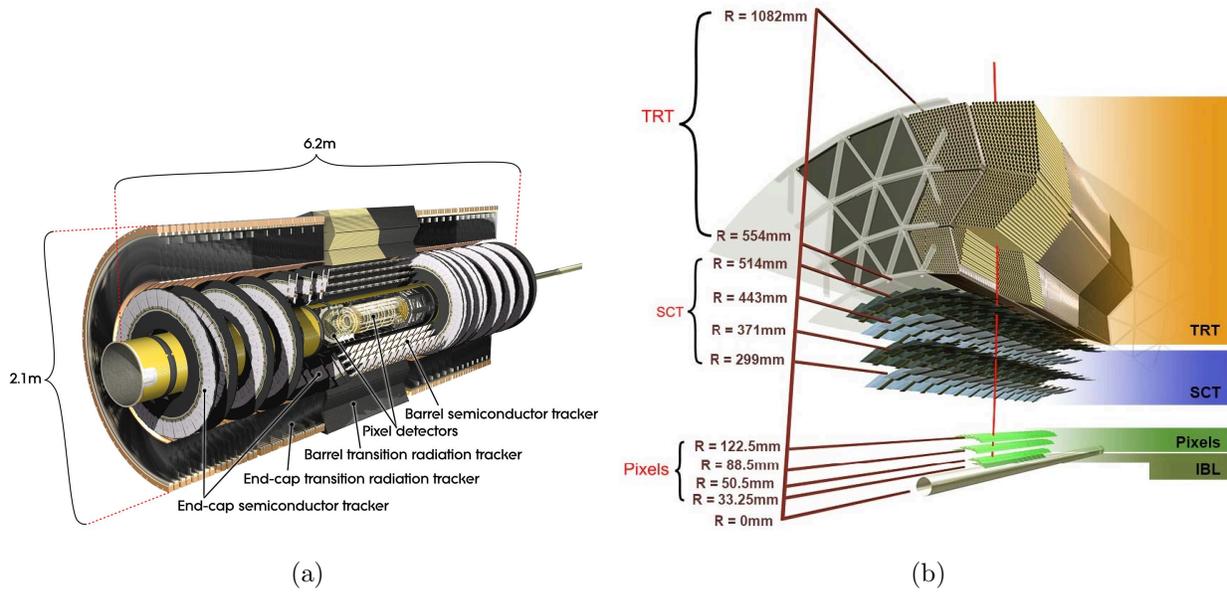


Figure 4.6: a) Schematic view of the ATLAS ID [83]. b) Cross sectional view of the ID barrel region traversed by a charged track (in red) [76]. From the collision point, the track traverses the beryllium pipe, the three silicon pixel layers, four SCT barrel layers of silicon micro-strip sensors, and ~ 36 TRT straws.

installed during the Phase II upgrades.

Pixel detector

The pixel system [84] is composed of three layers in the barrel region and three disks at each end-cap, and covers the region $|\eta| < 2.5$. The basic building block of the detector is a module composed of pixel sensors, front-end electronics, and control circuits. The nominal pixel size is of $50 \mu\text{m}$ in the ϕ direction and $400 \mu\text{m}$ in the z or r direction, resulting in 67 million pixels in the barrel and 13 million in the end-caps, for a total of ≈ 80 million readout channels. The detector was designed to provide at least three points per charged track, with intrinsic accuracies of $10 \mu\text{m}$ in the $(R - \phi)$ plane and $115 \mu\text{m}$ along the z direction.

The SemiConductor Tracker

The SCT [85] is composed of four cylindrical layers in the barrel and nine discs in each end-cap. Silicon micro-strip sensors are connected to 6.3 million readout channels. Each track crosses eight strip layers, giving four two-dimensional space points and providing an intrinsic accuracy of $17 \mu\text{m}$ in the $R - \phi$ plane and $580 \mu\text{m}$ along the z or r direction.

Transition Radiation Tracker

The TRT [86, 87] is the outermost layer of the ID and provides continuous tracking coverage within the $|\eta| < 2$ range, with an average of 36 hits per track. While the TRT provides only $R - \phi$ measurements and has a lower accuracy of $130 \mu\text{m}$, it provides a larger number of hits and longer track lengths, complementing the precision trackers. The TRT consists of about 350,000 straw tubes, 4 mm in diameter and with a $31 \mu\text{m}$ diameter gold-plated tungsten wire at their center. The straws are arranged parallel to the beam line in the barrel and in a wheel-like shape parallel to the transverse plane in the end-cap. A charged particle passing through a straw ionizes the gas producing some primary ionization electrons, which are accelerated by an electric field towards the wire, inducing an electron avalanche that produces a detectable signal. The straws are interleaved with transition radiation (TR) material. The amount of TR is proportional to the relativistic factor $\gamma = E/m$ of the incident particle, so that a particle as light as the electron will produce significantly more TR photons than a pion or muon. In fact, an essential function of the TRT is electron identification [88].

Insertable B-Layer

The first layer of the Pixel detector closest to the beam pipe was originally designed to be regularly replaced. However, changes to the detector system that became necessary during ATLAS operation made the extraction of the layer no longer possible. In order to retain a high tracking performance until the end of Run 3, it was therefore decided to add a new innermost layer, the Insertable B-Layer (IBL) [89, 90]. The IBL was installed at 3.3 cm from the beam axis, between the existing pixel detector and a smaller beam-pipe, and provides a longitudinal coverage of $|\eta| < 3$. A combination of planar and 3D sensor technologies are used and the pixel cells size is reduced to $50 \times 250 \mu\text{m}^2$. The new layer was installed during the first long shutdown (LS1) before Run 2 and has been successfully operating since 2015.

4.2.4 The calorimeters

Calorimeters [91, 92] are detectors used to measure the energy of incident particles via their total absorption. When particles are stopped in the detector volume, showers are initiated. The incoming high energy particle is converted into two or more lower-energy particles, which in turn produce more daughter particles. The cascade process stops when the final particles have an energy smaller than what would be needed to produce further particles. The shower evolution differs according to whether the incident particle interacts electromagnetically or hadronically, requiring different types of detectors. Calorimeters are particularly well-suited for a high-energy multipurpose experiment, as their energy resolution improves with energy as $1/\sqrt{E}$ (in contrast to, for instance, a magnetic spectrometer, whose

momentum resolution deteriorates linearly with increasing particle momentum), they provide sensitivity to both charged and neutral particles, and can provide indirect neutrino detection, measure the arrival time of particles, and provide fast signals for triggering. As they are typically segmented transversely and longitudinally, they also allow to measure the position of particles and discriminate between different particle types according to the shape of the shower. Triggering and particle identification are particularly relevant for this work.

Electromagnetic showers

An electromagnetic (EM) calorimeter measures EM showers induced by electrons and photons. At energies above ~ 100 MeV, electrons lose their energy almost exclusively via bremsstrahlung, while photons via electron-positron pair production. The depth and width of the showers can be expressed in terms of the radiation length X_0 of the detector material³. An electron traveling in a material covers an average distance of $x = X_0$ before its energy is reduced by $1/e$, while a photon travels a distance $x = 9/7X_0$. The resulting EM showers tend to be quite contained in width and length, requiring a smaller detector volume, but a higher granularity for precision measurements of the shower position. The energy of the incident particle initiating an EM shower is proportional to the energy deposited by the charged particles in the shower through ionization and excitation.

Hadronic showers

Hadrons interact with the detector material mostly via the strong interaction and thus exhibit different shower characteristics, often expressed in terms of the interaction length λ (the mean free path). The hadronic cascade presents two types of processes. The first type results in the production of high-energy secondary hadrons, typically at the GeV scale, with $\lambda \approx 35A^{1/3} \text{ g cm}^{-2}$. This is generally larger than the radiation length X_0 , requiring larger detector volumes. The second type consists in nuclear interactions with large transverse momentum transfers, such as excitation or nucleon evaporation, which produce particles in the MeV scale and broaden the shower shape. The soft spectrum of these inelastic processes is dominated by neutrons, photons, and electrons, while the energetic component is populated by pions and, in lower quantities, by kaons, nucleons, and other hadrons. One third of the pions produced are neutral pions, which quickly decay into two photons before they have a chance to interact hadronically, initiating electromagnetic sub-cascades within the hadronic

³The radiation length for a material with atomic number Z and atomic weight A is given by [92]

$$X_0(\text{g/cm}^{-2}) \simeq \frac{716 \text{ g cm}^{-2} A}{Z(Z+1) \ln(287/\sqrt{Z})}.$$

shower. The total fraction of energy in a hadronic shower that comes from an EM shower is called the *electromagnetic fraction* f_e . As the energy of the incident particle increases, the number of energetic hadronic interactions does as well, inducing a larger f_e .

A large fraction of the energy in a hadronic shower escapes detection: part of the energy is used to break up nuclear bonds; some energy goes into short-range nuclear fragments absorbed before they get to the active layers; long-lived or stable neutral particles, such as neutrons, K_L^0 and neutrinos can escape from the calorimeter, while muons produced by pions and kaons decays can deposit only part of their energy. The lower the electromagnetic fraction of the shower, the larger the fraction of this *invisible energy* from hadronic interactions. As this form of invisible energy fluctuates between events, it will affect the energy resolution. Because f_e , and therefore the fluctuation, is energy dependent, the calorimeter response will be nonlinear with energy. *Compensating calorimeters* are detectors that compensate for the loss of this invisible energy.

Sampling calorimeter

In a *sampling calorimeter* the functions of particle absorption and energy measurement are performed by different components (in contrast to a *homogeneous calorimeter*, where only one medium is used). This allows to choose the optimal material for each task, at the expense of an increase in the fraction of unmeasured energy. Sampling calorimeters are generally built in alternating layers of heavy absorbing material, such as lead, and layers of active material. This makes them more easily segmented longitudinally and radially, resulting in better space resolution and particle identification. The shower generation starts in the absorber, while the active layer generates and measures the detectable signal. Sampling calorimeters are classified according to the type of active material, where the deposited energy can be measured by collecting either the light produced in a scintillating material or the charge produced by ionization.

The ATLAS Calorimeter

The ATLAS calorimeter [62, 93, 94] surrounds the ID and is a hybrid system consisting of EM and hadronic sampling calorimeters covering the range $|\eta| < 4.9$. A schematic view is shown in Fig. 4.7. A finely segmented EM calorimeter closest to the beam-line provides precision measurements of electrons and photons. This is surrounded by a coarser granularity hadronic calorimeter (HCal), which provides sufficient resolution for jet and missing-transverse-energy measurements. A forward calorimeter in the end-cap closest to the interaction region extends the detector acceptance to high η . Three cryostat systems surround the calorimeter components that use LAr technology: one cryostat houses the barrel EM (EMB) calorimeter,

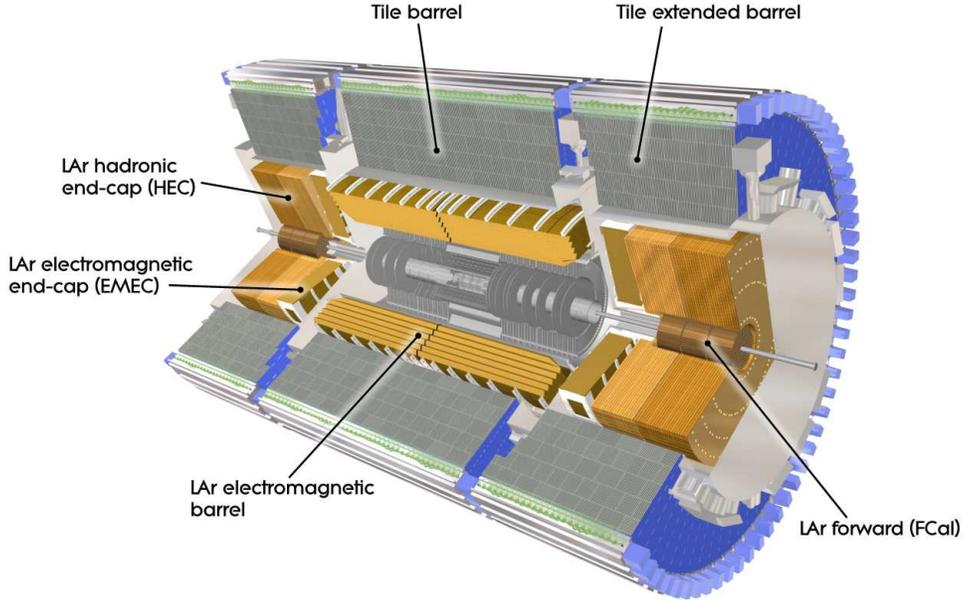


Figure 4.7: Schematic view of the ATLAS calorimeter [62].

while a cryostat at each end-cap contains the forward calorimeter (FCal), the EM end-cap (EMEC) calorimeter, and the hadronic end-cap (HEC) calorimeter. Hadronic calorimetry in the barrel is based instead on a steel-scintillator Tile calorimeter (TileCal). All cryostats are vacuum insulated and maintain a temperature of -184°C . The barrel cryostat is a 6.8 m long cylinder with inner and outer radii of $1.15\text{ m} < r < 2.25\text{ m}$, while the end-cap cryostat has length 3.17 m and a radius of 2.25 m, the same as the barrel. The superconducting solenoid is placed in the same insulation vacuum as the LAr system.

The electromagnetic calorimeter

The EM calorimeter is a high-granularity LAr sampling calorimeter, covering the range $|\eta| < 3.2$ and providing excellent energy and position resolution. In the barrel, the calorimeter consists of two half-cylinders joined at $z = 0$ and covering together the region $|\eta| < 1.475$. In the end-cap, the EMEC consists of a 63 cm thick wheel covering the region $1.375 < |\eta| < 3.2$. Each wheel is partitioned into two co-axial wheels that are joined at $|\eta| = 2.5$, matching the acceptance of the ID. The absorber plates are made of lead with a thickness of 1.1 – 2.2 mm, according to the η -region. These are interleaved with thin layers of LAr and readout electrodes. An accordion geometry is used in both the barrel and the end-caps, providing a uniform performance as a function of ϕ , and a fast signal readout. In the region $|\eta| < 2.5$, where precision studies were considered possible, the EM calorimeter is segmented into three sampling layers, while two layers with coarser granularity are used in the end-caps.

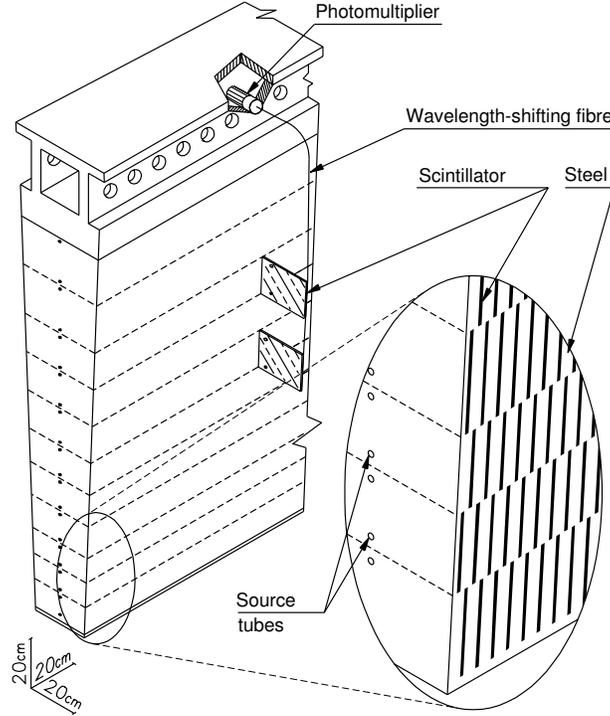


Figure 4.9: Schematic view of the tile calorimeter showing the various components of the optical readout: tiles, fibers, and photo-multiplier tubes [62].

$|\eta| < 1$, and two extended barrels of 2.6 m in length, up to $|\eta| < 1.7$. It has a radial depth of approximately 7.4λ . It is composed of 64 modules made of steel absorber plates and plastic scintillator tiles as the active medium. A schematic view of the optical readout is shown in Fig. 4.9. Ionizing particles that cross the tiles induce the production of ultraviolet scintillation light. Wavelength-shifting optical fibers, in contact with the tile edges, collect the scintillation light, convert it to a longer wavelength, and transmit it to the photo-multiplier tubes located at the outer edge of each module. The grouping of the 540,000 readout fibers into bundles provides a three-dimensional cell structure with three layers in depth, of 1.5, 4.1 and 1.8λ thickness at $\eta = 0$, and a $\Delta\eta \times \Delta\phi$ granularity of 0.1×0.1 in the first two layers and 0.2×0.1 in the last one.

The hadronic end-cap calorimeter (HEC) is a LAr sampling calorimeter that uses copper as the absorber material. Each HEC is designed as two wheels, with the outer wheel built with thicker copper layers (50 mm instead of 8.5 mm). The HEC has a depth of approximately 10λ . A granularity of 0.1×0.1 is used up to $|\eta| = 2.5$, while for higher η values it is reduced to 0.2×0.2 .

The forward calorimeter

The FCal provides both EM and hadronic energy measurements, and extends the detector

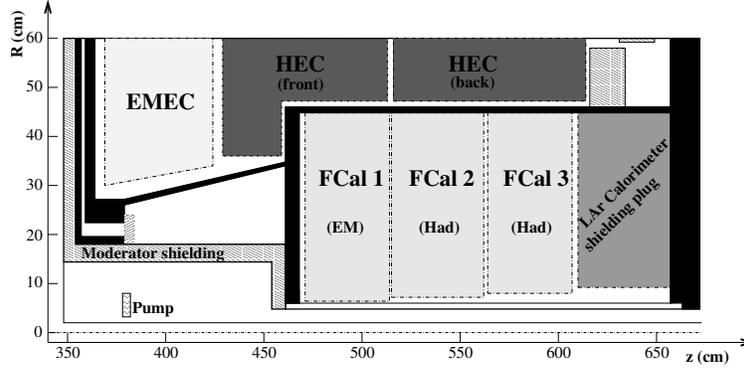


Figure 4.10: Schematic view of the FCal calorimeter inside the end-cap cryostat, showing the three FCal modules, the shielding layers, and the cryostat walls in black [62].

acceptance up to $|\eta| = 4.9$. The FCal is the innermost layer of the end-cap detectors, positioned along the beam axis at 4.7 m from the IP. The high levels of radiation in this region make this a particularly challenging detector. In order to reduce the neutron albedo in the ID cavity, the front face of the detector is moved back by about 1.2 m with respect to the EMEC edge. This is a trade-off in longitudinal length, requiring a very high density detector to integrate the full interaction length of the forward particles and preventing energy spills and pile-up contamination into the surrounding detectors. The design, as shown in Fig. 4.10, consists of three sections: the first section, closest to the IP, uses copper as the absorber to optimize detection of EM radiation; the two outer layers use tungsten for hadronic calorimetry.

Calorimeter read-out

The building block of the calorimeter readout is a *cell*, defined by the total integrated energy deposited in its volume, and by its (η, ϕ) coordinates and the sampling layer where it is located. The dynamic range for the energy of the cells goes from ~ 10 MeV up to 3 TeV. The lower limit is set by the irreducible thermal noise in the calorimeters, also referred to as *electronic noise*. The other source of noise in the cell readout is *pileup noise*, caused both by in-time and out-of-time pileup, for a total noise given by:

$$\sigma_{\text{noise}}^{\text{total}} = \sqrt{(\sigma_{\text{noise}}^{\text{electronic}})^2 + (\sigma_{\text{noise}}^{\text{pileup}})^2}. \quad (4.3)$$

Prior to 2011, the total noise was driven by the electronic noise, but with the increasing luminosity the pileup term has now become the dominant source [95]. Its effect is not homogeneous in the detector volume, as shown in Fig. 4.11 for a $\langle \mu \rangle = 200$ simulation. The majority of the energy flow is absorbed by the LAr calorimeters, while the Tile calorimeter

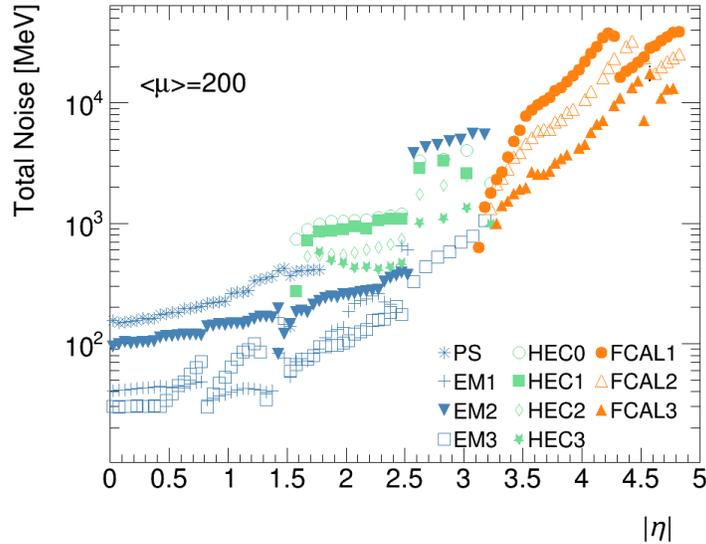


Figure 4.11: The total energy-equivalent cell noise at the EM scale as a function of η for the different detector sampling layers for a HL-LHC simulation with $\mu = 200$.

has little sensitivity. The energy deposited in the calorimeters is processed by the on-detector (*front-end*) and off-detector (*back-end*) electronics. The necessity for low electronic noise and low latencies favored the choice of a readout architecture with analog processing close to the detectors. The amplification, processing, and digitization of the analogue signals is therefore performed directly by the front-end electronics, which in turn required custom designed radiation-tolerant ASIC's. The back-end system is instead located in the main services cavern (USA15), 70 m away from the detector, and is made of commercial components.

4.2.5 The muon spectrometer

The muon spectrometer [62, 96] detects the momentum and trajectory of charged particles escaping the hadronic calorimeter. In the energy range of the LHC, muons behave like minimum ionizing particles. This characteristic behavior makes them easily distinguishable in the detector, making muons essential pieces in many analyses. However, this also requires their momentum to be inferred from the curvature of their trajectory, rather than from the energy deposition in a calorimeter, a measurement that degrades with increasing energies. Accurate measurement of muons was an important design goal for the experiment and significantly shaped the design of the entire detector, starting from its size.

A schematic view of the ATLAS muon system is shown in Fig. 4.12. Three superconducting air-core toroids provide strong bending power within a large open volume. High-precision tracking chambers provide excellent muon momentum measurement within $|\eta| < 2.7$, while

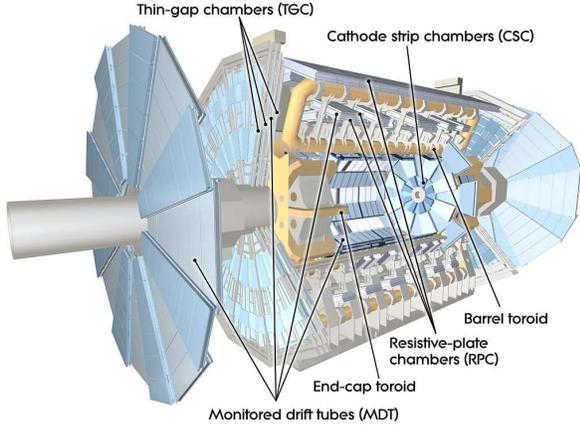


Figure 4.12: Schematic view of the muon spectrometer [62].

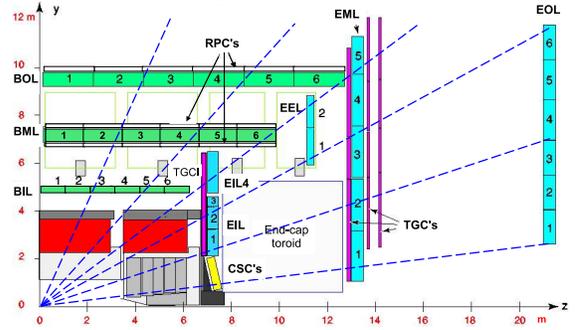


Figure 4.13: Cross-section of the muon spectrometer along a plane containing the beam axis. The straight lines represent infinite momentum muons, which traverse three muon stations [62].

trigger chambers with position and timing resolution cover the range $|\eta| < 2.4$. A cross-sectional view of the different detector components is shown in Fig. 4.13. Accurate muon momentum reconstruction relies on the precise alignment between the muon chambers and on an accurate magnetic field reconstruction. For this purpose, a high-precision optical alignment system monitors the relative positions and possible deformations of the MDT chambers, while approximately 1800 Hall sensors monitor the magnetic field throughout the spectrometer volume.

The precision chambers

The precision-tracking chambers are positioned in three concentric cylindrical layers in the barrel, in between the coils of the barrel toroid magnet, and in three parallel wheels in the end-caps, in front and behind the end-cap toroids. The precision measurement of muon momentum is performed by the Monitored Drift Tubes (MDTs) in almost all the spectrometer volume, covering the region $|\eta| < 2.7$. In the region $2 < |\eta| < 2.7$, the innermost layer is made of Cathode Strip Chambers (CSCs), a finer granularity detector required by the higher background rates in this region. The CSCs are multi-wire proportional chambers, where cathode planes made of strips are positioned in orthogonal directions to provide measurements of both coordinates.

The muon triggers

The precision-tracking system is complemented by fast trigger chambers in the region $|\eta| < 2.4$. The trigger chambers look for high transverse momentum muon tracks and deliver track information to the Level-1 muon trigger within a few tens of nanoseconds from the passage

of a particle. Reconstructed tracks are required to originate from approximately the IP and to pass certain p_T thresholds. Additionally, the trigger provides bunch-crossing identification and measurements of both coordinates, complementing the MDTs measurement. In the barrel ($|\eta| \leq 1.05$), three layers of Resistive Plate Chambers (RPCs) are operated in avalanche mode. In the end-cap region ($1.05 \leq |\eta| \leq 2.4$), the trigger is composed of four layers of Thin-Gap Chambers (TGCs), multi-wire chambers operated in saturated mode.

4.2.6 The forward detectors

In addition to the main detector systems, ATLAS is equipped with a set of smaller sub-detectors located in the very forward region $|\eta| > 5$ on both sides of the IP. Moving away from the IP, the first system is the LUCID (LUminosity measurement using Cerenkov Integrating Detector) detector [62]. Located at ± 17 m from the IP, LUCID primarily provides online relative luminosity monitoring. Next is the Zero-Degree Calorimeter (ZDC) [62], located at approximately zero degrees to the incident beam, ± 140 m from the IP, and whose primary purpose is to detect neutral forward ($|\eta| > 8.3$) particles. Coincidence requirements on the two ZDC systems are also used to suppress beam-induced backgrounds and provide some knowledge of the vertex location. The ALFA (Absolute Luminosity For ATLAS) detector [62], located at ± 240 m, is used to measure the absolute luminosity. The AFP (ATLAS Forward Proton) detector [97] was installed in 2017 at ± 204 m and ± 217 m from the IP to measure diffractive protons scattered at small angles ($100 \mu\text{rad}$), where one or both protons remain intact.

4.3 The ATLAS Trigger

The ATLAS Trigger [62, 98] system is responsible for the selection of the subset of events to be stored on disk and used in physics analyses. With an LHC event rate of 40 MHz and an event size of approximately 1.5 MB, the ATLAS trigger has to handle a data volume of 60 TB/s. Storing all this data is not only unfeasible, but it is also not desirable, as the events that are interesting for physics analyses are orders of magnitude rarer than the large background of QCD jet production and pileup, as shown in Fig. 4.2. The role of the TDAQ system is therefore to process the live stream of data coming from the detectors and select the most interesting events to study, while rejecting the remaining 99.9975%. As the events that are rejected are lost, this is a crucial step for the ATLAS experimental program.

In order to handle the large data flow, while keeping high signal efficiency and background rejection, the ATLAS trigger is a two-level system. The first pass is a hardware-based trigger,

executing fast algorithms on custom electronics for a first coarser selection. The reduced event rate is then processed by the second step in the trigger chain, which can run more complex algorithms on commercial software. During Run 2, the first trigger stage was called *Level-1* (L1) and the second step was called the *High-Level Trigger* (HLT)⁴. The flow of data from the detectors, through the trigger chain, up to when the data is written to disk, is controlled by the Data Acquisition (DAQ) system. The full Run 2 Trigger and Data Acquisition (TDAQ) system is shown in Fig. 4.14. The L1 trigger receives partial event data from the detector. If the event passes the L1 trigger (L1-Accept), the full event data is read-out by the front-end electronics of all the detectors and sent to the ReadOut Drivers (RODs), which perform an initial processing and pass it to the ReadOut System (ROS). The ROS buffers the data and sends it to the HLT on HLT request. Events that pass the HLT selection (HLT-Accept) are transferred to local storage, ready for offline reconstruction.

4.3.1 The Level-1 trigger

The L1 trigger is a hardware-based system that reduces the LHC event rate of 40 MHz down to the maximum detector read-out rate of 100 kHz. In addition to rejecting events, the L1 trigger identifies Regions of Interest (RoI) in $\eta \times \phi$ to be used by the algorithms in the next stage of the trigger chain.

The 25 ns interval between collisions is too short for the processing and evaluation of the trigger decision. Therefore, while the trigger decisions are being formed, the collision data is stored in memory buffers. These memories are contained on electronics on or near the detector, where radiation is high and costs and readout reliability put constraints on the amount of time the data can be stored for. For this reason, the maximum L1 latency, defined as the time between the pp collision of interest and the moment the L1 trigger decision is made, is required to be less than 2.5 μ s. Custom-built electronics are needed to satisfy these requirements.

The L1 trigger receives reduced-granularity data from the calorimeter and muon detectors, with the two detector systems handled by separate trigger components, the L1 Muon trigger (L1Muon) and the L1 Calorimeter trigger (L1Calo). The results are passed to the L1 Topological (L1Topo) processor, which was added during the first Long-Shutdown (LS1) in order to cope with the increased event rates by providing more sophisticated topological selections. The final step in the L1 trigger chain is the Central Trigger Processor (CTP),

⁴During Run 1 the HLT was based on two separate farms: the Level-2 (L2) trigger requested reduced event data and provided a first coarse selection. The reduced event rate was then processed by the Event Filter (EF), which had access to the full event information and longer latency. For Run 2, the L2 and EF were merged into a single system to allow better resource sharing and simplify the hardware and software [98].

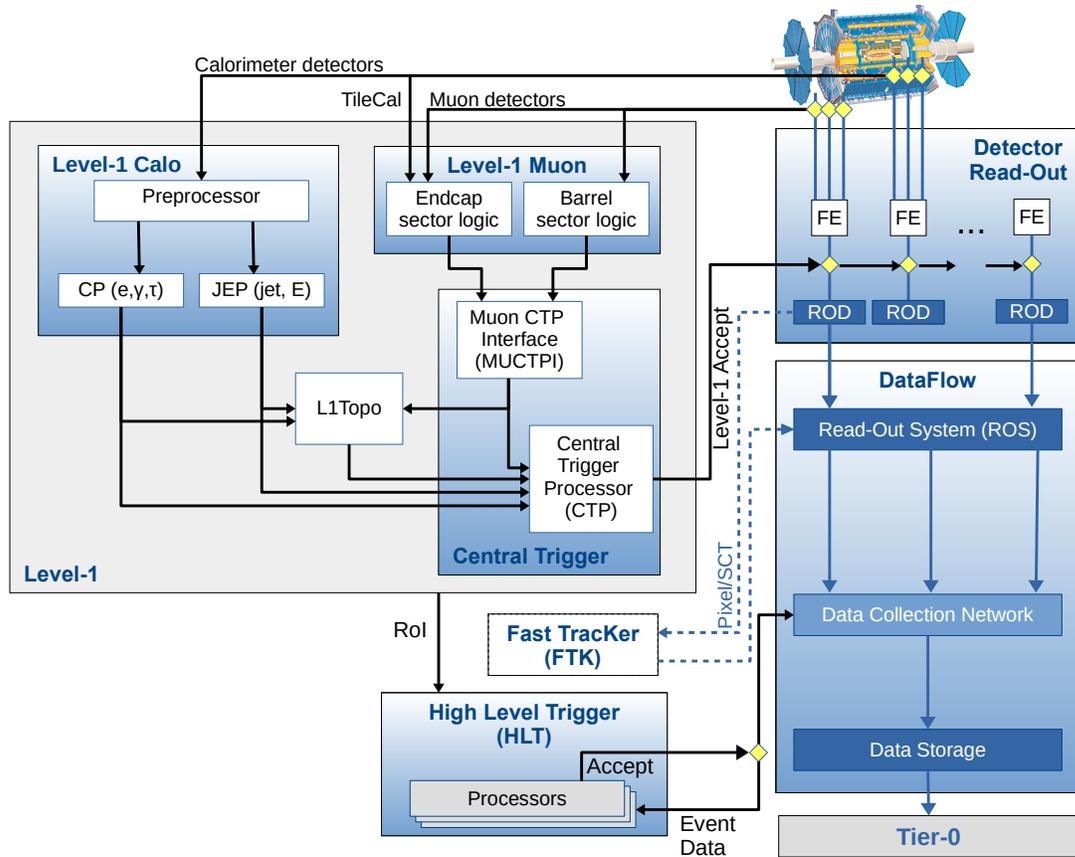


Figure 4.14: The ATLAS TDAQ system in Run 2, showing the relevant L1 and HLT trigger components, as well as the detector read-out and data flow to permanent storage on L1- and HLT-Accept [99]. Note that the Fast Tracker project was canceled and should be ignored in this figure.

which provides the L1 trigger decision to the TDAQ system.

L1Muon

The L1Muon [100] uses the hits from the RPC and the TGC muon triggers to apply coincidences requirements and identify high p_T muon candidates. The results from L1Muon are sent to the CTP via the Muon Central Trigger Processor Interface (MUCTPI).

L1Calo

The L1Calo [101] receives signals from all the calorimeter detectors and uses information about the energy deposits to identify high E_T objects or energy sums of interest. The input data consists in trigger towers of coarser granularity than the calorimeter cells, mostly 0.1×0.1 in $\Delta\eta \times \Delta\phi$, with larger sizes in the end-caps. A tower takes up the full depth of

each EM or hadronic calorimeter. The number of cells used to form a tower varies with the granularity of the calorimeter element, and it goes from a few in the end-caps, up to 60 in the LAr EM barrel. In the TileCal, most towers are built by summing the signal from five photo-multiplier tubes. The analogue trigger-tower signals are carried from the front-end electronics of the calorimeters to the L1Calo system located fully off-detector in the USA15 cavern. The L1Calo system consists of three main sub-systems. The Pre-Processor [102] digitizes the analogue calorimeter signals, identifies the bunch-crossing they originated from, and performs a series of operations to clean and calibrate the signals. The data is then transmitted in parallel to the Cluster Processor and the Jet/Energy-sum Processor (JEP), which use sliding window algorithms to identify energy depositions of interest. See Ref. [101] for a comprehensive description of the algorithms.

The Cluster Processor identifies clusters of energetic towers that could be associated to an electron, photon, or tau that pass a programmable E_T threshold and, if desired, isolation requirements. It operates within the region $|\eta| < 2.5$, corresponding to the boundary of high-precision data from the ID and EM calorimeter. The e/γ algorithm looks for narrow high E_T deposits in the EM calorimeter. To suppress hadronic jet background, the deposits are required to be isolated in the transverse plane and to not penetrate into the HCal. The τ /hadron algorithm, looks for collimated hadronic τ decays with looser isolation requirements and allowing deposits in the HCal.

The JEP is similarly used to identify jets and produce total, missing, and jet transverse energy sums. The jet trigger uses data up to $|\eta| < 3.2$, the limit of end-cap acceptance, while the energy sums extend up to $|\eta| < 4.9$, including also FCal information. For the purposes of jet and energy sum reconstruction, a coarser granularity can be used, and the EM and hadronic calorimeters do not need to be considered separately. The JEP towers, called *jet elements*, are the sum of 0.2×0.2 in $\Delta\eta \times \Delta\phi$ windows in the EM calorimeter and in the hadronic calorimeter. The jet algorithm calculates the E_T sums in windows of 2×2 , 3×3 , and 4×4 jet elements and compares them to programmable thresholds specifying the minimum E_T requirement and the window size. The different sizes are sensitive to different signatures: smaller windows are better suited to discriminate nearby small-radius jets, while larger sizes are more efficient for individual energetic large-radius jets. As the windows are overlapping, a jet can exceed the energy threshold in more than one window. In order to avoid double-counting of jets, a 2×2 window is required to be a local maximum compared to its eight neighboring jet elements. This is used to also define the η and ϕ coordinates of the RoI.

The Common Merged Module (CMM) merges the results from the Cluster Processor and JEP modules and sends the information to the CTP in the form of Trigger Objects (TOBs),

described by the E_T , η and ϕ coordinates, and the isolation threshold when used.

L1Topo

The L1Topo [103] consists of two modules of Field Programmable Gate Arrays (FPGAs). The modules are provided with the same TOBs from the L1Calo and MUCTPI systems, and execute parallel and independent algorithms. To reduce the combinatorics, part of the computational time is dedicated to produce reduced lists of sorted TOBs. The remaining time is used to evaluate the algorithms on the reduced lists. Various algorithms are available: angular separations in $\Delta\phi$, $\Delta\eta$, and ΔR ; energy thresholds of objects inside a cone; selection on invariant, transverse, or effective mass; event-hardness selections; corrections to the E_T^{miss} . L1Topo can also apply requirements on triggers from adjacent bunch crossings. The L1Topo decisions are transmitted to the CTP after \approx ns.

Central Trigger Processor

The L1 trigger decision is formed by the Central Trigger Processor (CTP) [104]. The CTP receives inputs from the L1Calo, L1Muon through the MUCTPI, and L1Topo, as well as from some detector subsystems. The trigger decision is implemented as a logical combination of the L1 outputs according to the *trigger menu* (see Sect. 4.3.3). The CTP is also responsible for applying pre-scales on certain menu items and for applying the *deadtime*, a mechanism used to prevent the detector front-end buffers from overloading by limiting the number of L1-Accepts. If an event passes any of the L1 trigger items, a L1-Accept signal is sent. On L1-Accept, L1 trigger decisions and RoIs are sent to the HLT.

4.3.2 The High-Level Trigger

The HLT reduces the event rate from 100 kHz down to 1 kHz. The HLT has access to the full granularity calorimeter information, data from the muon spectrometer precision chambers, and tracking information from the ID. The processing sequence consists in a first step, in which fast algorithms provide a fast coarse rejection, followed by a finer selection using CPU-intensive algorithms similar to the ones used in offline reconstruction. The algorithms are based on the offline software Athena [105] and are run on a farm of more than 40,000 Processing Units (PUs), which are continuously replaced with newer hardware throughout operations. Some algorithms use the L1 RoIs as seeds, requiring event data only around the RoI, while others require data from the full detector. The HLT algorithms were developed to be as close as possible to their offline versions. For instance, jet reconstruction is performed using the anti- k_t algorithm [106] with a radius parameter R of 0.4 or 1.0 (see Sec. 5.2). A detailed description of the HLT algorithms can be found in Ref. [98]. The HLT latency is of

a few hundred milliseconds. On HLT-Accept, the events are transferred to local storage and are ready for offline reconstruction.

4.3.3 Trigger operations

During detector operations [99] the trigger configuration determines the active triggers. For an event to be accepted, it has to pass one L1 trigger, referred to as *L1 item*, and one HLT trigger. A *trigger chain* is defined by a combination of one L1 item and one or more HLT selections. Trigger names are usually given by the name of the trigger level (L1 or HLT), followed by the object multiplicity, the particle type (e.g. j for jet, or \mathbf{x} e for E_T^{miss}), and the p_T threshold. Some triggers are prescaled in order to adjust the rate of accepted events: a prescale value of n means that an event that passes the given trigger is retained with a probability of $1/n$.

Each chain targets a specific physics signature and will be used by a physics analysis to recover events with the desired topology. A share of the rate budget is assigned to each chain according to the physics goals of the collaboration, and the threshold requirements of the L1 and HLT triggers are set to keep the expected rate within this budget. The list of trigger chains forms the *trigger menu*. The most significant constraints on the trigger menu design during Run 2 were the limits on the L1 and HLT output rates of 100 kHz and 1kHz, respectively. The design of the Run 2 trigger menu was shaped by the goal of maintaining the unprescaled single-electron and single-muon trigger p_T thresholds around 25 GeV, in order to preserve the trigger efficiency for events with W and Z boson leptonic decays. The trigger menu was adjusted several times during the course of Run 2 in response to changes in LHC bunch filling patterns and bunch intensities, which affected the peak luminosity and average number of pileup interactions. The physics trigger menu and operations for 2015 data-taking can be found in Ref. [98].

Different types of triggers and trigger menus exist. The *primary* triggers are used to select events of interest for physics analyses and are usually unprescaled. These cover all the signatures relevant to the ATLAS physics program, such as electrons, photons, taus, muons, jets, E_T^{miss} , and b-jets, all necessary for SM measurements and BSM searches. Other examples are *calibration* triggers, which store only partial event information while operating at high rates, and *support* triggers, which are used for monitoring and are usually prescaled. For trigger algorithm development and rate predictions, a special menu is used called *minimum bias* [107]. To estimate the rate of events that would pass any given trigger one needs an unbiased data sample, such as the one collected by a trigger that fires at random. However, most selections in ATLAS are interested in rare events with small cross sections, and require

some level of event activity, such as a high p_T lepton or jet. To reduce the amount of data necessary to have sufficient statistics for these rare events, a minimum bias sample is used. This is obtained by using a collection of several L1 trigger items targeting various signature types. The resulting sample re-introduces some bias for these harder events, while still not favoring any particular signature, and results in a mixture of soft and hard processes, with soft events dominating. Because the correlation between the triggers is preserved, a re-weighting of the events allows to recover a zero-bias sample.

4.3.4 The Phase I trigger upgrade

The ATLAS Phase I upgrades [76, 108] were installed before the start of Run 3 in order to cope with the concomitant LHC upgrade during the second Long-Shutdown (LS2) and as a first step in preparation for the HL-LHC. The LS2 upgrades to the injection system allow for lower emittance and higher intensity bunches. The full beam intensity attainable with these upgrades will be usable only after the final upgrades for the HL-LHC, mainly because of heating limitations of the inner-triplet magnets. Nonetheless, the improvements will provide 60% more intense beams already in Run 3 and luminosity leveling will allow to remain at a peak luminosity of approximately $2.4 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$ for up to 10 hours during an LHC fill, increasing the average pileup to $\langle \mu \rangle \approx 60 - 70$ [76]. In order to sustain the higher rates and radiation conditions, several detector systems were upgraded, including significant upgrades to the LAr calorimeter electronics to provide finer granularity and energy resolution to the trigger system, and the New Small Wheel (NSW) muon detector that replaced the inner end-caps of the muon spectrometer. The TDAQ system had to be upgraded to adapt to the new detectors, as well as to handle the higher event rates and pileup levels, both in terms of resources and algorithm performance. In particular, the DAQ system had to handle a 30% larger event size at L1 (2.1 MB at $\langle \mu \rangle \approx 60$), while the latency and output rate of the L1 trigger were fixed by the original specifications of the detector to 100 kHz. At the same time, the HLT system had to target an output rate to disk of 3 kHz and the DAQ system had to sustain a maximum throughput of 8 GB/s, a factor of two improvement in performance with respect to Run 2 [76]. The increased pileup levels were expected to degrade the calorimeter resolution and object isolation, which would result in a decreased trigger efficiency and higher rates, pushing the trigger thresholds up. In order to retain the physics reach in the near-threshold regime, a more refined data processing was necessary, obtained via improved trigger algorithms with access to higher granularity. The Run 3 TDAQ system is shown in Fig. 4.15. In the following, the upgrades to the L1Calo jet trigger algorithms are discussed, as these are relevant for this work.

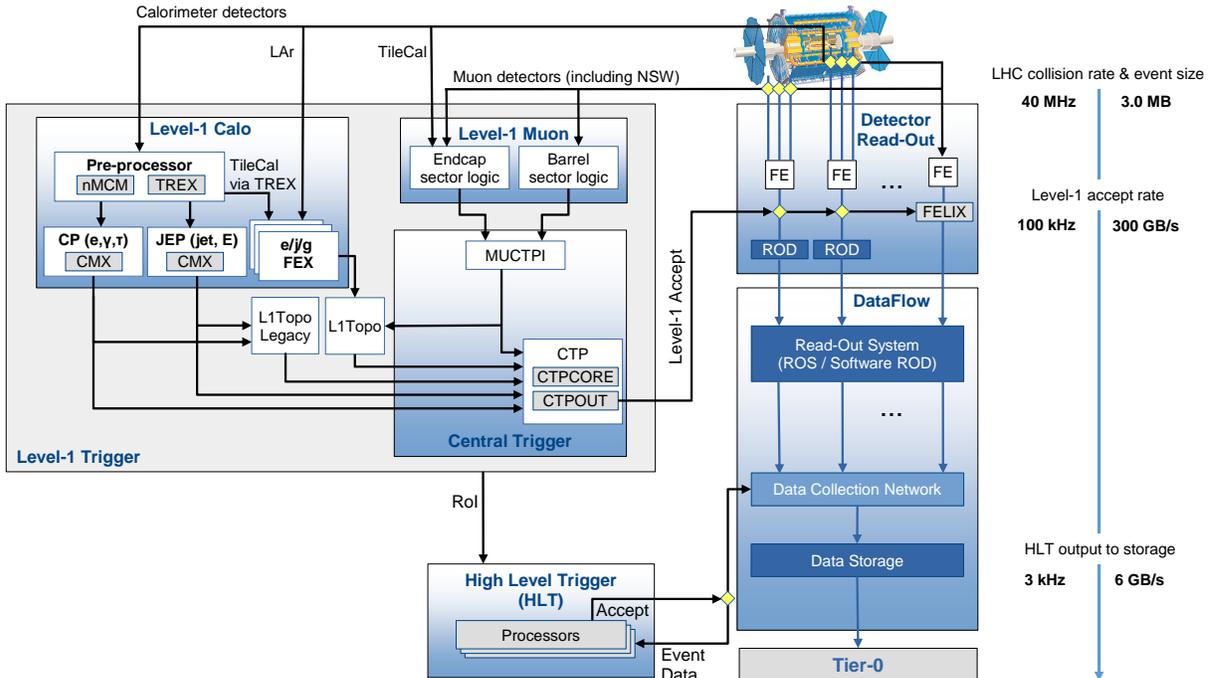


Figure 4.15: Schematic view of the Trigger and Data Acquisition system at the beginning of Run 3 [76].

L1Calo

The Run 3 L1Calo system was equipped with new Feature EXtraction (FEXs) algorithms running on FPGA modules and with access to a finer granularity calorimeter information. In place of the 0.1×0.1 trigger towers of Run 1 and 2, the LAr processing system now sends the information along the trigger path in the form of *Super Cells* containing sums of four or eight calorimeter cells (the maximum granularity of the detector front-ends). Fig. 4.16 shows an example of a 0.1×0.1 trigger tower in the EM Barrel calorimeter, now containing ten Super Cells. Different FEX algorithms are used to reconstruct different TOBs. The electron feature extractor (eFEX) module performs e/γ and hadronic τ identification, with coverage limited to the tracking acceptance of $|\eta| < 2.5$. The jet feature extractor (jFEX) system identifies small and large-radius jets in the region $|\eta| \leq 4.9$, hadronic τ decays in the region $|\eta| \leq 2.5$, and electrons in the forward region outside the eFEX acceptance. It also computes energy sums and applies pileup and noise subtraction cuts. Lastly, the global feature extractor (gFEX) processes data from the entire calorimeter on a single module and performs full-scan algorithms to identify large-radius jets with pileup suppressed energies

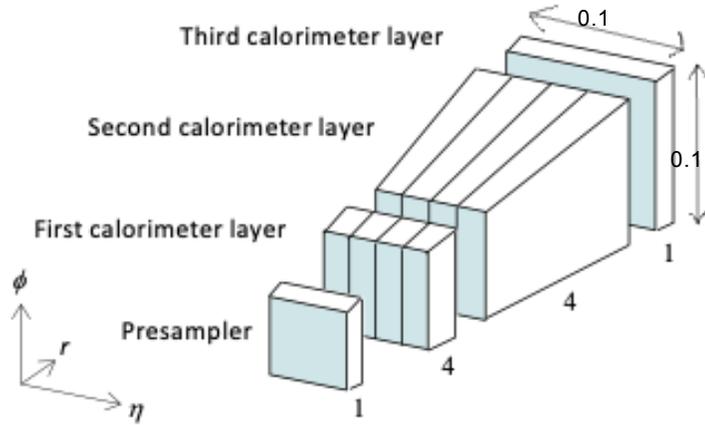


Figure 4.16: Example of an EM barrel 0.1×0.1 trigger tower containing ten Super Cells [76].

and global observables, such as E_T^{miss} . The full Super Cell granularity is available only to the eFEX algorithms, while the jFEX and gFEX systems have access to 0.1×0.1 towers, still an improvement with respect to the 0.2×0.2 resolution of the jet elements in Run 2. A brief summary of the jFEX small-R jet algorithm is given next, while a detailed description of all the FEX algorithms can be found in Ref. [76].

The jFEX small-radius jet algorithm

The calorimeter inputs to the jFEX algorithm are 0.1×0.1 trigger towers in the region $|\eta| < 2.5$, with slightly coarser granularity in the end-cap and forward regions. Each jFEX module covers an η slice of the calorimeter while providing full ϕ coverage. Each of the four FPGAs in a module is assigned a slice in $\eta \times \phi$, with overlap areas between FPGAs to correctly handle objects located on the edges.

The jFEX small-radius jet algorithm is a sliding-window algorithm, with the main steps shown in Fig. 4.17. The search window consists in 5×5 trigger towers (0.5×0.5 in $\eta \times \phi$). The seeds are constructed as the sum of 3×3 tower blocks centered on each tower in the search window. Comparative operators, which take care of the possibility of comparing equal digital values, are used to find the seed with the maximum energy in the search window. The tower at the center of the maximum energy seed is chosen as the center of the jet. The already computed energy sum inside the seed (shown in red in Fig. 4.17) is added to the energy ring including all the towers within a radius of $0.2 \leq R < 0.4$ (shown in purple). The final jet consists of 45 towers forming an approximately round shape of $R = 0.4$.

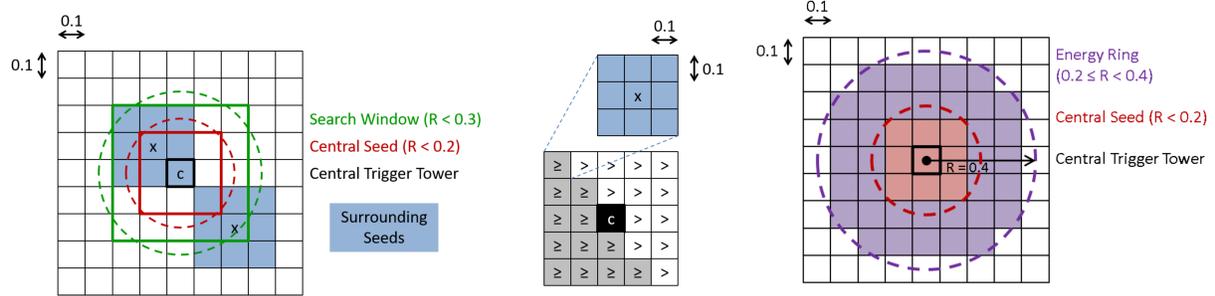


Figure 4.17: The jFEX small-radius jet algorithm. From left to right: the seed finding process with identification of local maxima; comparative operators used to identify a local maxima; the final small- R jet centered on the trigger tower at the center of the maximum energy seed and built with all the towers within $R < 0.4$ [76].

4.3.5 The Phase II trigger upgrade

The HL-LHC will run at a peak instantaneous luminosity of 5 to $7.5 \times 10^{-34} \text{cm}^{-2} \text{s}^{-1}$ and is expected to collect between 3000 to 4000fb^{-1} of data, ten times the entire data set collected up until Run 3. This will allow the ATLAS experiment to substantially extend its physics program, by opening up the possibility of high precision measurements of SM observables and giving access to previously prohibitively small cross sections. At the same time, the increase in luminosity will result in unprecedented levels of radiation and pileup, with up to 200 simultaneous pp interactions per bunch crossing. This extreme environment will pose new constraints on the ATLAS detector and TDAQ systems and will require extensive upgrades that will be installed during the LS3, referred to as Phase II upgrades. The main detector upgrades will be described here briefly, but one is referred to the corresponding Technical Design Reports for more information. The current ID will be fully replaced by the ITk, which will extend the η -coverage up to $|\eta| = 4.0$ (compared to the current $|\eta| = 2.5$). A new detector, the High-Granularity Timing Detector [109], will be added between the ITk and the LAr end-cap calorimeter to provide precision hit-timing information to aid with pileup mitigation and luminosity measurement. The LAr [110] and tile calorimeters readout electronics will be upgraded to improve the current limitations on the L1 trigger latency. Similarly, the MDTs front-end electronics will be replaced to handle the higher rates and provide MDTs hit information to the first step of the trigger chain. The rest of the MS upgrade will focus on upgrading the electronics of the RPC and TGC trigger chambers and adding new RPC detectors to increase the solid angle coverage.

The Phase II TDAQ upgrade [10] is required to adjust to the new detector systems and to the harsher data taking conditions. Without an upgrade of the TDAQ system, the high

levels of pileup would significantly degrade the performance of the current trigger algorithms. The larger backgrounds would also result in higher trigger thresholds to keep the rates under control, which would reduce the sensitivity of physics analyses, as discussed more in detail in Chap. 8. An upgrade of the trigger is necessary to retain ATLAS physics goals, which are summarized in Fig. 4.18, together with the triggers and hardware systems required to achieve these goals. For instance, the Global Trigger (see below) is necessary for improved multi-jet triggers needed to achieve sensitivity to non-resonant $HH \rightarrow b\bar{b}b\bar{b}$. The architecture of the

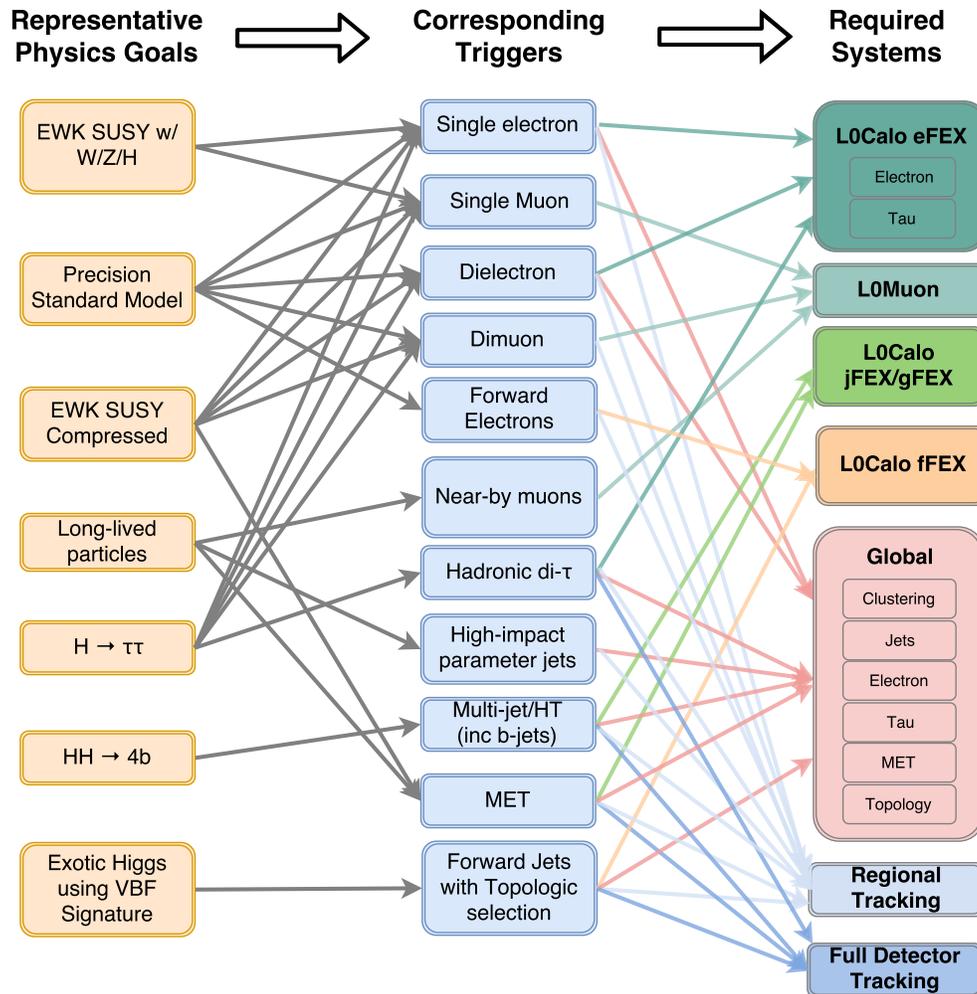


Figure 4.18: Diagram showing the relationship between ATLAS physics goals, required triggers, to the related trigger components of the Phase II trigger system [10].

Phase II TDAQ system is shown in Fig. 4.19. The trigger will still be a two-level system, with a first hardware-based trigger, now called *Level-0* (L0), and a second software-based trigger, now named *Event Filter* (EF). The DAQ system will handle the data flow from the detector electronics, through the trigger chain, up to permanent storage. The L0 trigger will still receive data at the LHC event rate of 40 MHz, but the new detector readout electronics

will allow a L0 output rate of 1 MHz (from the 100 kHz of Run 3). The HLT will also have an increased output rate of 10 kHz (from the 3 kHz of Run 3). With a predicted event size of 6 MB, the total output bandwidth will be 60 GB/s.

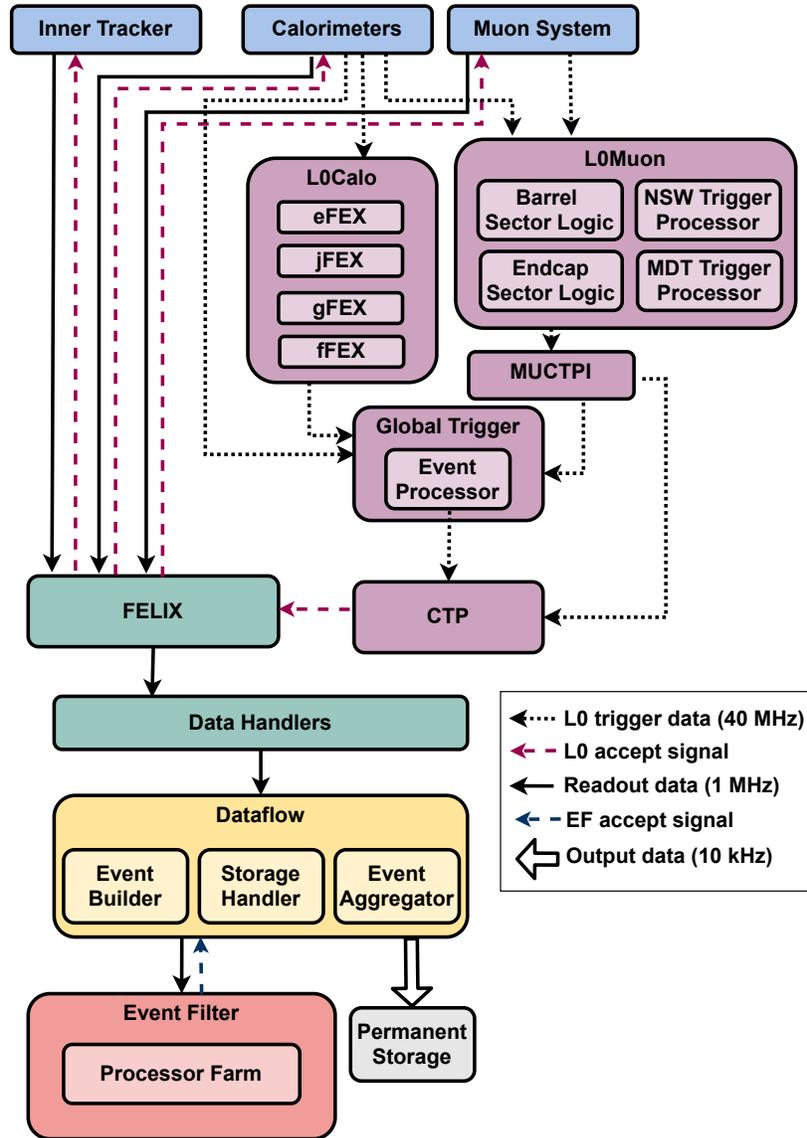


Figure 4.19: Design of the Phase II TDAQ system with its three main systems: L0 Trigger, DAQ system, and Event Filter. The black dotted lines indicate the data flow at 40 MHz from the detectors to the L0 trigger system, which must produce a trigger decision within $10\mu\text{s}$. The red dashed arrow indicate the flow of the L0 trigger decision. The solid black arrows represent the detector and trigger data being transmitted through the DAQ system at 1 MHz. The EF makes the second level trigger decision reducing the event rate to 10 kHz. On EF-Accept events are transferred to permanent storage [111].

The Level-0 Trigger

The L0 trigger will be composed of the L0Calo, L0Muon, MUCTPI, and the CTP, inherited from the current trigger system, and a new addition, the Global Trigger.

The L0Calo and L0Muon sub-systems will receive reduced granularity information at 40 MHz from the calorimeter and muon detectors, respectively. They will be mostly similar to their Phase I predecessors. The L0Calo will run the FEX algorithms described in the previous section, with the addition of a forward FEX (fFEX) for reconstruction of forward electromagnetic (jet) objects in the region $2.5 < |\eta| < 4.9$ ($3.2 < |\eta| < 4.9$). The L0Muon sub-systems will receive inputs from all the muon detector systems and the Tile calorimeter. New additions will be the inclusion in the trigger decision of precision MDTs momentum measurements and signals from the RPC inner stations. The MUCTPI will calculate multiplicities of high energy muons, check for double-counting of muon candidates, and interface the L0Muon with the Global Trigger (GT) and the CTP.

The GT will be an entirely new addition to the trigger system that will bring EF-like capabilities to the L0 trigger by running offline-like algorithms on custom FPGA hardware. The GT will have access to full granularity calorimeter data, as well as TOBs from the L0Calo and L0Muon. The refined TOBs produced by the GT will be available as input to topological algorithms, as the GT will replace and extend the functionalities of the L1Topo system. The new TOBs and trigger conditions will be sent to the CTP for evaluation of the final trigger decision. The development of firmware algorithms for the GT is a major part of the work presented in this thesis, so the GT will be discussed in detail in Sec. 8.1.

The Event Filter

The EF system will still consist in a large CPU-based processing farm running offline-like reconstruction algorithms. Most importantly, the EF will have access to tracking information, which will allow to perform track reconstruction and implement vertex-finding and particle-flow-like algorithms to significantly reduce the rates by improving pileup mitigation, the identification of b -jets, and the E_T^{miss} calculation. A first fast initial rejection will be provided by regional tracking based on TOBs received from the L0 trigger. The reduced event rate will then be input to a global tracking performed over the full ITk detector. Note that the plan for the EF tracking Phase II upgrade has evolved since the original Technical Design Report [10], which was superseded by Ref. [111].

4.4 ATLAS Event reconstruction

The reconstructed final state of a collision in ATLAS includes electrons, photons, muons, τ -leptons, jets, and missing transverse energy. Except for muons, reconstruction of all the other objects requires calorimeter information. Fig. 4.20 shows the paths that different types of particles follow in the detector systems. Charged particles, such as electrons, protons, and muons, leave curved tracks in the inner detector (ID). Thanks to the solenoidal magnetic field, the particles are bent and the direction and radius of curvature of the tracks provides charge and momentum information. Neutral particles, like photons and neutrons, do not interact with the ID. Electromagnetic interacting particles, like electrons and photons, are stopped in the electromagnetic calorimeter, while hadronically decaying particles deposit most of their energy in the hadronic calorimeter, where they are stopped after a longer and wider shower. Muons interact with the ID but, behaving as minimum ionizing particles, usually escape the calorimeters and leave tracks in the muon spectrometer bent by the toroidal magnetic field. Neutrinos escape the detector volume undetected, but their presence in the event is inferred from a momentum imbalance on the transverse plane. In the following, the

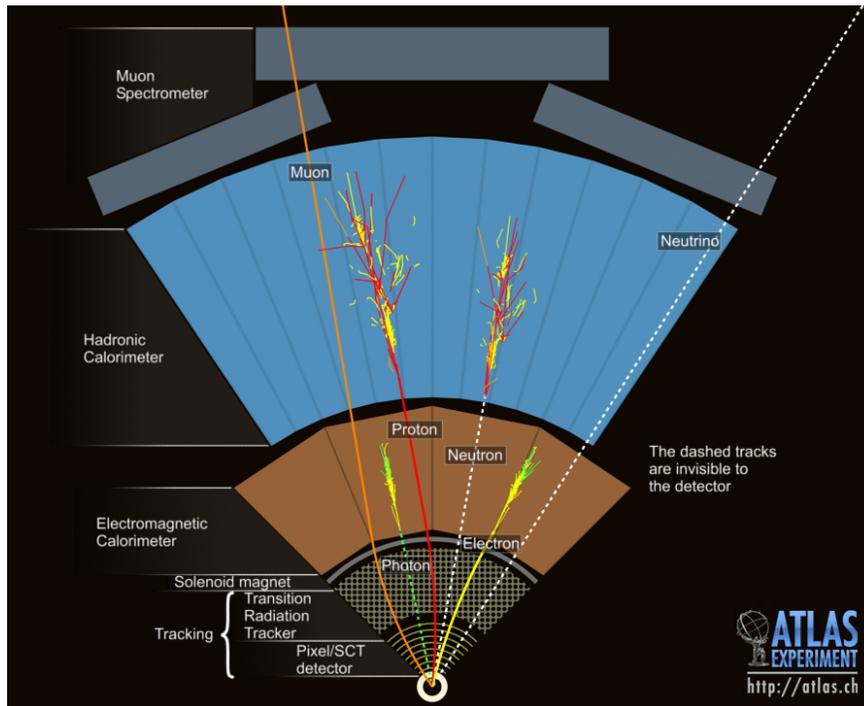


Figure 4.20: Cross-section of the ATLAS detector with simulated particles trajectories [112].

reconstruction of the physics objects relevant for this thesis is briefly reviewed. Jets will be treated separately and more extensively in the next chapter. The algorithms discussed here

are developed by the ATLAS Combined Performance groups, which provide working points (WPs), calibrations, and general recommendations for all physics analyses.

4.4.1 Tracks and vertices

The track reconstruction [113] algorithm reconstructs the trajectory of charged particles from the electronic signals, or *hits*, left in the ID. Tracking is a pattern recognition task made more difficult by the busy environment of the ID, including in-time and out-of-time pileup, and the possibility of collimated tracks. Track reconstruction starts from seeds made of tracks with three hits recorded in the Pixel or the SCT detectors. The seeds are then extended to include further hits to create track candidates. An ambiguity resolution step removes overlaps or wrongly assigned hits. Finally, a χ^2 -based track fit is performed and only tracks with $p_T > 400$ MeV and passing quality selection criteria are retained. The final track is specified by the collection of hits assigned to it and the associated parameters describing the particle's trajectory: the transverse and longitudinal impact parameters d_0 and z_0 , the azimuthal and polar angle ϕ and θ , and the charge to momentum ratio $\frac{q}{p}$.

The tracks are also used to reconstruct the *primary vertices* [114], by iteratively associating the reconstructed tracks with $p_T > 500$ MeV. Primary vertex candidates are required to have at least two reconstructed tracks with $p_T > 500$ MeV and to be compatible with the interaction region. The *hard-scatter* vertex is the vertex with the highest p_T^2 sum of the tracks associated to it. The other primary vertices are assumed to be produced by in-time pileup. Different track and track-vertex-association quality criteria WPs are provided.

4.4.2 Electrons

Several analyses, including the one discussed in this thesis, rely on the efficient identification of prompt electrons originating from decays of W and Z bosons, from electrons produced by photon conversions, misidentified hadrons, and non-isolated electrons from heavy-flavor decays. Electrons and photons travel through the ID, where only electrons leave tracks, and are then stopped in the EM calorimeter (EMCal). Almost 40% of photons convert to electron-positron pairs (*converted photons*). An electron can lose energy through bremsstrahlung radiation due to the interaction with the different detector materials, with the radiated photon also possibly decaying to an electron-positron pair. These interactions can occur already in the beam pipe or in the ID, producing multiple tracks in the ID, or they can occur in the EMCal, where they are contained in the EM shower. A schematic view of the path of an electron traveling through the detector is shown in Fig. 4.21.

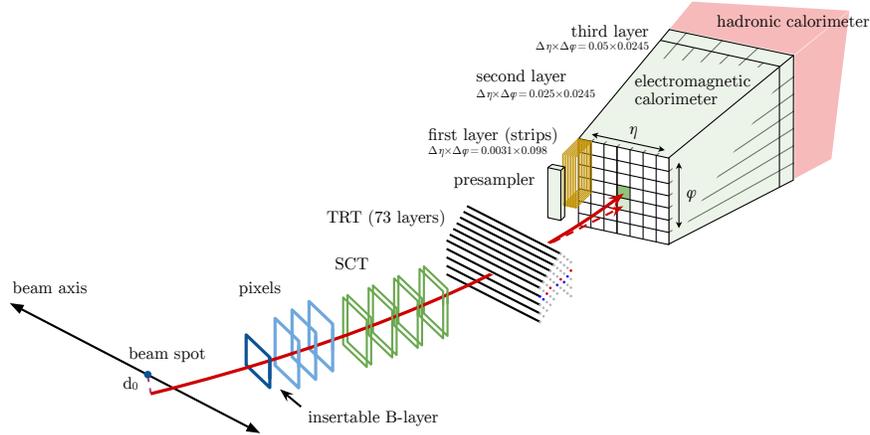


Figure 4.21: Illustration of the trajectory of an electron through the detector (solid red), with a photon emitted via bremsstrahlung radiation (dashed red) [115].

In 2015 and 2016 electrons and photons were reconstructed using a sliding window algorithm seeded by calorimeter towers [115]. A new algorithm [116, 117] based on topological clusters was introduced in 2016. The variable-size topoclusters, as opposed to fixed-size towers, are better fit to capture the dynamic shape of the EM shower, subject to bremsstrahlung photon emission and photon conversions. The algorithm starts by selecting the subset of the 4-2-0 topoclusters (as described in Sec. 4.4.4) that are primarily generated by showers in the EM calorimeter, by requiring the EM fraction $f_{EM} > 0.5$. A set of the EM clusters is selected as seeds of possible electrons and photons, and superclusters are formed by associating nearby EM clusters that originate from the same vertex, in the case of an electron and a bremsstrahlung photon, or that originate from a displaced vertex, in the case of a converted photon. In general, an electron is defined as a supercluster in the calorimeter matched to a track in the ID; a converted photon as a supercluster in the calorimeter matched to a conversion vertex; and an unconverted photon as a supercluster matched to neither a track nor a vertex. The reconstructed electrons are further cleaned via quality criteria based on a likelihood discriminant. Four sets of electron identification criteria with increasing background rejection power are provided: *VeryLoose*, *Loose*, *Medium*, and *Tight*. Isolation requirements are also defined to suppress background from hadrons faking electrons.

4.4.3 Muons

Muon reconstruction [118] is based on detector information from the muon spectrometer (MS), the inner detector (ID), and the calorimeter. The primary reconstruction strategy looks for reconstructed tracks in the MS, which are then matched to ID tracks. A combined fit of the MS and ID tracks, which takes into account the energy loss in the calorimeter, gives

the final *combined muon*. Other reconstruction strategies are available to retain efficiency. *Inside-out* muons are reconstructed from extrapolating ID tracks into the MS, where they are required to match with three MS hits included in the final fit. This allows to recover efficiency in regions of low MS acceptance or for low p_T muons. *Muon spectrometer extrapolated* tracks are reconstructed from only MS tracks and are used to extend the acceptance outside the $|\eta| < 2.5$ region covered by the ID. *Segment-tagged* muons are reconstructed from ID tracks that satisfy tight matching requirements on hits in the MS, but only the ID information is used to obtain the muon parameters. Lastly, *calorimeter-tagged* muons are reconstructed by matching ID tracks to energy depositions in the calorimeter consistent with a minimum-ionizing particle signature, and compensate for the MS inefficiency in the $|\eta| \sim 0$ gap region.

After reconstruction, identification criteria are applied to select the highest quality tracks. Muon candidates are separated into prompt muons originating from the interaction vertex, and non-prompt muons originating from secondary decays. The WPs used in ATLAS are, in order of decreasing efficiency and increasing purity of prompt muon selection: *Loose*, *Medium*, and *Tight*. Special WPs are provided for analyses targeting more exotic regions of phase space. These are the *Low - p_T* and *High - p_T* WPs, respectively.

4.4.4 Topological clustering

Topological clusters, or *topoclusters*, are clusters of topologically connected calorimeter cells that are used for the reconstruction of isolated hadrons, jets, and E_T^{miss} . Each topocluster is three-dimensional, thanks to the longitudinal segmentation of the sampling layers, and can contain the full or partial response to one or multiple signal particles. The *topoclustering* algorithm [95] starts by evaluating the significance of each cell,

$$\mathcal{S} = \frac{E_{\text{cell}}^{EM}}{\sigma_{\text{cell noise}}^{EM}}, \quad (4.4)$$

where $\sigma_{\text{cell noise}}$ is estimated for each run according to Eq. (4.3). Both the cell energy and noise are evaluated at the EM energy scale, which is the scale at which photon and electron energy depositions are reconstructed correctly. The algorithm proceeds by identifying the seed cells, defined as those cells with $|\mathcal{S}| > 4$. Each seed cell represents a protocluster, which is progressively grown in volume. For each protocluster, the algorithm finds all the neighboring cells (cells adjacent to the seed either in the same sampling layer, or in adjacent layers and overlapping in the (η, ϕ) plane) with $|\mathcal{S}| > 2$. These cells are added to the protocluster and the step is repeated until no cells with $|\mathcal{S}| > 2$ adjacent to the protocluster are left. If a cell with $|\mathcal{S}| > 2$ is assigned to two protoclusters, the protoclusters are merged.

Lastly, an outer layer of cells adjacent to the protocluster and satisfying $|\mathcal{S}| > 0$ is added. The resulting clusters have a high \mathcal{S} core, which differentiates them from background noise, while the softer outer layer allows to retain signals that are closer to the noise level. A representative simulation of the three stages of the clustering process is shown in Fig. 4.22.

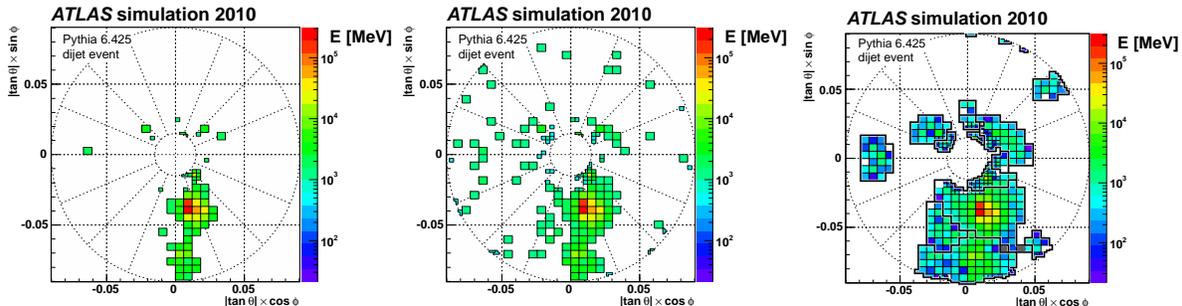


Figure 4.22: Stages of topological clustering in the first FCal layer for a simulated di-jet event with at least one jet entering the FCal and no pileup. From left to right: all seed cells with $|\mathcal{S}| > 4$ starting a protocluster; all neighboring cells with $|\mathcal{S}| > 2$ are added recursively to the protocluster; all neighboring cells with $|\mathcal{S}| > 0$ are included. Topocluster fragments not associated to a seed are seeded in a surrounding calorimeter layer [95].

Due to the shaping of the calorimeter signal, it is possible for calorimeter cells to have negative energy signals if induced by out-of-time pileup that occurred 100 ns before the event. Out-of-time pileup can also cause positive energy signals, when this comes from collisions in closer bunch crossings. It is therefore desirable to include negative energy cells in the clustering process for these positive and negative noise fluctuations to cancel each other out, providing an implicit noise suppression. However, this can result in negative energy clusters, especially when the seed itself was a large negative energy cell. Negative energy clusters are not used as input to jet reconstruction, as they represent pileup-induced energy fluctuations with no real correlation with the particle that is being reconstructed.

The kinematics of the final clusters are obtained from a sum of the four-vectors of the associated cells. Including the negative energy cells would distort the calculation, to the point of projecting clusters to the opposite side of the detector, while not including them would result in a bias from the positive fluctuations. A special recombination scheme is therefore used that includes all cells, but avoids biasing in either direction and is described in Ref. [95]. Once the basic kinematic variables ($\eta_{\text{clus}}, \phi_{\text{clus}}, E_{\text{clus}}^{EM}$) are calculated, the final four-vector is obtained by interpreting the topocluster as a massless pseudo-particle.

At this point the energy of the topoclusters is still at the EM scale, which does not account for the non-compensating calorimeter response to hadrons. The topoclusters need therefore to be calibrated to properly represent the hadronic energy scale. The calibration compensates also for the inefficiency due to the loss of low \mathcal{S} signal clusters because of

the pileup-dependent clustering strategy. The calibration is referred to as *Local hadronic Cell Weighting* (LCW). It consists in a series of corrections to iteratively reweight the cells energy, and is performed using simulation of neutral and charged pions, representative of electromagnetic and hadronic showers, respectively. The final calibrated cluster energy is typically $E_{\text{clus}}^{\text{LCW}} \geq E_{\text{clus}}^{\text{EM}}$.

4.4.5 Missing transverse momentum

Missing transverse momentum (E_T^{miss} or MET) [119] is an important proxy to identify the production in the hard scatter of stable weakly interacting particles, which escape the experimental volume without leaving any detectable signal and include neutrinos, as well as possible new BSM particles. Indicating the contributions from all the observable electrons, photons, taus, muons, and jets, and the non-observable (invisible) particles, the vectorial sum of the transverse momenta of all the objects emerging from the hard scatter \mathbf{p}_T^{HS} is calculated as,

$$\mathbf{0} = \mathbf{p}_T^{\text{HS}} = \underbrace{\sum \mathbf{p}_T^e + \sum \mathbf{p}_T^\gamma + \sum \mathbf{p}_T^\tau + \sum \mathbf{p}_T^\mu + \sum \mathbf{p}_T^{\text{jet}}}_{\mathbf{p}_T^{\text{obs}}(\text{observable})} + \underbrace{\sum \mathbf{p}_T^\nu}_{\mathbf{p}_T^{\text{inv}}(\text{not observable})}. \quad (4.5)$$

By conservation of momentum in the transverse plane, any significant deviation from zero indicates the presence of a particle that eluded detection with transverse momentum $\mathbf{p}_T^{\text{inv}} = -\mathbf{p}_T^{\text{obs}}$. In practice, due to limitations of the detector acceptance and experimental inefficiencies in the reconstruction of the hard objects, only a proxy of $\mathbf{p}_T^{\text{obs}}$ can be measured, referred to as $\mathbf{E}_T^{\text{hard}}$ and which includes only the reconstructed objects that pass kinematic selection and reconstruction quality criteria. In general $\mathbf{E}_T^{\text{hard}} < \mathbf{p}_T^{\text{obs}}$. To partially recover this loss, an additional soft-term $\mathbf{p}_T^{\text{soft}}$ is included, built from reconstructed charged-particle tracks coming from the hard-scatter vertex, but not associated to any hard object⁵. As the hard objects are reconstructed and calibrated independently, it is possible that different objects share energy contributions, such as a topocluster contributing both to a jet and to an electron. For this reason, a *signal ambiguity resolution* procedure is implemented. The missing transverse momentum observable $\mathbf{E}_T^{\text{miss}} = (E_x^{\text{miss}}, E_y^{\text{miss}})$ is then calculated as,

$$\mathbf{E}_T^{\text{miss}} = -(\mathbf{p}_T^{\text{hard}} + \mathbf{p}_T^{\text{soft}}), \quad (4.6)$$

⁵A less commonly used definition calculates the soft-term from the unmatched topoclusters in the calorimeter, which includes neutral particles, but suffers from a large residual dependence on pileup.

with magnitude and azimuthal angle,

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}, \quad \phi^{\text{miss}} = \tan^{-1} \left(\frac{E_y^{\text{miss}}}{E_x^{\text{miss}}} \right). \quad (4.7)$$

Important quantities used to estimate the event hadronic activity are $\sum E_T$, the scalar sum of the transverse momenta of all the hard and soft contributions to the E_T^{miss} calculation, and H_T , the scalar sum of the transverse momenta of all the hard objects only.

In practice fake E_T^{miss} can arise due to limited detector acceptance, signal fluctuations in the detector response, and fluctuations in pileup contribution. The level of agreement between the observed non-zero E_T^{miss} value and the hypothesis of true E_T^{miss} is given by the significance \mathcal{S} . This is calculated with respect to the event activity as,

$$\mathcal{S} = \frac{E_T^{\text{miss}}}{\sqrt{H_T}} \quad \text{or} \quad \mathcal{S} = \frac{E_T^{\text{miss}}}{\sqrt{\sum E_T}}. \quad (4.8)$$

Another more recent object-based definition [120] calculates the significance as a likelihood ratio to test the hypothesis $\mathbf{p}_T^{\text{inv}} = 0$ and $\mathbf{p}_T^{\text{inv}} \neq 0$, and is the one used in the analysis discussed in this thesis to select events with true neutrinos.

4.4.6 *b*-tagging

The identification of jets⁶ containing *b*-hadrons is an important step in ATLAS physics [121, 122], as top quark and Higgs boson decays proceed almost exclusively via bottom quarks. Jets originating from *b*-quarks can be identified by exploiting the distinct features of such decays. A *b*-quark hadronizes into a *B* meson – a meson composed of a *b*-quark and a *u*-, *d*-, *s*-, or *c*-quark. The lifetime of a *B*-meson is of the order of 1.5 ps ($\langle c\tau \rangle \sim 4.5$ mm), which corresponds to a mean flight length of $\langle l \rangle = \gamma\beta c\tau$ before decaying. At LHC energies, this is of the order of a centimeter [123], a sizable distance observable in the ID as a displaced vertex: a certain number of tracks points to a secondary vertex, with large longitudinal and transverse impact parameters. The decay of the *B*-meson is well described by the decay of the *b*-quark inside the hadron (spectator model), which proceeds predominantly via $b \rightarrow cW^-$, with the virtual *W* decaying either leptonically into $l\bar{\nu}$, or into a pair of quarks, which then hadronizes. The transition $b \rightarrow c$ is favored for the hadronic decay path by the CKM matrix, so that hadronic decays of *B*-mesons typically produce at least one *c*-flavoured hadron (a *D* meson), which then decays further, also with an appreciable lifetime, resulting in a characteristic topological configuration with two secondary vertices.

⁶Jets are discussed in detail in Chap. 5

The tagging of b -jets in ATLAS relies on the track reconstruction of the displaced B -meson decay and is a two-stage approach. In the first step, a series of low-level algorithms [124] exploits the characteristic features of the decay: the IP2D and IP3D track-based impact parameter taggers; the SV1 secondary vertex reconstruction algorithm; and **JetFitter** algorithm for a topological reconstruction of the full b - and c -hadron decay chain. The discriminating variables produced by these algorithms provide complementary information and are used in the second stage as inputs to the DL1r[122] algorithm⁷, a high-level tagger which includes as input also the output probabilities from the RNNIP algorithm [125]. The algorithm output is multidimensional and provides the probability of the jet to be a b -jet (p_b), a c -jet (p_c), or a light-flavor jet (p_{light}), with the final b -tagging discriminant

$$D_{\text{DL1r}} = \ln \left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{\text{light}}} \right), \quad (4.9)$$

where f_c gives the percentage of c -jets in the background hypothesis and can be optimized at the physics analysis level. Different WPs at fixed signal efficiency are provided.

⁷Historically, two high-level taggers were available: the MV2 boosted decision tree classifier, and the DL1 artificial neural network. The DL1 algorithm was introduced for Run 2 and has now evolved into the DL1r. The latter achieves the best tagging performance and is the current recommendation for physics analyses.

Chapter 5

Hadron collider physics

Thanks to the unprecedented center-of-mass (CoM) energies of the LHC, the ATLAS and CMS experiments can probe the SM over scales ranging from 10 GeV up to 10 TeV. The vast experimental reach relies on two fundamental principles of collider physics [126]: i) the higher the energy, the smaller the length scale one can probe, according to de Broglie relation $\lambda = \frac{h}{p}$; and ii) particles interacting at high energies should enable the production of heavier particles, according to Einstein's equation $E = mc^2$. At the same time, understanding the final states of these high-energy collisions is challenging both theoretically and experimentally. In particular, the evolution of hadron-hadron collisions is tightly connected to the nature of the QCD interaction and its running coupling.

When two protons collide, a hard scattering event — an event with a large momentum transfer — will involve only one parton from each proton. At the energy scale of the hard scatter, QCD can be treated as a perturbative quantum field theory and the matrix element for any hard process can be calculated systematically at fixed order using the standard Feynman diagrammatic techniques. The hard process results in the production of a few energetic or heavy particles — whether quarks, leptons, or bosons — and, if these are short-lived, their resonant decays. These particles usually represent the process of interest that one would like to study. However, on top of this hard process several effects related to QCD have to be taken into account.

The primary partons from the hard scatter will have a non-zero probability to split further into mostly soft and collinear gluons and quarks, resulting in a parton shower. The evolution produces progressively softer and smaller angles partons, down to a scale where QCD becomes non-perturbative. At this point, when the momentum transfers are small and the QCD running coupling is large, hadronization occurs: the connected partons combine into color singlet states, with unstable hadrons decaying further. The final stable hadrons (with lifetimes $\tau > 10$ ps) are the physical particles that interact with the detector. These neutral and charged hadrons are stopped in the hadronic calorimeter, leaving a cone like energy deposition that is reconstructed as a jet. For an experimentalist, jets are one of the main means to gain insight into what happened in the hard scatter. As such, they are part

of almost any physics analysis and are fundamental for the study of the SM, as well as for the search of new BSM phenomena. For a theorist, jets offer a rich playground where to test QCD predictions at high energies. Additional soft physics arising from interactions between the colliding proton remnants, as well as pileup radiation, can contribute to the final state, making the reconstruction of the event more difficult.

The partons confined in the incoming protons, as well as the hadronization of the final state partons, occur at a much lower scale than the hard scattering process, of the order of 1 GeV, where the validity of QCD as a perturbative theory comes short. These processes cannot be calculated theoretically, and have to be modelled and fit to data. Additionally, the final state of such collisions typically involves hundreds of particles. The high final state multiplicities make matrix element computations in the perturbative regime often too complex to be calculated exactly. Despite these difficulties, theoretical predictions can still be obtained thanks to the factorization of the contributions from the different scales and the use of Monte Carlo methods.

5.1 From QCD to jets

5.1.1 The strong coupling

The strong coupling $\alpha_s = \frac{g_s^2}{4\pi}$ is the fundamental parameter governing QCD interactions. The strong coupling “runs”, meaning that the effective strength of the strong interaction changes with the physics scale Q of the process in question. The running of the strong coupling is governed by the renormalization group equation (RGE),

$$Q^2 \frac{\partial \alpha_s}{\partial Q^2} = \beta(\alpha_s) = -\alpha_s^2 (b_0 + b_1 \alpha_s + b_2 \alpha_s^2 + \dots), \quad (5.1)$$

where $b_0 = (33 - 2n_f)(12\pi)$ and n_f is the number of quark flavors relevant at the given scale. The RGE allows to take the known value of the coupling at a given scale and find the value at any other scale. Numerically, this is often done with respect to the known value $\alpha_s(M_Z^2) = 0.12$, so that Eq. (5.1) is solved as

$$\alpha_s = \alpha_s(M_Z^2) \frac{1}{1 + b_0 \alpha_s(M_Z^2) \ln \frac{Q^2}{M_Z^2} + \mathcal{O}(\alpha_s^2)}. \quad (5.2)$$

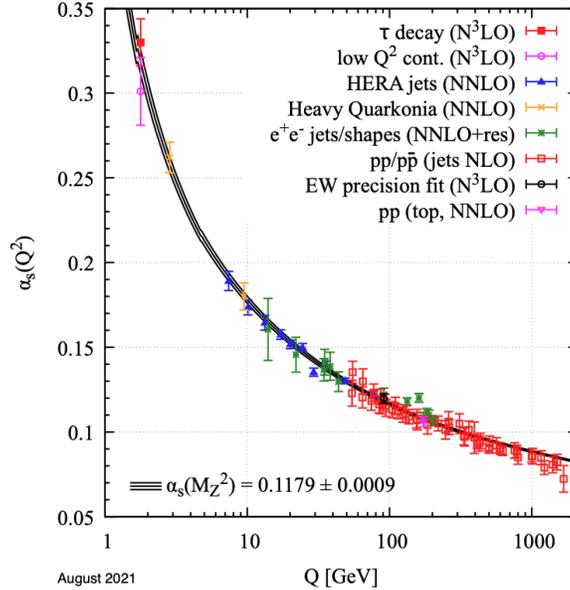


Figure 5.1: Measurements of α_s as a function of the energy scale Q . The order in perturbative QCD used in the extraction of α_s is indicated in parentheses [123].

The negative sign in Eq. (5.2) causes the coupling to decrease with increasing energy, as shown in Fig. 5.1. For large momentum transfers, or small distances, QCD becomes almost a free theory, a phenomenon known as *asymptotic freedom*. In this regime, $\alpha_s \ll 1$ and perturbation theory is valid. Conversely, at small momentum transfers, or large distances, the coupling diverges and, at the scale $\Lambda_{QCD} \approx 200$ MeV, QCD becomes non-perturbative. The fact that the coupling diverges at large distances prevents quarks from ever being observed alone, but only as color singlet bound-states, mesons or baryons. This phenomenon is called *confinement*. In the context of LHC physics, confinement plays a fundamental role in the evolution from the free quarks and gluons produced in the hard scatter to the hadrons actually observed in the detector.

5.1.2 The hard-scatter cross section

While high energy proton collisions involve, by definition, high momentum transfers, the partons confined in the incoming protons interact at a much lower scale – of the order of 1 GeV – where QCD is non perturbative¹. It follows that in an LHC collision there are two scales at play: one is the soft long-distance physics of the proton structure, and the other

¹The proton can be described as a sea of strongly interacting quarks and gluons, where $q\bar{q}$ pairs and gluons carrying a small fraction x of the proton’s momentum are constantly being produced and absorbed. The three quarks that define the hadron type (two up-quarks and a down-quark in the case of a proton) are the *valence quarks*, which can be described, at first approximation, as the quarks whose net number is non-zero.

is the high energy short-distance physics of the hard process. While the latter is calculable in perturbative QCD, the former is too low for perturbative methods to work. This issue is resolved by the *factorization theorem*, which theorizes the independence of the short and long distance physics. The total cross section for the process $pp \rightarrow f$ can be expressed as [127],

$$\sigma_{pp \rightarrow f} = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij \rightarrow f}(x_1 p_1, x_2 p_2, \mu_F^2), \quad (5.3)$$

where p_1 and p_2 are the colliding protons, $\hat{\sigma}_{ij \rightarrow f}$ is the parton-level cross section for the production of the final state f through the initial partons i and j , and the functions $f_i(x_n, \mu_F^2)$ are the PDFs. The PDFs represent (at first approximation) the number density of partons of type i carrying a fraction x_n of the momentum of the proton p_n , when the proton is probed at the factorization scale μ_F .

Parton distribution functions

The factorization theorem can be intuitively understood by the fact that the hard interaction occurs over a much shorter timescale than the fluctuations inside the proton structure, so that from the point of view of the hard scatter, the quark sea appears frozen [128]. The PDFs are in fact decoupled from the short distance physics and their shape can be treated as universal, or process-independent. Therefore, although not calculable from first principles, the PDF shape as a function of x can be modelled and constrained by fitting cross sections to experimental data. Once this is performed at a given μ_F , the result can be derived for a different scale by renormalization group evolution². Fig. 5.2 shows the PDFs behavior of gluons and sea-quarks inside the proton as a function of x , for $Q = 10 \text{ GeV}$ and $Q = 100 \text{ GeV}$, where μ_F is taken to be equal to Q . While the valence up and down quarks carry a significant portion of the proton momentum, at high Q the sea-quarks and gluon contributions become enhanced, even if with smaller x values. The strong enhancement of gluon PDFs towards low x at increasing Q is particularly relevant for LHC physics: as can be observed in Fig. 4.2, cross sections for gluon-initiated processes have a steeper slope of increase with increasing CoM energy than quark-initiated ones [126]. It should be noted that different collaborations use different functional forms for the PDFs and may also constraint the fit using different datasets. In the context of ATLAS physics, PDF modelling contributes to the systematic uncertainties of many analyses.

Cross section

The cross section of an interaction is calculated using two main ingredients: the matrix

²Specifically, the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) renormalization group equations.

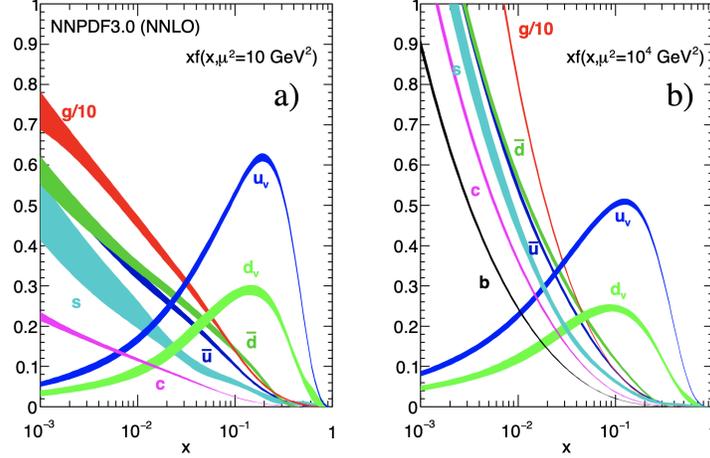


Figure 5.2: Parton distribution functions obtained by NNLO NNPDF3.0 global analysis illustrating the gluon and quark flavor contributions to the proton composition as a function of x at $Q = \mu_F = 10 \text{ GeV}$ (a) and $Q = \mu_F = 100 \text{ GeV}$ (b). Note that a factor of 0.1 is applied to the gluon PDF [32].

element \mathcal{M} and the phase space integral. The matrix element represents the probability amplitude for the transition from an initial state i to a final state f to occur. The phase space integral represents the kinematics available to the participating particles for the interaction to occur. In practice, the parton-level cross-section $\hat{\sigma}_{i \rightarrow f}$ is obtained by taking the absolute value squared of the matrix element, summing over all possible polarizations and color states, and integrating over the phase space. The cross section can be calculated at fixed order in perturbation theory, where one approximates the series up to a given order n in the strong coupling α_s^n , with the assumption that the contribution from the omitted higher orders should be small. Each power of α_s corresponds to the addition of new diagrams including an extra real or virtual emission, starting from the leading order (LO) diagram with no emissions. The cross section calculated at next-to-leading (NLO) accuracy contains the contribution from diagrams with one emission or one loop. These diagrams introduce different types of divergences, which have to be regulated to preserve unitarity. Diagrams with loop corrections introduce ultraviolet (UV) divergences. Because QCD is a renormalizable theory, these divergences can be treated by first regularizing and then renormalizing the theory, i.e. the divergences can be absorbed in the redefinition of the parameters.

Both virtual and real-emission diagrams exhibit infrared and collinear (IRC) divergences when the emitted gluon is soft or collinear. However, according to the Kinoshita-Lee-Nauenberg (KLN) theorem, order-by-order unitarity implies that the singularities coming from integration over unresolved real emissions must cancel, order by order, with the equal but opposite sign singularities generated by integrating over the virtual loop corrections. As

long as both contributions are included, the calculation of the n^{th} order is finite [128]. In principle, the energy and spatial resolution of the detector acts as a regularizer, by making these corners of phase space not detectable and therefore not contributing to the total observable cross section. However, it is desirable not to have theoretical calculations based on experiment-dependent parameters [127]. It is therefore preferable to study observables for which the KLN theorem holds, called *IRC-safe* observables.

5.1.3 Showering and hadronization

The scattering of any charged particle leads to the emission of radiation, called bremsstrahlung. This occurs both in QED and QCD, with photon and gluon emission respectively. Unlike photons, however, gluons carry themselves color charge and will give rise to further gluon radiation and parton multiplication [129]. A parton produced in the hard scatter will start at the scale of the hard process and move towards a lower scale, with predominantly soft and collinear emissions. This process is called *fragmentation* and continues until the partons are resolved at a scale of $Q_{had} \sim 1$ GeV. At this point, confinement requires these particles to undergo some transition from free colored partons to color singlet hadrons. This non-perturbative process is called *hadronization*.

Although the quark and gluon emissions occur in the perturbative regime, the high parton multiplicities would require matrix element calculations to very high orders, a task in most cases not solvable analytically. At the same time, the hadronization process is non-perturbative and not very well understood theoretically. Event generator packages use therefore an alternative approach, where the perturbative emissions are treated as a probabilistic process, referred to as *parton shower*. A shower of soft and collinear quarks and gluons is simulated to accompany the partons participating in the hard scatter, in practice providing approximations of the higher-order real-emission corrections [130].

Matrix element calculations provide an exact solution at fixed order for hard wide-angle emissions, but can only handle a few for the problem to be analytically solvable. On the other hand, parton showers describe well regions of the phase space dominated by soft and collinear gluon emissions, but fail to model hard wide-angle gluon emissions. In order to simulate pp collisions, both methods are typically combined, as discussed later.

5.1.4 Soft physics

The *Underlying Event* UE describes any process that accompanies a hard inelastic scatter. Due to the composite nature of protons, each pp collision can contain several few-GeV collisions between secondary partons from the same colliding protons, referred to as *multi-*

parton interactions (MPIs). In addition, each colliding proton may also leave behind a beam remnant, which does not take part in the initial state radiation or hard-scattering process, but still remains color connected to the rest of the event. The contribution of the UE to the final state is understood only phenomenologically from data due to its non-perturbative nature. In particular, the UE is related to the *pedestal effect*: kinematic distributions of hard jets display a constant E_T plateau that is significantly higher than what is observed for a minimum bias event³. The larger activity is explained by a trigger-induced bias. The trigger selection of a hard jet biases the event selection towards more central collisions, associated to a larger number of MPIs and increased event activity [129].

5.1.5 Monte Carlo event generators

Monte Carlo (MC) simulations are an essential part of the ATLAS physics program, as they allow to develop new analysis methods, isolate specific physics signatures with targeted phase space selections, perform calibrations between data and MC, and provide the distributions for the background-only hypothesis in any statistical fit. A MC sample is a set of events representing a given process. Each event represents the same hard interaction, but the kinematics of the final state objects varies event-by-event according to the true probability distributions. The MC generation path is generally composed of independent steps carried out by different MC simulation programs. The objects output at each of these steps are said to be at the *parton level*, *hadron-level*, or *reco level*. In the context of ATLAS performance studies, reconstruction algorithms can be fed input objects from any of these stages, according to the need.

Parton-level

Matrix element generators are used to simulate pp collision events at the parton-level. One of the most widely used is MADGRAPH. In the first step, MADGRAPH calculates the matrix element, which provides the mathematical description of the interaction and is a function of the momenta of the final state particles. This is usually performed to the highest possible order, although this often remains the LO. The result is then convoluted with the chosen PDF set describing the partonic structure evaluated at the LHC CoM energy. Short-lived particles produced in the hard-scatter are decayed. When referring to the *parton-level*, one refers to the particles output by the matrix element calculation. The phase space integral is

³*Minimum bias* refers to a data sample collected by an experiment using a “minimum bias trigger” (see Sec. 4.3.3). The resulting sample includes a mixture of soft and hard processes, with a prevalence of soft events. While the processes that make up the UE are similar to the soft interactions that dominate a Minimum Bias sample, the two are not the same, as the definition of the UE requires a hard scatter to have occurred, resulting in increased event activity.

then computed using numerical integration to obtain real predictions for the cross section. The result is a statistically representative sample of parton-level events for the given process. The parton level gives a good description of the momenta of the outgoing particles. However, fragmentation and hadronization have to be simulated in order to correctly reconstruct the interaction with the detector.

Hadron-level

Common event generators are PYTHIA, HERWIG, and SHERPA, which can be used to simulate the parton shower, the decay of unstable particles, the formation of hadrons, and multiple pp interactions. The packages differ in the type of algorithm used for showering and hadronization. For instance, PYTHIA showering algorithm is based on the *Lund string model*, where quarks are thought of as strings and quark confinement is represented as a string potential. As the quarks at the endpoints of the string move apart, the potential energy increases until enough energy is available for a new $q\bar{q}$ pair to be created, breaking the string into two separate color singlet pieces. At the end of this fragmentation process the color connected partons are combined to create hadrons.

The most common method to simulate pp collisions is to combine LO matrix element predictions with parton showers. Another possibility is to start from NLO (or higher) matrix element calculations before interfacing with a parton shower generator. Such approaches are used by MC@NLO and POWHEG. This is advantageous, as one can benefit from the higher accuracy and smaller normalization uncertainty of NLO predictions. However, when combining NLO matrix element calculations with parton showering, special care has to be taken to avoid double counting in overlapping regions of phase space, a process called *merging*.

The stable hadrons at this point are referred to as the *hadron-level* or *particle-level* of the MC simulation. A particle is considered stable if its lifetime is long enough for it to interact with the detector. Although the actual lifetime cutoff is somewhat arbitrary, the convention used by ATLAS in MC simulation is $\tau > 10$ ps.

Reco-level

The stable particles output by the event generators are passed through the detector simulation GEANT4 [131], which simulates their interaction with the different detector materials. Next, a digitization step reproduces the detector's response and readout. At this point, the simulated events are in the same format as any real data event recorded during operations. The only difference is that the MC simulation retains the *truth information* about the hard process, including particle types, four-momenta, and decay chains [126]. The same

object reconstruction algorithms are run on data and MC events. The objects output by the reconstruction step are said to be at the *reco-level* or *detector-level*.

5.1.6 Jets

A *jet* is a collimated spray of particles resulting from the showering and hadronization of high-energy quarks and gluons. As discussed in the previous section, the hadronic final state of a hard scatter can be described on three levels: the final state partons of the hard process (*parton-level*), the final stable hadrons before interaction with the detector systems (*hadron-level*), and the observable energy depositions in the detector (*reco-level*). A jet algorithm takes a list of input objects — at the level of particles, hadrons, or energy depositions — and returns a list of new objects called *jets*. The processes that relate these three levels are complex and result in jets whose composition — in terms of type, multiplicity, and momenta of the particles associated to each jet — varies between events. Nonetheless, the direction of the jet, built from the four-momentum sum of its constituents, is generally a good representation of the original direction of the parent parton [13]. In principle, therefore, there is a close correspondence between these three levels of description, as represented in Fig. 5.3. This makes jets important proxies to study the partonic dynamics of the collision and ubiquitous tools in collider physics. However, in practice soft non-perturbative physics, such as pileup and UE, as well as additional hard QCD emissions, can blur the picture, making the task of a jet algorithm more complicated. As an example, consider a simple di-jet event, where two quarks are produced in the hard scatter accompanied only by soft and collinear emissions. The event will have two cone-shaped energy depositions in the detector associated to the two quarks and can be reconstructed in a straightforward way. In contrast, consider an event where one quark emitted a hard wide-angle gluon. There is somewhat an ambiguity on whether this should be considered a single jet or two jets. The decision of when an emission is deemed hard enough for it to be considered a separate jet depends on what physics question one wishes to study and is made via the choice of a jet algorithm [127]. The presence of extra radiation in the final state, including pileup, can also affect jet physics, as it can modify jet properties. The subject of jet reconstruction and identification and of pileup suppression are relevant for this work and will be discussed in detail in the following sections.

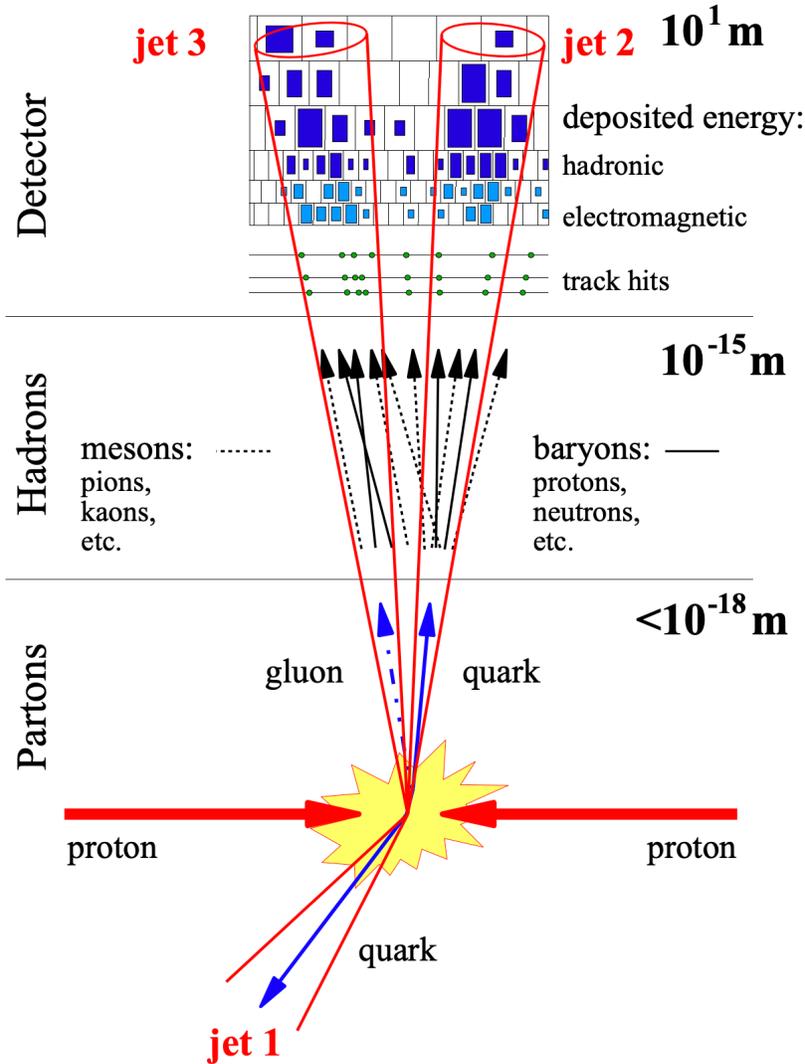


Figure 5.3: Illustration of the correspondence between a jet and the possible types of objects associated to it: the partons produced in the final state, the hadrons resulting from showering and hadronization, and the energy depositions in the calorimeter. *Reproduced with permission from Springer Nature from Ref. [130], Fig. 5.2.*

5.2 Jet reconstruction algorithms

The reconstruction of jets depends on the *jet definition* and the algorithm inputs. The input particles are described by their four-vectors and can be partons, hadrons, or energy deposits. The jet definition is determined by the jet algorithm, or the rules used to combine particles into groups of objects, and by the *recombination scheme*, the rules used to combine the momenta of the grouped objects into the momentum of the final jet. The standard recombination scheme is the $E - scheme$, where the four-vector of the jet is given by the

sum of the components of the four-vectors of its constituents. All jet algorithms can be classified according to two broad categories [127]:

- *Cone algorithms* rely on an event-level (top-down) approach where the jets are viewed as dominant directions of energy flow.
- *Sequential-recombination algorithms* have a bottom-up approach, where the closest particles – according to some predefined metric – are recombined iteratively, as if reproducing in reverse the fragmentation process.

The choice of which algorithm to use is based on physics and practical considerations, including the requirement of infrared-collinear (IRC) safety, the dependence of the boundary of the jets on soft emissions, and the computational time. Until the first years of LHC operation, cone algorithms were favored despite being IRC unsafe, because of the well-defined circular shape of the output jets, less sensitive to non-perturbative effects and easier to calibrate. This changed, however, with the development of the anti- k_T algorithm [106], as it provided both an IRC safe and soft-resilient shape option, making it the current standard of jet reconstruction.

5.2.1 Infrared-collinear safety

Ideally the set of hard jets reconstructed in an event should be insensitive to the random soft and collinear emissions characterizing the showering process. Experimentally, the detectors' resolution acts already as a regularizer, as below a certain scale one has no way of distinguishing a parton from a parton plus a collinear or soft emission. However, this is detector-dependent and can make it difficult to connect the experimental measurement to theoretical predictions. From the theoretical side, as discussed in the previous sections, fixed-order perturbative QCD calculations used to make these predictions remain finite thanks to the cancellation of divergent contributions from real and virtual emission diagrams. Observables where this cancellation is guaranteed are said to be *IRC-safe*. In general, an observable is IRC-safe if its value remains unchanged under any number of soft or collinear splittings. In other words, if \vec{p}_i is the momentum of any particle entering the definition of an observable, the observable must be invariant under the branching $\vec{p}_i \rightarrow \vec{p}_j + \vec{p}_k$, whenever $\vec{p}_j \parallel \vec{p}_k$ (collinear) or $\vec{p}_j \rightarrow 0$ (soft) [127].

IRC-safety for jet algorithms is necessary for any QCD precision studies. The preferred IRC safe algorithm for jet reconstructions in ATLAS is currently the anti- k_T algorithm, while for jet substructure the k_T or Cambridge/Aachen are generally used, as they are more sensitive to QCD branching. Nonetheless, non-IRC safe jet algorithms can still give good

and not necessarily worse predictions. Most cone jet algorithms used up until recently fall into this category.

5.2.2 Cone algorithms

Cone algorithms rely on the idea that soft and collinear emissions will not modify the main features of an event and define jets as angular cones around dominant directions of energy flow [127]. In order to reduce the computational time, cone algorithms are typically *seeded*. A proto-jet is built around the seed, whose constituents are selected by drawing a cone of radius R around it⁴. The four-momentum of the jet is calculated from the constituents according to the recombination scheme used. Iterative procedures are usually implemented to select stable cones: a cone is stable when the axis (usually given by the four-vector sum of its constituents) points in the same direction as its seed.

Typical issues with these algorithms are the problem of overlapping cones and IRC unsafety. The overlap of two cones is an issue for the reason that energy is being double-counted. Cone algorithms can be subdivided into two classes according to how they deal with this. Some algorithms, including the old ATLAS iterative cone with split-merge (IC-SM) algorithm, implement a split-merge approach: if two overlapping jets share more than a fraction f of their energy, the jets are merged, otherwise the constituents are split among the two jets. Other algorithms build the cones starting from the hardest seed, and once the stable cone is found, its constituents are removed from the event before moving on to the next seed. This results in hard jets always being perfectly circular. An example of this type is the old CMS iterative cone with progressive removal (IC-PR) algorithm.

The problem of IRC-unsafety typically arises from the seeding procedure: the selection of seeds according to their hardness is problematic, as particles p_T 's are not collinear safe quantities. If a hard particle, which under the no-emission scenario would result in a hard seed, undergoes a resolvable collinear splitting, the result will be two lower energy seeds. This can result in different seed choices and hence in different jets. An attempt at avoiding selecting seeds according to their hardness was made by building all the possible stable cones to then select the hardest ones. However, this was shown to be unsafe under soft emissions. Consider two hard particles at a distance $R < \Delta R < 2R$, where the cones built on them do not overlap. If a soft emission occurs at a distance R between the two, it will produce a cone including both jets that could be harder than the two jets alone. In 2007 an IRC-safe cone

⁴As defined in Sec. 4.2.1, the angular distance between two objects i and j in the detector is given by $\Delta R_{i,j} = \sqrt{\Delta(y_i - y_j)^2 + \Delta(\phi_i - \phi_j)^2}$. Drawing a cone of radius R around the seed means selecting all the objects with $\Delta R_{\text{seed,object}} < R$.

algorithm was developed called *SISCone* (Seedless Infrared Safe Cone) [132]. However, this option loses one of the advantages of cone algorithms, as it produces irregular jet boundaries due to soft radiation. For a comprehensive list of cone algorithms see Ref. [133].

5.2.3 Sequential-recombination algorithms

The most widely used sequential-recombination algorithms today belong to the family of the k_t algorithms. These algorithms introduce a new distance metric between particles and iteratively combine the closest pair of particles until not particles are left. The inter-particle distance is given by,

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = k_{ti}^{2p}, \quad (5.4)$$

where k_{ti} is the transverse momentum of particle i , ΔR_{ij} is the distance in the rapidity and azimuth plane between particle i and j , R is the radius parameter of the algorithm, and p is an input parameter. The algorithms differ in the value of the parameter p , which determines the momentum weighting: $p = 2$ for the k_t algorithm [134, 135], which combines soft and collinear particles first; $p = 0$ for the Cambridge/Aachen algorithm [136, 137], which clusters particles together only based on angular proximity; and $p = -2$ for the anti- k_t algorithm [106], which preferentially combines hard particles.

The recombination is an iterative procedure:

1. Start with a list of input objects
2. For each particle i , calculate the distances d_{ij} from every other particle and the distance d_{iB} of the particle from the beam.
3. Find the minimum distance d_{min} in the set of $\{d_{ij}\} \cup \{d_{iB}\}$. If $d_{min} \in \{d_{ij}\}$, combine particles i and j into a new particle, remove them from the list of input objects, and add the new particle to the list. If $d_{min} \in \{d_{iB}\}$, call particle i a jet and remove it from the list of inputs.
4. Repeat from step 2.

Originally these algorithms were considered very slow, as naively the algorithmic complexity scales like N^3 : one has to calculate N^2 distances and repeat for N iterations. However, it was later shown that the speed can be greatly improved with geometrical arguments [138]. First, one can prove that the d_{ij} distance in step 2 does not need to be computed for every pair of particles i and j , but only for particle i and its nearest neighbor, so the total complexity

is reduced to $O(N^2)$. This can be further improved by making the finding of the nearest neighbor more efficient. Using the Voronoi diagram technique from computational geometry one can reach an algorithm complexity of $O(N \ln N)$.

This class of algorithms is usually trivially made IRC safe. Consider the case of the anti- k_t algorithm and assume a new soft particle i is produced in the event. If d_{iB} is the minimum distance, this will produce a new jet with $p_T \rightarrow 0$. If instead the particle is closest to another particle j , the d_{ij} distance will be dominated by the $1/k_{ti}^2$ term, so that $d_{ij} \rightarrow \infty$ and the soft particle will be clustered last. Similarly, a particle originating from a collinear emission will have $\Delta R_{ij} \rightarrow 0$, so it will cluster first to the hard jet and not change its coordinates. Either way, the addition of a soft or collinear particle has no effect on the hard jets found in the event.

These algorithms implicitly produce a clustering sequence for the event. In the case of the k_t algorithm, this is closely related to the probabilistic emissions in the parton shower: the pair that recombines first is the one with the highest probability of having been produced by the same splitting. For this reason, the k_t algorithm is often used for substructure studies of hadronic decays of boosted massive particles, such as top-quark and Higgs, W , and Z bosons. A draw back of the k_t algorithm is that the shape of the resulting jets is sensitive to soft radiation, resulting in irregular boundaries. This is caused by the fact that soft particles are clustered together first, so the presence of a soft jet around the boundary can affect whether close-by particles get assigned to the jet or not. Similar conclusions hold for the Cambridge/Aachen algorithm.

Conversely, the anti- k_t algorithm clusters first hard objects that are close together, which ensures that the jet grows around a hard core, but does not bring information about the substructure. As new particles are added to the proto-jet, the jet axis can move slightly but, in the absence of other nearby hard particles, the final shape will be a perfect cone of radius R . The result is that of an ideal stable cone algorithm, making it the most accurate algorithm to resolve jets. Anti- k_t also automatically takes care of the potential issue of two hard particles at a distance ΔR that is $R < \Delta R < 2R$, a situation that would produce overlapping cone jets. Considering the two extreme situations:

- If $k_{t1} \gg k_{t2}$, the jet around particle 1 will be conical, while the second jet will lose some of its constituents.
- If $k_{t1} = k_{t2}$, the boundary will be a straight line equidistant between the two jets.

5.3 Jets in ATLAS

Being a pp collider, jets are ubiquitous in LHC physics and are essential components to many SM measurements and searches for new phenomena. On average, two-thirds of the visible jet energy is contributed to by charged particles, predominantly by charged pions, a quarter is composed of photons from neutral hadron decays, and the remainder consists in neutral hadrons [139]. Jets interact therefore with the inner detector (ID), before being stopped in the calorimeter, where they leave a cone-like energy deposition.

The standard for jet reconstruction in ATLAS is the anti- k_t algorithm with radius parameter $R = 0.4$ for small-R jets, and $R = 1.0$ for large-R jets, the latter used in the reconstruction of boosted hadronic decays of massive particles. The inputs to the jet algorithm consist in a list of four-vectors, which can describe charged particle tracks from the ID or energy deposits in the calorimeter, or a combination of the two. Stable particles from MC generators at the parton- or hadron-level can also be used for MC studies. Jets produced with different inputs are referred to as *jet collections*. Jets built from detector inputs have to be calibrated to compensate for several factors, including detectors inefficiencies (particularly the non-compensating nature of the hadronic calorimeter) and electronic and pileup noise. The calibration chain is different for small-R [140] and large-R [141] jets, and is performed independently for any given jet collection. As large-R jets are used to reconstruct decays of boosted massive particles, where the jet mass is well-defined, their calibration includes both energy and mass corrections. Different techniques for pileup mitigation can also be used, both directly on the set of objects input to the jet algorithm, and on the reconstructed jets. Lastly, an essential step in most physics analyses is the identification of the true particle from which a given jet originated, a procedure called *jet tagging*. Several algorithms have been developed in ATLAS in the context of heavy flavor (see Sec. 4.4.6) and boosted large-R jet identification. In the following, the steps of the ATLAS jet reconstruction process most relevant for this thesis are discussed, including jet collections and methods for boosted jet tagging. The topic of pileup suppression is discussed in the next section.

5.3.1 Jet algorithm

In ATLAS, jets are reconstructed using the anti- k_t jet algorithm, as implemented in the FASTJET package [142]. The standard radius parameter for jet reconstruction is $R = 0.4$. These are referred to as *small-R jets* and are used to reconstruct jets originating from individual partons, such as a hard quark produced via the strong interaction or the two b -quarks from a resolvable Higgs boson decay. The average transverse distance between two

particles coming from the decay of a particle of mass m and transverse momentum p_T is approximately [126],

$$\Delta R \approx \frac{2m}{p_T}. \tag{5.5}$$

For example, a Higgs boson with $p_T = 250(500)$ GeV produces a jet contained on average within a cone of radius 1.0 (0.5). In other words, the larger the transverse momentum of a particle, the more its decay products are collimated. For large boosts, the decay products can be sufficiently collimated that they are not resolvable as separated jets anymore. *Large- R jets* with $R = 1.0$ were introduced to recover efficiency in the reconstruction of boosted decays of massive particles such as top-quark and W , Z , and Higgs bosons.

5.3.2 Jet inputs and jet collections

EMTopo and LCTopo

During Run 2, the standard inputs for jet reconstruction were topological clusters (see Sec. 4.4.4). According to whether the topoclusters are input at the EM scale or are LCW calibrated, the corresponding jet collection is referred to as *EMTopo* or *LCTopo*. This was possible thanks to the excellent ATLAS calorimetry, which provides clusters with high energy resolution. However, in the increasingly dense environments of the LHC, several improvements can be gained with a *particle flow* approach that makes use of both calorimeter and track information. Two such reconstruction strategies were developed at the end of Run 2.

Particle Flow

The Particle Flow (PFlow) algorithm [139] relies on tracking information to improve the performance of the reconstruction of low p_T charged particles. Tracks provide a superior momentum resolution for low p_T particles, and better angular resolution that allows to recover low p_T charged particles swept outside the jet cone by the magnetic field before reaching the calorimeter. They also allow to reject charged pileup particles not originating from the primary vertex. The algorithm associates individual well-reconstructed tracks to single topoclusters in the calorimeter and then finds the best position and energy measurement for each track-cluster system according to which detector has a better resolution in the given energy regime. Different processing steps account for the possibility of overlapping showers or of a track contributing to more than one cluster. The final input objects (PFOs) consist of tracks, the remaining modified clusters, and the clusters not matched to any track that are considered originating from neutral particles. The resulting *PFlow jets* show a superior performance at low p_T . Originally, the tracking resolution deteriorated at higher p_T ,

but more recent developments obtained a resolution compatible to *EMTopo* jets [140].

Track-CaloClusters

The *Track-CaloClusters* (TCCs) [143] algorithm focuses on combining the spatial information of the tracker and the energy measurement of the calorimeter at high p_T . This method also improves the identification of substructure in large-R jets, as it can resolve distinct particles associated to a single topocluster. However, TCC jets suffer from pileup instabilities and their performance is typically worse than the standard jets at low p_T .

Unified Flow Objects

In 2021 a new input definition called *Unified Flow Objects* (UFOs) was developed [144], which combines the desirable aspects of PFlow and TCC reconstruction for an optimal overall performance across the full p_T regime. The resulting jet collection is referred to as *UFO jets*, and has a superior performance to TCC jets at high p_T , while retaining a similar performance to PFlow jets at low p_T .

Track jets

Track jets are built from ID tracks. These jets are primarily used in the context of b -tagging of subjects contained in large-R jets. At the beginning of Run 2, the standard radius parameter for track jets was $R = 0.2$. This was later changed to a variable radius parameter $R(p_T) = \frac{30\text{GeV}}{p_T}$, inversely proportional to the p_T of the jet, which better describes the p_T -dependence of the angular spread of a jet according to Eq. (5.5). The algorithm [145] has two additional parameters, R_{\min} and R_{\max} , to set the lower and upper limits on the jet size. The resulting jets are referred to as *variable-radius (VR) track jets*.

Truth jets

Truth jets take as input hadron level stable particles (see Sec. 5.1.5). Truth jets can only be reconstructed in simulation, but are essential for performance studies, such as algorithm development or calibration. In this context, it is often important to know the true generator-level parton from which a reconstructed jet originated. This is typically found via *truth matching*: the reconstructed jet is matched to the closest truth jet, and the truth jet is matched to the closest stable particle, where matching generally consists in a minimum ΔR requirement. The label of the reconstructed jet will be the type of the matched particle.

5.4 Boosted jet tagging

A tagger attempts to identify the true particle from which a jet originated. Several BSM models predict new heavy resonances with masses around 1 TeV and with significant decay branching ratios into highly Lorentz boosted SM bosons (see Chap. 3). Since in more than 60% of the cases W , Z , and Higgs bosons decay hadronically into a pair of quarks, boosted jet tagging plays an essential role in these searches, including the one discussed in this thesis (see Chap. 7).

Most forms of jet tagging are a form of supervised learning, so that a method needs to be established to provide true labels for the jets. For boosted heavy particles, such as top-quark and W , Z , and Higgs (H) bosons, the radiation pattern is generally isolated from the rest of the event, although some ambiguity remains related to whether the full radiation originating from the particle is contained in the jet. This is more complicated for jets originating from colored particles, for which a formal separation of the decay from the rest of the event is not possible [146]. The labeling of the training samples typically involves truth-matching together with some containment criteria based on the truth information from the parton-level of the MC simulation. Defining the true particle type as signal, and the rest as background, the performance of a tagging algorithm is quantified in terms of the *signal efficiency* ϵ_s — the probability of correctly tagging a signal jet — and of the *background efficiency* or *mis-tag rate* ϵ_b — the probability of incorrectly identifying a background jet as signal. One often quotes also the background *rejection factor*, defined as $1/\epsilon_b$ ⁵.

The reconstructed mass of a jet is one of the most important discriminants between jets of different origin. For a jet originating from a heavy particle, the jet mass has a scale associated to the mass of the particle, while for a q/g -induced jet, the mass scales as the product of the jet p_T and radius. Important information about a large-R jet is contained also in its internal structure, as different particle origins will determine different multiplicities and kinematic distributions of the jet constituents. Jet substructure [146, 147] is a field that aims at exploiting the radiation pattern inside jets as a tool for boosted jet-tagging, as well as to perform precision tests of QCD. An important feature that is usually exploited by jet substructure techniques is the number of prongs in the jet. For instance, $H/W/Z$ boson hadronic decays typically display a two-prong structure, with two subjets evenly sharing the momentum of the mother particle. Similarly, a large-R jet fully containing a hadronic top-quark decay will have a three-prong structure. On the other hand, a q/g -initiated jet is generally one-pronged and, in the case of a real emission, the second prong is usually significantly softer. In fact, several substructure observables rely on the identification of

⁵In other fields background rejection is more commonly defined as $1 - \epsilon_b$.

dominant directions of energy flow inside the jet. Some techniques look explicitly for hard subjets contained in the jet. These include N -subjettiness observables, which rely on the identification of explicit axes associated to N -prong decays, and declustering techniques, such as k_t splitting scales, which identify the subjets by walking in reverse the jet clustering history. Other jet-shapes methods, such as energy correlation functions and Fox-Wolfram moments [148], quantify the energy dispersion of the jet constituents in an axis-independent way. In the following, a subset of these substructure observables and boosted jet tagging methods relevant for this thesis are discussed.

k_t splitting scales

The k_t splitting scales [149] are obtained by reclustering the jet constituents using the k_t algorithm, which clusters harder constituents last, and then look at the k_t distance at a given step of the clustering history. The *splitting scale* variable d_{ij} is defined as,

$$\sqrt{d_{ij}} = \min(p_{T,i}, p_{T,j} \times \Delta R_{ij}). \quad (5.6)$$

In particular, the $\sqrt{d_{12}}$ variable refers to the splitting scale at the last clustering step for the two hardest subjets. Similarly, $\sqrt{d_{23}}$ is given by the second-to-last clustering step for the second and third hardest subjets. The variables $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ are helpful in identifying the two- and three-prong decays of heavy particles, which show a more symmetric energy sharing between the subjets than the splittings in q/g -jets.

N -subjettiness

The N -subjettiness [150] observables τ_N are also obtained by reclustering the jet constituents using the k_t algorithm to identify the N hardest subjets. The variable τ_N is calculated as,

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}\}, \quad (5.7)$$

where d_0 is the normalization factor $d_0 = \sum_k p_{T,k} R_0$ and the sum runs over the k jet constituents. The result can be interpreted as a metric of how good is the hypothesis that the jet has N hard subjets. For a jet with N or fewer true subjets, $\tau_N \approx 0$, as all the jet constituents are aligned with one of the N directions. On the other hand, jets with more than N true subjets, will have $\tau_N \gg 0$, as a larger number of constituents will be at a larger distance from the identified axes. For instance, a jet originating from a W decay, with two subjets, will have $\tau_1 \gg 0$ and $\tau_2 \approx 0$. A QCD jet containing two hard quarks can in principle have $\tau_2 \approx 0$. However, such a QCD jet is accompanied by significantly

more wide-angle radiation, determining a correlation between τ_2 and τ_1 . For this reason, the N -subjettiness ratio $\tau_{21} = \tau_2/\tau_1$ has a greater discrimination power to identify two-hard-prongs decays from a q/g -initiated jet. Similarly, the ratio $\tau_{32} = \tau_3/\tau_2$ is used to identify three-pronged top-decays.

Energy correlation functions

Energy correlation functions (ECFs) are used to identify N -prong substructure in a similar manner to N -subjettiness ratios, with the main difference being that ECFs do not require finding subjets. For a hadron collider, the N -point ECF is defined as

$$ECF(N, \beta) = \sum_{i_1 < i_2 < \dots < i_N \in J} \left(\prod_{a=1}^N p_{Ti_a} \right) \left(\prod_{b=1}^{N-1} \prod_{c=b+1}^N \Delta R_{i_b i_c} \right)^\beta, \quad (5.8)$$

with the corresponding one-, two-, and three-point ECFs given by,

$$ECF(1, \beta) = \sum_{i \in J} p_{Ti}, \quad (5.9)$$

$$ECF(2, \beta) = \sum_{i < j \in J} p_{Ti} p_{Tj} (\Delta R_{ij})^\beta, \quad (5.10)$$

$$ECF(3, \beta) = \sum_{i < j < k \in J} p_{Ti} p_{Tj} p_{Tk} (\Delta R_{ij} \Delta R_{ik} \Delta R_{jk})^\beta, \quad (5.11)$$

where the sums run over the jet constituents and β is a parameter to be optimized. The ECF is IRC safe for any value of $\beta > 0$. Different ECFs are useful according to the application. In practice, if a jet J has N subjets, then $ECF(N+1) \ll ECF(N)$. It follows that the ratio $r_N = ECF(N+1)/ECF(N)$ behaves very similarly to the N -subjettiness observable τ_N , while the energy correlation double ratios $\frac{r_N}{r_{N-1}}$ behave like N -subjettiness ratios. Two important dimensionless ratios of ECFs (proposed in Refs. [151] and [152], respectively) are,

$$C_2^\beta = \frac{ECF(3, \beta) ECF(1, \beta)}{ECF(2, \beta)^2}, \quad (5.12)$$

$$D_2^\beta = \frac{ECF(3, \beta) ECF(1, \beta)^3}{ECF(2, \beta)^3}. \quad (5.13)$$

These are useful for the identification of 2-pronged substructure and are used for boosted $W/Z/H$ vs. q/g jet discrimination, with the former signatures having predominantly lower values of C_2 and D_2 .

Grooming algorithms

Grooming techniques aim at cleaning the jet of soft and wide-angle radiation in order to enhance the hard radiation pattern inside the jet, with the overall effect of also reducing the sensitivity to radiation that does not originate from the final state, such as pileup and the UE. In the context of jet tagging, grooming has proven a useful tool in identifying jet substructure, and it is often used in combination, or as input, to jet tagging algorithms. The main difference between a groomer and a tagger is that a tagger provides a classification of the true jet origin and is optimized to increase the signal-to-background ratio, while a groomer returns the cleaned (groomed) jet and is optimized in order to improve the resolution of the jet kinematics and properties [74]. Several grooming algorithms are used in ATLAS. All share the common idea of reclustering the jet constituents using the k_t or C/A algorithms and then use the output clustering history to remove soft components. Here, the soft drop algorithm [153] is described, as it is the one relevant for this work.

In the first step, the anti- k_t $R = 1.0$ jets are reclustered with the C/A algorithm. The angular-ordered clustering sequence history is then reversed. The last stage of C/A clustering $p_{i+j} \rightarrow p_i + p_j$ is undone by breaking the jet p_{i+j} into two subjets p_i and p_j . The soft drop condition is then evaluated:

$$\frac{\min(p_{T,i}, p_{T,j})}{p_{T,i} + p_{T,j}} > z_{\text{cut}} \left(\frac{\Delta R_{ij}}{R_0} \right)^\beta, \quad (5.14)$$

where R_0 is the jet radius and β and z_{cut} are parameters to be optimized for the algorithm. If the condition is satisfied, the declustering is stopped and the jet $i + j$ is taken as the final jet, otherwise only the subjet with larger p_T is kept and the procedure is repeated. The parameter z_{cut} determines the choice of what emissions should be deemed soft and excluded, while the angular exponent β weights the soft threshold according to the angular separation between the two subjets. In ATLAS the current recommendation is $\beta = 1.0$ and $z_{\text{cut}} = 0.1$.

5.5 Pileup suppression

*Pileup*⁶ refers to the simultaneous pp collisions that occur per bunch-crossing (see Sec. 4.1.3). The average pileup multiplicity was already $\langle N \rangle = 20$ in Run 1, $\langle N \rangle = 50$ at the end of Run 2, and is expected to go up to $\langle N \rangle = 200$ at the HL-LHC. As every pileup collision adds tens of soft hadrons to the final state, the net effect is that of adding hundreds to thousands of soft particles on top of the decay products of a hard collision of interest [74].

⁶The material in this section is based primarily on Ref. [74]

Mitigating the effect of this extra radiation is one of the main challenges for trigger and data analysis at the LHC.

In the context of jet reconstruction, pileup contamination has two main consequences: a bias and a smearing of measured kinematic quantities. Consider the pileup contribution at any given (η, ϕ) location as sampled from a Gaussian of mean ρ and standard deviation σ . The mean represents the average positive bias induced by the increased hadronic activity on the quantities measured. For example, the transverse momentum of a jet increases with increasing pileup proportionally to the jet area. The variation σ parametrizes the fluctuations in the pileup-induced bias ρ per event and across the detector volume. The fluctuations are a form of noise that blurs the reconstructed quantities reducing their resolution. One third effect is the impact of the particles originating from pileup interactions on the jet clustering procedure itself, as jets built with and without pileup will look slightly different due to different clustering histories. However, this effect is generally negligible.

All pileup mitigation techniques aim at reducing these effects, but the approaches differ according to which object is “corrected”. Historically, the standard methods included event-by-event and jet-by-jet algorithms. However, to address the new challenges posed by the increasing levels of pileup, new approaches have been developed based on the correction of the jet algorithm inputs. In this section, some of the most common techniques of pileup suppression are discussed, with a focus on those that are most relevant for this thesis.

5.5.1 Area-median subtraction

The most widely used event-level scheme is the area-median subtraction approach. This was the standard in ATLAS during Run 1, and it was still extensively used during Run 2. The algorithm is based on the fact that, if one draws a grid on the $y - \phi$ plane for a given event, most patches will not contain any particle from the hard scatter, so that their momentum flow p_T/A_{patch} is a good estimate of the pileup transverse momentum density in the event ρ . The algorithm is therefore split in two steps. In the first step, one finds an estimate of ρ by breaking the event into patches of similar areas and taking the median $p_{T,i}/A_i$ of all patches. The use of the median instead of the average makes ρ less sensitive to outliers, such as very hard jets. The second step is to subtract from the kinematic distribution of each jet the correction ρA_{jet} , where A_{jet} is the catchment area of the jet⁷. Computing ρ event-

⁷The *catchment area* of a jet is defined as the area in $y - \phi$ space where the jet would contain infinitely soft particles. For pileup subtraction, the *active area* is generally used, obtained by running the jet clustering algorithm on all the particles in the event plus a dense coverage of ghost particles (particles with infinitely small momentum) distributed evenly in $y - \phi$ space. Assuming an IRC-safe algorithm, the addition of these soft particles does not affect the momentum of the output jets. Moreover, for the anti- k_t algorithm, the boundary of the jets will always be circular and approximately independent of the initial set of ghosts. If

by-event and A jet-by-jet results in jets with better resolution compared to other methods that subtract the average pileup contamination per-vertex or per-event, as averaging usually introduces extra resolution degradation. The area-median subtraction approach has proven to be a robust method, that leaves on average an unbiased transverse momentum of the jet. However, it also leaves a residual p_T resolution degradation from pileup fluctuations (σ) across the detector for a given event. The smearing was sufficiently small during Run 1 and 2 not to be an issue. However, as the levels of pileup keep increasing, these effects will become non-negligible, particularly on low p_T jets essential for certain measurements, such as di-Higgs production.

5.5.2 Grooming

Pileup mitigation techniques at the jet level usually focus on large-R jets, as the larger area makes them more sensitive to pileup or the UE. In this context, grooming techniques (discussed in the previous section) can be a useful tool. A fundamental difference exists between groomers and an approach like the area-median subtraction method. The latter aims at reducing the positive bias due to the pileup contribution independently of the hard process. In the case of the measurement of the mass of a boosted top large-R jet, this approach will apply the same correction to a top jet and a QCD jet, reproducing on average the correct top mass, but including a smearing effect coming from pileup fluctuations. A groomer, on the other hand, aims at reducing the smearing as much as possible to improve the jet kinematics resolution. This is at the expense of always introducing a negative bias, as the output jet is always pruned of some of its constituents even in the absence of pileup. In the case of a top-quark decay, a groomer would therefore retain the three hard prongs of the decay, while cleaning the jet of the extra radiation, resulting in a sharply peaked mass distribution, with little bias from pileup. For a QCD jet, on the other hand, a groomer would remove a significant portion of the soft radiation in the jet, hence strongly reducing the jet mass. While this represents a large negative bias, it is desirable in this case, as it allows to identify the QCD jet as background. The study of the interplay between grooming, pileup removal, and jet tagging algorithms is an active area of study.

5.5.3 Constituent-level

More recent approaches attempt to explicitly remove pileup contributions in a noise suppression fashion: the inputs to the jet algorithm themselves are pileup suppressed, by being re-

the number of ghosts per unit area is ν_g and $\mathcal{N}_g(J)$ is the number of ghosts contained in jet, then the scalar active area of the jet is given by $A(J) = \frac{\mathcal{N}_g(J)}{\nu_g}$ [74].

moved or by having their energy adjusted, thereby automatically improving jet observables, independently of the jet clustering algorithm. Different algorithms have been developed, including Soft-Killer (SK) [154] used by ATLAS and PUPPI [155] used by CMS. Other algorithms, such as Voronoi Subtraction [156] and Constituent Subtraction (CS) [157], extend the area-based subtraction method to particle-level pileup mitigation and are often used as a pre-processing step to adjust the constituents four-vectors before these are input to the former algorithms. In the following, the algorithms relevant for this work are discussed.

Soft-Killer

The SK algorithm relies on the idea that the most important discriminant between a particle originating from a pileup interaction and a particle coming from the hard scatter is its transverse momentum. The algorithm consists in calculating an event-dependent p_T^{cut} threshold quantifying the hadronic activity and removing particles that have a transverse momentum below this cutoff. This is similar to the pileup suppression strategy implicit in the ATLAS topoclustering algorithm 4.4.4, where the energy cut-off on the input cells is determined by the event-dependent pileup noise, so that as pileup increases, the noise threshold increases as well.

In practice, the value p_T^{cut} is found as the p_T threshold that gives $\rho = 0$, where ρ is the transverse-momentum-flow density used in the area-median approach. In practice, the event is divided into patches of area A_i , and ρ is set to the median transverse-momentum-flow density $\frac{p_{T,i}}{A_i}$ of all the patches:

$$\rho = \text{median}_{i \in \text{patches}} \left\{ \frac{p_{T,i}}{A_i} \right\}. \quad (5.15)$$

The value p_T^{cut} is found by increasing the p_T threshold until exactly half of the patches contain no particles. In practice, this is fast to compute, as it is equivalent to taking the median of the p_T 's of the leading particles in each patch:

$$p_T^{\text{cut}} = \text{median}_{i \in \text{patches}} \left\{ p_{T,i}^{\text{max}} \right\}. \quad (5.16)$$

The value of p_T^{cut} computed as a function of pileup vertices was shown to be slightly above 2 GeV at the HL-LHC $\langle \mu \rangle = 200$ conditions [154].

Two types of biases can arise with this method: a positive bias caused by energetic pileup particles that are above threshold and do not get removed, and a negative bias from soft true signal particles that get suppressed. The jet energy scale will not be affected only if the two biases cancel each other out. Similarly, the energy resolution will not suffer only if the

fluctuations in the biases are not large.

When the particles in question are at the detector-level (towers or topoclusters), the issue arises that a single particle may contribute in energy to different signals, or a given signal may receive contributions from different particles. In particular, in the case of a hard particle sharing a topocluster/tower with pileup particles, the pileup contribution will never get removed. In this case, the SK algorithm can be adapted: for a tower of area A , the p_T is adjusted as $p_T^{sub} = \max(0, p_T^{\text{tower}} - \rho A)$. The standard SK algorithm in Eq. (5.15) is then applied on the subtracted towers.

Voronoi and Constituent Subtraction

Voronoi (Vor) [156] and Constituent Subtraction (CS) [157] are both extensions of the area-based subtraction method to particle-level pileup mitigation. Consider the case of the particle type being topoclusters.

In Voronoi subtraction each individual topocluster is assigned a Voronoi cell, defined as all the points in space that are closer to the topocluster than to any other particle. The area of each cell is called the *Voronoi area* A_{Vor} . Each topocluster receives a correction to its transverse momentum as $p_T^{\text{corr}} = p_T - \rho \cdot A_{\text{Vor}}$, where ρ is the per-event pileup density as in the jet-area correction method.

The CS subtraction method uses ghost particles with $p_T^g = A_g \times \rho$ uniformly covering the (η, ϕ) plane. Every particle-ghost pair $i - k$ is then considered in ascending order of $\Delta R_{i,k}$, and the following correction to each particle and ghost p_T is applied:

$$\begin{aligned}
 \text{If } p_{T,i} \geq p_{T,k}^g: \quad & p_{T,i} \rightarrow p_{T,i} - p_{T,k}^g, \\
 & p_{T,k}^g \rightarrow 0 \text{ GeV} \\
 \text{otherwise:} \quad & p_{T,k}^g \rightarrow p_{T,k}^g, -p_{T,i}, \\
 & p_{T,i} \rightarrow 0 \text{ GeV}.
 \end{aligned} \tag{5.17}$$

Fixing the maximum ΔR values to be considered allows to tune the maximum jet area. The original method similarly provides also a mass correction, but this is not applicable when using topoclusters, as they are massless.

After the energy has been corrected, different techniques can be used to remove pileup-like topoclusters, going from simply removing any topocluster with $p_T^{\text{corr}} < 0$ GeV to feed the correct topoclusters to a more sophisticated algorithm, such as SK.

Chapter 6

Concepts of statistics and machine learning

6.1 Statistical inference

Consider¹ a set of independent and identically distributed (i.i.d.) data $X = (x_1, \dots, x_n)$ that is assumed to be sampled from a probability density function (PDF) $p(x|\theta)$, dependent on some parameters θ . In a standard inference problem, one wants to find the estimate of θ .

Frequentist vs. Bayesian

Statisticians use probability to quantify uncertainties, but they do not all agree on the interpretation. Let \mathcal{A} be an element of the sample space. The *frequentist* interpretation views probability as a limiting relative frequency. From a frequentist point of view, \mathcal{A} represents a possible outcome of a measurement assumed to be repeatable and, given N measurements, the probability of \mathcal{A} is given by

$$P(\mathcal{A}) = \lim_{n \rightarrow \infty} \frac{\text{number of occurrences of outcome } \mathcal{A}}{N}.$$

From a Bayesian perspective, \mathcal{A} represents a statement that can be either true or false, and $p(\mathcal{A})$ represents the degree of belief that hypothesis \mathcal{A} is true. Note that since the statement that “an experiment yields a given outcome a certain fraction of the time” can be regarded as a hypothesis, the framework of Bayesian probability includes the frequentist interpretation.

Frequentist inference

The most common method of frequentist inference is based on Maximum Likelihood Esti-

¹This section draws on notes taken throughout the years, particularly from Refs. [158–160].

mation (MLE). Given the observed data X , one first builds the likelihood function,

$$L(\theta) = p(X|\theta) = \prod_{i=1}^N p(x_i; \theta), \quad (6.1)$$

where the second equality holds because the measurements are assumed to be independent. The likelihood attains higher values for the choices of θ that are closer to the true distribution $p(x_i; \theta^{\text{true}})$. The best estimate of θ is then found by maximizing the likelihood or, equivalently, by minimizing the negative log-likelihood (NLL):

$$\hat{\theta}_{MLE} = \arg \max_{\theta} \prod_{i=1}^n p(x_i|\theta) = \arg \max_{\theta} \sum_{i=1}^n \log p(x_i|\theta) = \arg \min_{\theta} - \sum_{i=1}^n \log p(x_i|\theta). \quad (6.2)$$

In the frequentist framework, the true parameter θ is assumed to be fixed, but unknown. The parameter estimate $\hat{\theta}_{MLE}$, on the other hand, is a function of the data and therefore is a random variable. In other words, one can define a bias and a variance for the estimator, which describe how the estimator is distributed if one repeats the experiment several times. The variance of the estimator quantifies the irreducible aleatoric uncertainty due to the inherent variability of a random variable. Note that, contrary to Bayesian inference, MLE does not offer a way to quantify possible sources of epistemic uncertainty. This has the consequence of making the MLE prone to overfitting and to be overconfident in its predictions when limited data is available.

Bayesian inference

When performing Bayesian inference one encodes previous knowledge (or guess) about θ in a prior distribution $p(\theta)$ and applies Bayes theorem to obtain the posterior distribution over θ ,

$$p(\theta|X) = \frac{\prod_{i=1}^n p(x_i|\theta)p(\theta)}{\int \prod_{i=1}^n p(x_i|\theta)p(\theta)d\theta}, \quad (6.3)$$

where the numerator is the product of the likelihood and the prior. The denominator can be regarded, in most cases, as a normalization factor. From a Bayesian perspective, θ is not simply unknown, but it is itself a random variable, and the posterior distribution $p(\theta|X)$ naturally expresses both the aleatoric and epistemic uncertainty over θ . One can still obtain a point estimate by taking the maximum of the posterior (MAP),

$$\hat{\theta}_{MAP} = \arg \max_{\theta} \log p(\theta|X) = \arg \max_{\theta} \log p(X|\theta) + \log p(\theta). \quad (6.4)$$

The objective function in Eq. 6.4 retains the log-likelihood term from MLE as in Eq. 6.2, plus what looks like a regularization term coming from the prior. As shown later, this is often related to the choice of loss function and regularization method when training a neural network. Note that in the case of unlimited training data X , the second term becomes negligible and $\hat{\theta}_{MAP} \rightarrow \hat{\theta}_{MLE}$.

6.2 Neural networks

In a supervised learning problem one has a training dataset D composed of input features X and input targets Y . Given $\mathbf{x} \in X$ and $y \in Y$, the goal is to find a mapping f such that $y = f(\mathbf{x})$. A machine learning algorithm \mathcal{A} , which is a function of some parameters, provides an approximation $\hat{f} = \mathcal{A}(D)$ of this mapping. Given a data point \mathbf{x} , the quality of the prediction $\hat{y} = \hat{f}(\mathbf{x})$ is measured by a loss function $J = J(y, \hat{y}(x))$, a metric of how close the predicted and expected target values are. The learning – or training – is then formulated as an optimization problem over the model parameters to minimize the loss function.

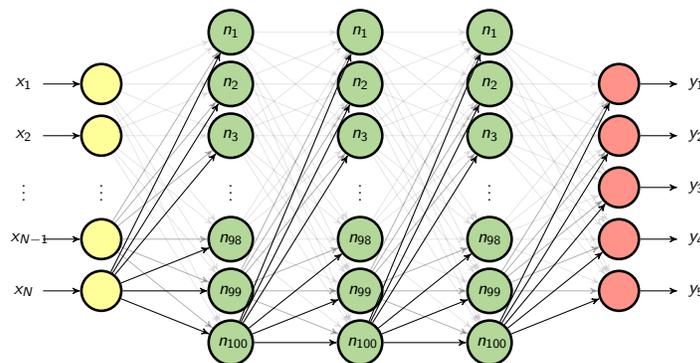


Figure 6.1: Architecture of a deep neural network with N input features, 3 hidden layers of 100 neurons each, and 5 output nodes.

A simple feed forward neural network architecture is defined by the number of hidden layers, the number of input, output, and hidden neurons, and the activation functions, as shown in Fig. 6.1. The output of the first hidden layer with n neurons is given by the affine transformation $A(\mathbf{x}) = W^T \mathbf{x} + \mathbf{b}$, followed by a nonlinear transformation by a monotonic activation function $h(\mathbf{x}) = s(A(\mathbf{x}))$. This operation is then cascaded over all the L hidden layers $\hat{f}(\mathbf{x}) = h_L(\dots(h_1(\mathbf{x})))$. The architecture is set by the choice of several hyperparameters, which are fixed before training. Thus, at the moment of training, the neural network output $\hat{f} = NN_{\mathbf{w}}$ depends only on the weights \mathbf{w}^2 . The training then consists in finding the optimal

²From now on, the term *weights* will refer to both weights and biases, as one can always redefine the weight matrix to include the bias terms in the first row and append a 1 at the top of the input vector \mathbf{x} .

weights.

Most neural networks are trained using MLE. In order to do this, one typically specifies a probabilistic model $p(y|\mathbf{x}, \mathbf{w})$. The cost function is taken to be the NLL of the conditional distribution $p(y|\mathbf{x}, \mathbf{w})$, and the optimization procedure consists in finding the parameters \mathbf{w} that minimize this objective function, also called the *loss function*,

$$\mathbf{w}_{MLE} = \arg \min_{\mathbf{w}} -\log p(Y|X, \mathbf{w}). \quad (6.5)$$

To prevent the model from overfitting to the training data, regularization can be implemented. This is often achieved by introducing a prior over the weights and finding the maximum posterior probability,

$$\mathbf{w}_{MAP} = \arg \max_{\mathbf{w}} \log p(\mathbf{w}|D) \quad (6.6)$$

$$= \arg \max_{\mathbf{w}} \log p(D|\mathbf{w}) + \log p(\mathbf{w}). \quad (6.7)$$

As mentioned before, the cost function penalizes deviations from the prior predictions.

Consider the case of a regression task, where $y \in \mathbb{R}$. To express the aleatoric uncertainty associated to a random variable, one could assume the target variable y to be given by a deterministic function $f(\mathbf{x}, \mathbf{w})$ with additive Gaussian noise,

$$y = f(\mathbf{x}, \mathbf{w}) + \epsilon, \quad \epsilon \in \mathcal{N}(0, \beta^{-1}). \quad (6.8)$$

The objective of the training of a neural network would then correspond to find the mean $\hat{y} = NN_{\mathbf{w}}(\mathbf{x})$ or, equivalently, to find the Gaussian conditional probability distribution

$$p(Y|X, \mathbf{w}, \beta) = \prod_{n=1}^N \mathcal{N}(y_n | NN_{\mathbf{w}}(\mathbf{x}), \beta^{-1}). \quad (6.9)$$

Defining the cost function as the NLL of Eq. 6.9 and removing the terms not dependent on \mathbf{w} , the optimal parameters are found by minimizing the following objective function:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} -\log p(Y|X, \mathbf{w}, \beta) = \arg \min_{\mathbf{w}} \frac{\beta}{2} \sum_{n=1}^N \|y_n - \hat{y}_n(\mathbf{x}, \mathbf{w})\|^2 \quad (6.10)$$

Hence, minimizing the NLL is equivalent to minimizing the Mean Squared Error (MSE). If one now assumes a Gaussian prior on the weights of the form $\mathcal{N}(\mathbf{w}; 0, \frac{1}{\lambda} I^2)$, it can be shown that the log-prior term in the cost function corresponds to the weight decay penalty $\lambda \mathbf{w}^T \mathbf{w}$.

In practice, the training is divided into two steps. During the *forward pass*, the network reads a set of inputs \mathbf{x} , produces the outputs $\hat{y}(\mathbf{x})$, and evaluates the cost function $J(y, \hat{y})$. In this step information flows forward through the network. During the *backward pass*, information from the cost function flows backward through the network to calculate the gradient. The *back propagation* algorithm is used to compute the gradient of the cost function with respect to the model weights. The gradient is typically estimated on a mini-batch of m examples as

$$\mathbf{g} = \nabla_{\mathbf{w}} J(\mathbf{w}) = \frac{1}{m} \nabla_{\mathbf{w}} \sum_{i=1}^m L(\mathbf{x}^{(i)}, y^{(i)}, \mathbf{w}). \quad (6.11)$$

Here L is the per-example loss. Stochastic gradient descent is then implemented to update the model parameters in the direction of decreasing loss as,

$$\mathbf{w}^{new} = \mathbf{w} - \alpha \mathbf{g}, \quad (6.12)$$

where α is the *learning rate*, a hyperparameter fixed before training.

6.3 Hypothesis testing with profile likelihood ratio

In particle physics experiments one often looks for new signal processes that have not been observed before³. In order to make conclusions regarding an excess over the background prediction, or lack thereof, a frequentist statistical test is performed, where one quantifies the level of agreement of the data with a given predicted hypothesis H . The hypothesis to be tested is generally referred to as *null hypothesis* H_0 . In order to make a statement about the viability of the null hypothesis, this is compared to an *alternative hypothesis* H_1 . In general, the null-hypothesis is the hypothesis one wants to exclude. For the purpose of claiming discovery of a new signal when a data excess is observed, the null hypothesis H_0 is the background-only hypothesis. If no excess is observed, exclusion limits are set where the null-hypothesis is the signal-plus-background hypothesis to be excluded with a given confidence level.

P-value and significance

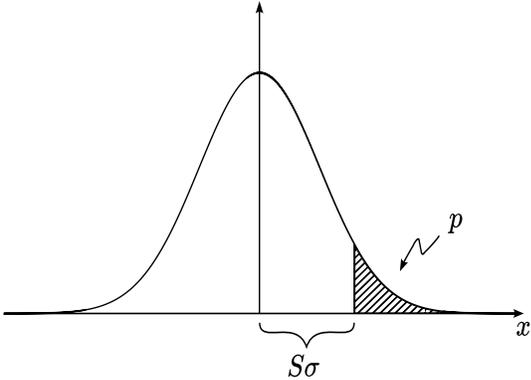
The concepts of significance and p -value are related and are often used in evaluating how well a given hypothesis describes the data. Suppose the background-only null-hypothesis H_0 and the new-physics-hypothesis H_1 predict two different PDFs, $f(x|H_0)$ and $f(x|H_1)$,

³This section is based on Ref. [161]

for a set of observations $\mathbf{x} = (x_1, \dots, x_N)$. Consider the observation of n events in data, which can consist of n_b events from known processes (background) and n_s events from a new process (signal). The background-only hypothesis predicts $n = n_b$, while the signal hypothesis predicts $n = n_s + n_b$.

The p -value of hypothesis H_0 is given by the probability, under the assumption of the hypothesis H_0 , to observe data with equal or of lesser compatibility with H_0 than the one actually observed (note that this is not the probability that H_0 is true). In their words, the p -value expresses the level of compatibility of the hypothesis H_0 with the observed data, and the weaker the compatibility, the more likely it is that H_0 can be rejected.

The significance S of a given p -value is often defined as the number of standard deviations that a Gaussian variable would fluctuate in one direction to give the same p -value:

$$\begin{aligned}
 p &= \int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - \Phi(S), \\
 S &= \Phi^{-1}(1 - p)
 \end{aligned}
 \tag{6.13}$$


where $\Phi(S)$ is the inverse of the cumulative distribution of the standard Gaussian. Note that the p value is defined for a standardized Gaussian centered at 0 and with $\sigma = 1$. The tradition in particle physics is that the threshold to report *evidence* of a new signal is $p < 0.003$, or a significance of $S = 3$, while it is $p < 2.87 \times 10^{-7}$, or a significance of $S = 5$, to report a *discovery*. To exclude a signal hypothesis one requires a p -value of 0.05, corresponding to a 95% confidence level and a significance of $S = 1.64$.

Likelihood parameters

The data is usually assumed to be a set of i.i.d. measurements x and the hypothesis is expressed as a PDF, with each hypothesis predicting a different PDF $f(x)$. Often a continuous set of hypotheses is considered $f(x; \mu)$, where each hypothesis is determined by the parameter μ , called the *parameter of interest*. For instance, μ could be the signal strength that relates the true to the simulated signal cross section $\sigma_s = \mu\sigma_s^{\text{MC}}$. In particular, $\mu = 0$ corresponds to the background-only hypothesis and $\mu = 1$ to the background-plus-nominal-signal hypothesis.

Once the model is fixed, a likelihood function can be constructed $L(f(x; \mu))$, giving

the probability of the data given the hypothesis $f(x; \mu)$. The value $\hat{\mu}$ that maximizes the likelihood is the best fit estimator of μ .

Experimental and theory systematic uncertainties can affect the PDF of x , both in terms of shape and normalization. Their effect is encoded in the model via a set of *nuisance parameters* θ . Their values are unknown and must be estimated in the fit together with the parameter of interest μ .

Binned likelihood

Consider the case of one signal and one background simulated samples and a variable of interest x , e.g. the invariant mass distribution of the reconstructed signal resonance. Signal and background will have different PDFs of the variable x , $f_s(x; \theta)$ and $f_b(x; \theta)$.

If one constructs a histogram $\mathbf{n} = (n_1, n_2, \dots, n_N)$ with N bins of the variable x , the expectation value of the number of events in a given bin i is given by $E[n_i] = \mu s_i(\theta) + b_i(\theta)$, where $b_i = b_{\text{tot}} \int_{\text{bin } i} f_b(x; \theta_b)$, and similarly for s_i . The number of entries n_i in each bin is generally assumed to be Poisson distributed with mean ν_i , so that the joint likelihood function for all bins is given by the product of the Poisson probabilities in each bin. The likelihood $L(\mu, \theta)$ can then be expressed as,

$$L(\mu, \theta) = \prod_{i=1}^N \text{Pois}(n_i | \mu s_i(\theta) + b_i(\theta)) \prod_{\theta_k}^{\theta} \mathcal{N}(\theta_k | \theta_k^0, \sigma_k), \quad (6.14)$$

where Gaussian priors are included to constraint the k nuisance parameters θ . The priors act as a penalty term in the maximum likelihood fit, as a postfit value $\hat{\theta}_k \neq \theta_k^0$ decreases the likelihood.

Profile likelihood ratio test

In order to test a hypothesized value of μ one needs a test statistics q_μ [161]. This is often obtained from the profile likelihood ratio $\lambda(\mu)$, given by

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})} \quad (6.15)$$

The numerator is the *profile likelihood function*. The quantity $\hat{\theta}$ represents the ML estimate of θ conditional on the specified value of μ , and thus depends on μ . The denominator is the maximized unconditional likelihood function, where $\hat{\mu}$ and $\hat{\theta}$ are set to their MLE estimators. The denominator represents the global maximum, so that the ratio is always ≤ 1 , with λ closer to 1 implying a better agreement between the data and the given hypothesis $f(x; \mu)$.

Generally one assumes that the presence of a signal could only increase the observed number of events, so one defines

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})}, & \hat{\mu} \geq 0, \\ \frac{L(\mu, \hat{\theta})}{L(0, \hat{\theta}(0))}, & \hat{\mu} < 0, \end{cases} \quad (6.16)$$

where it is assumed that the best level of agreement for an observed value of $\hat{\mu} < 0$ occurs for $\mu = 0$. The test statistic q_μ is given by $q_\mu = -2 \ln \tilde{\lambda}(\mu)$.

The test statistic q_μ depends on the data, so it is itself a random variable described by a PDF $f(q_\mu|\mu)$ under the assumption of μ . To quantify the level of disagreement with the data one would like to calculate the p -value of a given observed value $q_{\mu,obs}$, but in order to evaluate this one needs to know the PDF $f(q_\mu|\mu)$. This can be approximated via Monte Carlo methods, where pseudo-experiments are performed by sampling the likelihood and generating toy datasets. However, it can be shown that, under the Wald's approximation, the PDFs assume the shape of χ_1^2 functions, such that the p -value can be expressed in terms of the cumulative distribution of a standard Gaussian as

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu|\mu) dq_\mu = 1 - F(q_\mu|\mu) = 1 - \Phi(\sqrt{q_\mu}). \quad (6.17)$$

A predefined critical threshold $\alpha = 0.05$ is often chosen, so that if the p -value is found to be $p_\mu < \alpha$, then the value of μ is excluded at a confidence level (CL) of $1 - \alpha = 95\%$. The upper limit on the signal strength can be found by solving for the value of μ at $p_\mu = 0.05$.

For the discovery of a signal, one tests the background-only hypothesis $\mu = 0$. In this case,

$$q_{\mu=0} = \begin{cases} -2 \ln \lambda(0), & \hat{\mu} \geq \mu, \\ 0, & \hat{\mu} < \mu \end{cases} \quad (6.18)$$

where q_μ is 0 if the data fluctuates downward, as an observed value $\hat{\mu} < \mu$ is not regarded as less compatible with the background-only hypothesis. The p -value can be calculated as

$$p_0 = \int_{q_{0,obs}}^{\infty} f(q_0|0) dq_0. \quad (6.19)$$

If no excess is observed, exclusion limits are set on the signal strength μ by excluding the

signal-plus-background hypothesis at a given confidence level. The test statistic is given by

$$q_{\mu=0} = \begin{cases} -2 \ln \lambda(\mu), & \hat{\mu} \leq \mu, \\ 0, & \hat{\mu} > \mu, \end{cases} \quad (6.20)$$

where q_{μ} is zero if the data fluctuates upward, as an observed value $\hat{\mu} > \mu$ would not be considered less compatible with the signal-plus-background hypothesis. The corresponding p -value is

$$p_{\mu} = \int_{q_{\mu, \text{obs}}}^{\infty} f(q_{\mu} | \mu) dq_{\mu}. \quad (6.21)$$

The upper limit on μ is given by the largest μ such that $p_{\mu} \leq \alpha$. Setting $p_{\mu} = \alpha$ and solving for μ , one obtains

$$\mu_{up} = \hat{\mu} + \sigma \Phi^{-1}(1 - \alpha), \quad (6.22)$$

where σ is the standard deviation of $\hat{\mu}$ and can be obtained via Monte Carlo methods or from Asimov data.

Chapter 7

Search for new heavy resonances decaying to two SM bosons in semi-leptonic final states

The search for new heavy resonances has been the focus of intense efforts at the LHC since it began operations. If these particles are produced in an LHC collision, it should be possible to reconstruct the four-vectors of their decay products and they should appear as a narrow resonance on the invariant mass distribution of the final state particles over a smoothly falling background. However, if such collision events exist, they are very rare, making designing this type of searches a non-trivial task. In this chapter the search for new heavy resonances decaying to a pair of Standard Model bosons in semi-leptonic final states is presented, including the new deep learning techniques developed to enhance the sensitivity of this type of searches¹.

7.1 The search for new heavy resonances

Several well motivated extensions of the Standard Model predict the existence of new heavy resonances appearing at the TeV scale that can couple to the Higgs, W, and Z bosons and could be produced in pp collisions at the LHC.

A class of these models, motivated by naturalness arguments, predicts additional vector gauge bosons and include composite Higgs [45, 46] and little Higgs [44] models. As experimental searches are not sensitive to all the parameters of a theory, these models are studied experimentally in the context of a general Heavy Vector Triplet (HVT) model [3], which is parametrized by a simplified Lagrangian with an additional $SU(2)$ triplet (see Sec. 3.3). A second class of models, the two-Higgs-doublet models (2HDMs) [4], predicts the simplest extension of the SM scalar sector, by including an additional scalar $SU(2)$ doublet, resulting in

¹The analysis discussed in this chapter was still ongoing at the time of this writing. Certain details might therefore evolve before the analysis is published.

five physical scalars (see Sec. 3.4). In addition, Randall-Sundrum (RS) models with warped extra dimensions [162] or the bulk RS model, predict new particles, including a spin-0 radion and the spin-2 Kaluza-Klein excitation of the graviton, which are used as additional benchmark signatures in this type of searches.

Several of these new heavy resonances are predicted to decay with significant branching ratios (BRs) to a vector boson and a Higgs boson (WH , ZH) or to pairs of vector bosons (WZ , WW , ZZ). In the following, these will be referred to as VH and VV processes, respectively. According to the decay mode of the SM bosons, the final state of these processes is referred to as *fully leptonic* if both bosons decay to a pair of leptons, *fully hadronic* if both bosons decay to two quarks, or *semi-leptonic* when one boson decays leptonically and one hadronically. The semi-leptonic final state is particularly advantageous, as one can benefit from the higher decay BR of the hadronic decay, while keeping a high trigger and selection efficiency thanks to the cleaner leptonic signature.

Previous searches have been performed in ATLAS in semi-leptonic final states for VH and VV processes separately. The VH analyses were performed using the 3.2 fb^{-1} [163], 36.1 fb^{-1} [164], and 139 fb^{-1} [165] datasets. The VV searches were performed using the 36.1 fb^{-1} [166, 167] and the 139 fb^{-1} [168] datasets. Similar searches in semi-leptonic final states have been performed in CMS as well, with the latest analyses using the 137 fb^{-1} dataset being a search for a new resonance decaying to $WZ/WW/WH$ [169] and to ZH final state [170]. ATLAS has also performed searches for the same process in other final states, including two fully hadronic searches based on an integrated luminosity of 139 fb^{-1} for VH [171] and VV [172]. CMS has performed a fully-hadronic VH search [173] and a ZH search in final states with two taus and two light leptons [174] with the 35 fb^{-1} dataset.

Statistical combinations of the available searches for different processes and in the different final states have also been performed with the 36.1 fb^{-1} analyses [175] and the ongoing effort with the 136 fb^{-1} searches [176], which includes also decays of heavy resonances directly into a pair of leptons. CMS has performed similar combination efforts at the beginning of Run 2 [177] and with the 35.9 fb^{-1} dataset [178].

Several small excesses have been observed in the latest publications, all below a local significance of three standard deviations (σ). In the VH analysis, the largest deviations from the SM expectations in the latest publications have been observed in the search for a pseudoscalar A , where an excess was observed around a mass of 500 GeV of 2.1 (1.9) σ in the ggA (bbA) channel, primarily originating from the 2 b -tag category in the 2-lepton channel. A similar excess was observed in the Z' search at the same 500 GeV mass. Other smaller excesses of 2 standard deviations were observed at a resonance mass of 2.2 TeV for the Z' search and 400 GeV for the W' search. In the VV analysis, an excess around an RS

radion mass of 1.5 TeV with a local significance of 2.8 standard deviations was observed, induced by the merged HP region in the 0-lepton channel. The pseudoscalar excess is of particular interest, as it was already present in the 36.1 fb^{-1} publication with a local (global) significance of 3.6 (2.4) standard deviations and a disagreement around a similar mass was also observed both in the $A \rightarrow t\bar{t}$ CMS search with 35.9 fb^{-1} [179] with a local (global) significance of 3.5 (1.9), and a local excess above 2σ is observed in the $A \rightarrow \tau\tau$ ATLAS search with 139 fb^{-1} [180]. It should be noted that the excess is around the threshold for top-pair production, where higher-order electroweak corrections to the SM $t\bar{t}$ production can become important and could induce misinterpreted distortions [?]. Nonetheless, it is worth investigating further.

Fig. 7.1 shows a summary of all ATLAS searches interpreted in a benchmark scenario for the MSSM Higgs sector 7.1a and for a Type I 2HDM 7.1b, while Fig. 7.2 shows the latest ATLAS summary of the mass exclusion limits from diboson searches for new HVT and RS bosons with the full Run 2 dataset. These figures are representative of the effort that ATLAS has devoted to looking for these new particles.

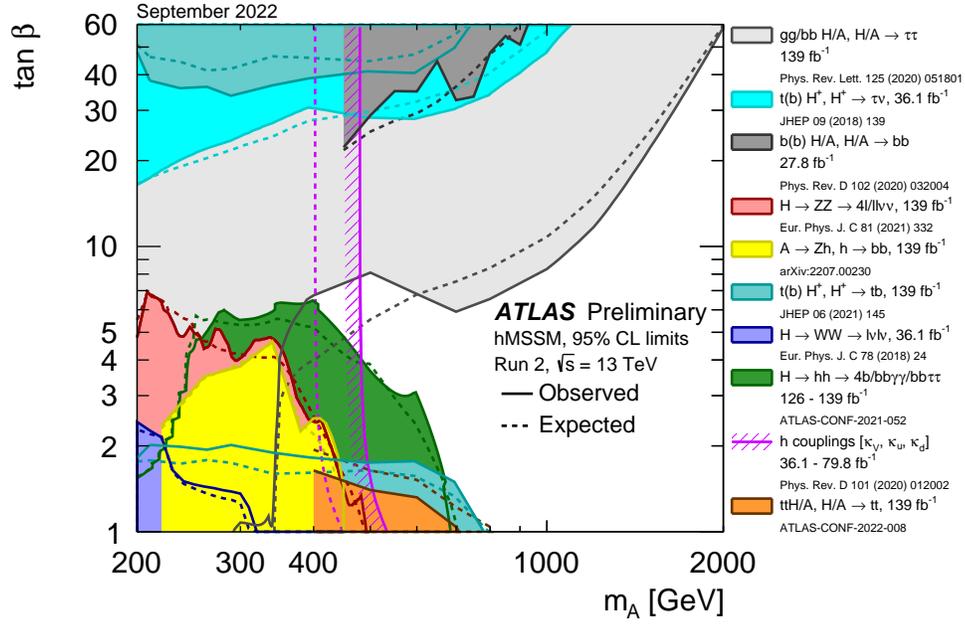
7.2 Analysis overview

This section presents the ongoing search for new heavy resonances decaying through VV and VH processes in the semi-leptonic final state. The leptonic decay of one vector boson proceeds as $Z \rightarrow \nu\bar{\nu}$, $W \rightarrow l^\pm \nu^{(\pm)}$, or $Z \rightarrow l^+l^-$, where l refers to a light charged lepton (electron or muon). The hadronic decay of vector bosons proceeds as $W \rightarrow qq$ and $Z \rightarrow q\bar{q}$, while the analysis targets only Higgs boson decays to a pair of b -quarks in order to capitalize on the large BR of the $H \rightarrow b\bar{b}$ decay channel ($\sim 57\%$)².

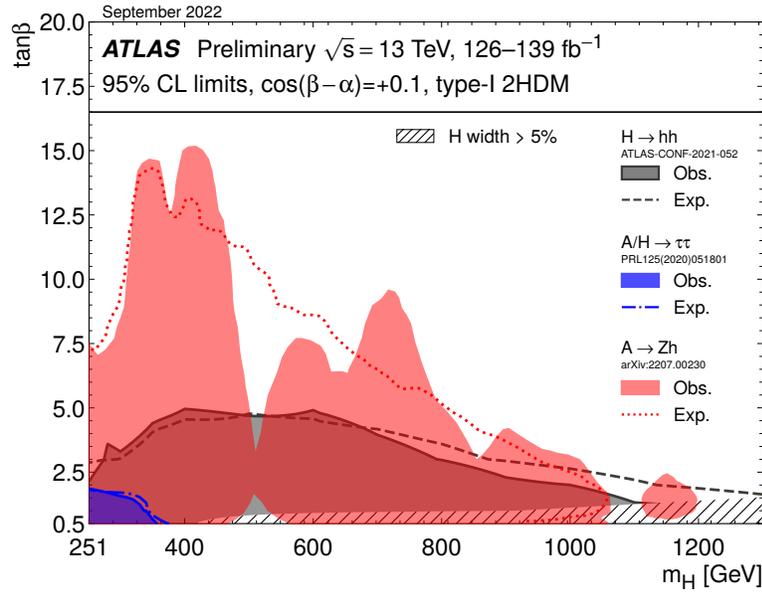
Different signal interpretations are considered. A search for new HVT bosons W' and Z' is performed in both the VV and VH final states, which motivated in part the combination of the VV and VH analyses into a single search. Specific to the VH analysis is also the search for a new pseudoscalar A predicted by the 2HDM model in the process $A \rightarrow ZH$. Lastly, specific to the VV analysis are the signal interpretations as a Kaluza-Klein Graviton and a RS radion, both of which can decay to WW and ZZ final states.

The search re-analyzes the 139 fb^{-1} dataset collected up until the end of Run 2. A new analysis of this dataset was motivated by several developments, including improved b -tagging algorithm, improved jet collections, and a new optimized event selection, including

²From now on, the references to particle/anti-particle state will be omitted, e.g. the W leptonic decay will be referred to as $W \rightarrow l\nu$.



(a)



(b)

Figure 7.1: Left: Regions of the $[m_A, \tan\beta]$ phase space excluded for a type of MSSM model by direct searches for new heavy Higgs bosons and by constraints from fits of the measured production and decay rates of the observed Higgs boson. Both the data (solid lines) and the expectation for the SM Higgs sector (dashed lines) are shown. Right: Regions of the $[m_H, \tan\beta]$ phase space excluded for a benchmark scenario of the Type I 2HDM by direct searches, comparing observed (filled) and expected (lines) limits. [181].

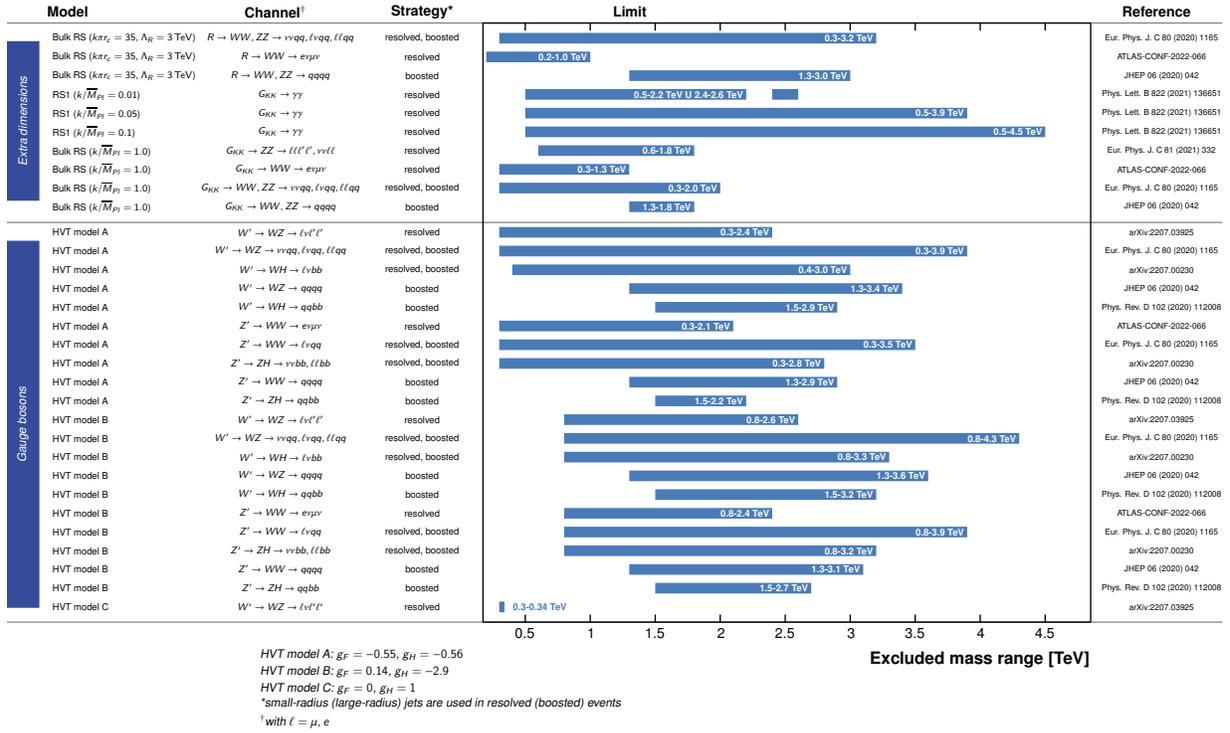


Figure 7.2: Summary of mass exclusion limits at 95% confidence level from ATLAS diboson searches with the full Run dataset [182].

the development of new machine learning techniques. In particular, the power of a search can be gauged by the expected significance, which can be approximated as,

$$\mathcal{S} = \frac{N_s}{\sqrt{N_b}} = \frac{\sigma_s \epsilon_s}{\sqrt{\sigma_b \epsilon_b}} \sqrt{\mathcal{L}}, \quad (7.1)$$

where \mathcal{L} is the total integrated luminosity, ϵ_s and ϵ_b are the signal and background efficiencies, and σ_s and σ_b are the predicted cross sections. By the end of Run 3, \mathcal{L} will have increased by a factor of 2. Assuming everything else constant, which means assuming that the hypothetical signal will be produced at the same rate as in Run 2, the observed Run 2 significances of ~ 2 will result in a significance of $2 \times \sqrt{2} = 2.8$, a very small increase that would still not qualify as evidence of new physics³. This value would be even lower if what was observed in Run 2 was just an upward fluctuation of the signal. It follows that in order to increase the physics reach of the search one cannot only rely on a larger dataset and methods to increase the analysis' signal efficiency are necessary. In practice, this means developing more

³The tradition in particle physics is that the threshold to report “evidence of a particle” is 3σ , and the standard to report a “discovery” is 5σ .

sophisticated data acquisition and analysis techniques. For this reason, while waiting for the delayed Run 3 dataset, the analysis effort focused on optimizing the analysis workflow, making extensive use of deep learning techniques, and preparing the groundwork for the future. In particular, a new analysis strategy based on deep-learning algorithms has been implemented, with several possible extensions envisioned for the future.

7.2.1 Analysis strategy

The analysis is conducted as a “bump search”, by looking for a localized data excess with respect to the known SM background in the distribution of the reconstructed resonance mass obtained from the selected $\nu\nu qq(bb)$, $\nu lqq(bb)$, or $llqq(bb)$ systems. In practice, the statistical interpretation is performed as a binned maximum likelihood fit (see Sec. 6.3) of the invariant mass distribution in all the signal regions (SR) and the background-dominated control regions (CRs).

The final regions of the analysis that enter the fit are defined via a series of cuts referred to as *event selection*. The selection is performed by applying requirements on the kinematic properties of the final state objects, or on event-level variables, to select regions of phase space close to what would be populated by the target signal. This process is dependent on the signal topology, which can be determined by the signal hypothesis, production mode, and final state. Each event topology has a specific event selection, which defines a *channel*. For each production mode, six channels are always defined according to the process being VV or VH, and to the number of charged leptons (0, 1, or 2) in the final state. Events in final regions can then be sorted into different categories, a process sometimes referred to as *categorization*. One reason to do this is to isolate particular signatures with different background contributions, which helps to better constraint the given background normalization. For example, separating according to the number of *b*-tagged jets allows to better isolate different V+jets contributions according to the number of light and heavy flavor jets, as shown later.

Different types of final regions are defined. *Signal regions* (SR) select events with the goal of maximizing the significance of the target signal, according to Eq. (7.1). *Control regions* (CR) are defined such that they target a region of phase space close to the SRs, but with negligible signal efficiency. This is usually done by using the same event selection as for the SRs, and then inverting one single cut that is expected to drastically remove signal. The use of CRs is two-fold. During the analysis optimization, when the analysis is blinded, they are used for the validation of background Monte Carlo (MC) modeling and of analysis techniques. They are then included in the final fit in order to constraint

background normalizations. *Validation regions* are defined similarly to CRs, but are used only for validation purposes and are not included in the fit.

Once the final regions have been defined, the discriminant distribution in each final region for data and background are input to the fit to test the background-only hypothesis. The background is provided by the MC simulation of the SM processes that pass the event selection and whose normalizations are mostly fixed by the fit using the CRs. The data is passed through the same event selection as MC. In case of an excess, the fit outputs a p-value, or the probability that the background can produce a fluctuation greater than the excess observed in data. When the background-only hypothesis cannot be excluded, upper limits on the signal cross section times branching fraction are set. Further constraints on specific model parameters can also be provided.

Once the channels have been analyzed independently, further improvements on the search sensitivity can be obtained by performing a combined fit, or a *combination*. For each signal hypothesis, a simultaneous analysis of the discriminants of all the channels sensitive to that hypothesis is performed. This provides several advantages. The first is an increase in the power of the search due to the fact that the total significance grows as the sum in quadrature of the significance in each bin entering the fit. Another advantage is the possibility to treat certain background contributions as correlated between different channels, which allows to better constraint their normalization and reduce the post-fit uncertainties, hence increasing the fit sensitivity. Different channels can also provide complementary information, so that a combined treatment results in an overall stronger sensitivity. This is the case, for instance, for the VH channel with 0 and 2 charged leptons, where the first provides stronger exclusion limits at high resonance masses, while the latter is more sensitive at low masses. Lastly, the inclusion of more bins also means stronger constraints on the parameters of the model under study.

7.2.2 Machine learning approach

Traditional *cut-based* analyses place hard cuts on individual variables to increase the signal purity of the final regions. While the significance is kept high, this is often at the expense of signal efficiency, which in previous searches of this type was below 20%. This is not surprising, as the final state of an LHC pp collision contains hundreds of particles. Even when one focuses on only a handful of objects, each of these is described by a four-dimensional four vector. The event selection therefore has to be optimized in $N > 16$ dimensions.

A more nuanced event selection can be obtained by moving to a deep-learning-based approach. Deep learning provides a way to optimally process a large number of correlated

inputs and find optimal decision boundaries. Consider a neural network (NN) with one output node that yields a probability score, serving as a classification metric. The score functions as a one-dimensional test statistic, $t(x)$, analogous to the traditional one-dimensional cuts. However, unlike simple threshold-based methods, the NN dynamically optimizes its internal weights to maximize the separation of the PDF of $t(x)$ for the signal and background hypotheses. In doing so, the NN is capable of learning any form of correlation among the input variables. In addition, a cut on $t(x)$ corresponds, in fact, to a non-linear decision boundary in the feature space, which can retain a higher signal efficiency and purity than a combination of hard cuts on the input variables. The strength of the NN lies therefore in its ability to map a high dimensional space onto a low dimensional output, while providing enhanced discrimination power. This motivated a more global machine learning approach in the analysis, including the development of a new multi-class NN jet classifier, one of the main contributions of this thesis.

The analysis makes also use of two other multivariate algorithms inherited and repurposed from previous publications. The vector-boson fusion (VBF) and gluon-gluon fusion (ggF)/Drell-Yan (DY) production modes are characterized by characteristic event topologies that require different final region definitions. In particular, the VBF topology has two additional jets, referred to as “VBF candidates,” that tend to be well separated in pseudorapidity and to have a large di-jet invariant mass. In the previous VV search [168], the VBF and ggF/DY final regions were made orthogonal via a Recursive Neural Network (RNN) [183], which uses as inputs the four-momenta of the small-R jets in the event identified as the “VBF candidates.” The same RNN is used in the current analysis, and is applied for the first time to VH final states as well. The analysis workflow is shown in Fig. 7.3.

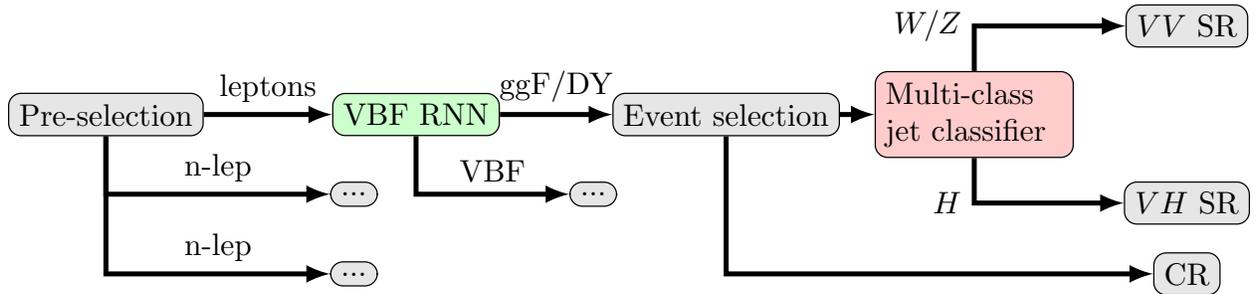


Figure 7.3: Analysis workflow displaying where the VBF-RNN and the MCT are applied.

7.2.3 The Multi-Class Tagger

A new five-class neural network was developed for the identification of the hadronic decay as coming from a Higgs boson, a W boson, a Z boson, a top-quark, or light-quarks and gluons

(QCD). The identification of the hadronic decay is an essential part of the search for new heavy resonances, both in the semi-leptonic and fully-hadronic final states. On one side, the SRs are generally defined to select hadronic decays of a H/W/Z boson. On the other, the major backgrounds that mimic these decays are mis-reconstructed top-quarks or energetic QCD jets, often with specific designated CRs. Having simultaneous access to the likelihood for all these hypotheses is therefore highly desirable.

The multi-class-tagger (MCT) was designed as a general tool in the context of boosted jet tagging. Individual taggers targeting specific signatures already exist in ATLAS, such as the top tagger [184] or the W/Z tagger [185], which attempt to identify a top quark or a vector boson from light jets. However, comparing scores from different taggers is potentially complicated, as the output scores are not correlated in a well understood way. The ambiguity in the interpretation of the scores is resolved if one moves to multi-class classification. Here, the output scores are by construction correlated, allowing for simultaneous scoring, in particular via the definition of likelihood ratios, as shown in Fig. 7.4b. Because the scores are correlated, the ratios are automatically well-defined and can bring significant improvements in terms of tagger performance, thanks to their ability to capture in part the uncertainty in the network predictions. As shown in Fig. 7.4c, likelihood ratios can also be used to access a multi-class space, which can provide further discrimination power. Multi-class classification was already used in ATLAS in the context of flavor tagging (see Sec. 4.4.6), but at the time of this work it had not been explored by ATLAS for boosted jet tagging. A multi-class approach has already been used by CMS [186].

The fact that the VV and VH final states are considered simultaneously in a single analysis was a second motivation for the development of the MCT. While harmonizing the efforts permits a better optimization of the event selection in anticipation of a combination of the results, considering different channels simultaneously can increase the complexity of the analysis. In particular, standard cut-based analyses can incur in the issue of overlapping selection criteria. In the case of the $VV + VH$ effort, the jet mass windows overlap, which requires an extra step to orthogonalize the final regions. Since the VV and VH processes differ only by the hadronic decay, the multi-class tagger provides an optimal way to solve this issue.

The next sections will discuss the development of the MCT for large-R jet classification, its extension to the resolved jets topology, and its deployment in the analysis to orthogonalize the VV and VH final regions. Although not discussed in this thesis, the way the MCT was envisioned allows for a straightforward extension to aid in the definition of top- and QCD-enriched control regions. Output scores of the MCT would also be candidate high-level inputs to a possible event-level classifier. These are ideas that will be explored in the future.

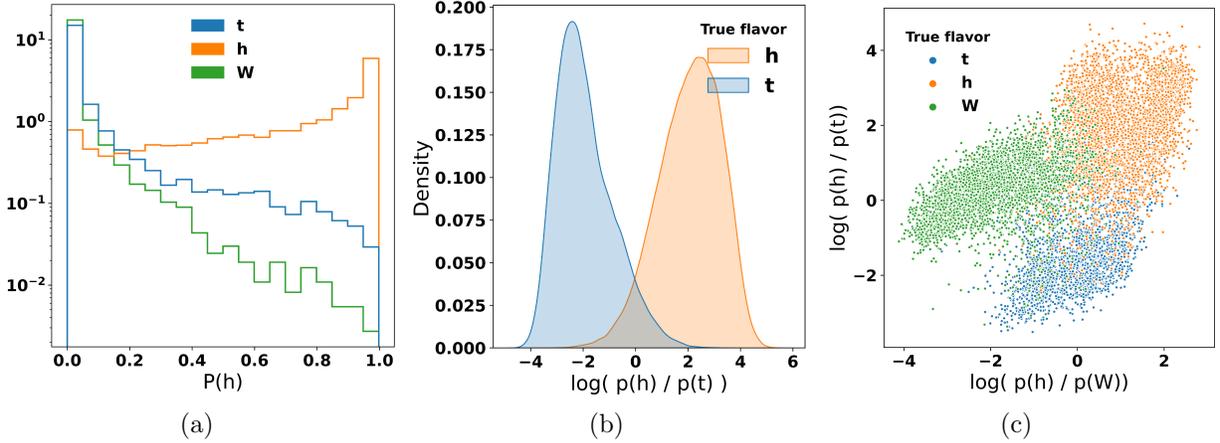


Figure 7.4: Example of multi-class classification showing: the raw scores for true Higgs, W, and top jets (left); the a likelihood ratio $\log(p(h)/p(t))$ for true Higgs and top jets (center); and the simultaneous evaluation of all the likelihoods in the two-dimensional space of likelihood ratios (right). These figures were produced with true Monte Carlo data and with the merged MCT described in the following section, but with an additional cut on the p_T of the large-R jet for displaying purposes.

7.3 Signal and background processes

The search targets different signal interpretations, with different production modes. Both DY (or quark-antiquark annihilation) and VBF productions are considered in the search for new HVT bosons. In the search for a pseudo-scalar A, ggF and b-quark associated production are considered. For the spin-0 RS radion and the spin-2 graviton, both ggF and VBF mechanisms are studied. Representative Feynman diagrams of the different production modes are shown in Fig. 7.5 for a general new resonance X .

In the HVT (see Sec. 3.3) signal interpretation, X can be a new electrically charged W' or electrically neutral Z' vector boson. The possible decay modes are $W' \rightarrow WZ$, $W' \rightarrow WH$, $Z' \rightarrow WW$, and $Z' \rightarrow ZH$. The two resonances are assumed degenerate, which favors a common interpretation of the results. The coupling of the new particles to the SM bosons is parametrized by the combination $g_H = g_V c_H$, while the couplings to fermions by $g_F = g^2/g_V c_F$. The parameters c_H and c_F are expected to be of order unity, so the parameter g_V represents the typical strength of the interaction. The results are interpreted with respect to three benchmark models. *Model A* predicts comparable fermionic and bosonic BRs and is representative of a weakly coupled model. *Model B* is representative of a composite model with the couplings to fermions suppressed. Lastly, *Model C* is representative of a fermiophobic scenario, with $g_V = g_H = 1$ and the couplings to fermions set to zero. For Models A and B, the W' and Z' bosons are produced mainly via

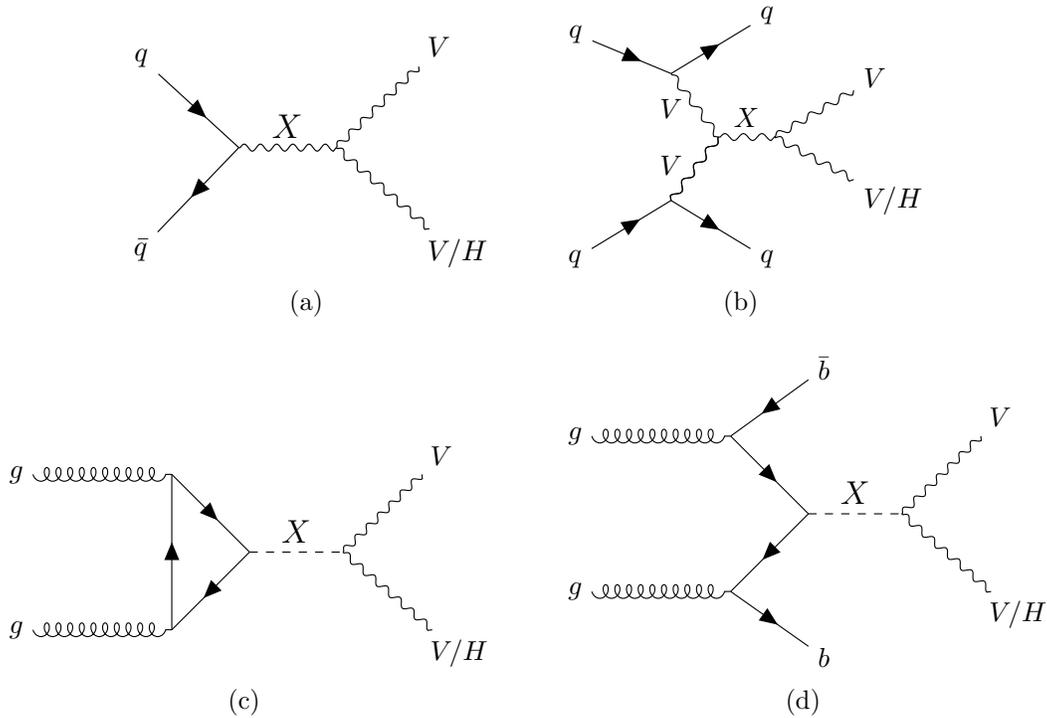


Figure 7.5: Representative Feynman diagrams for Drell-Yan (a), vector-boson fusion (b), gluon-gluon fusion (c), and b -quarks associated production (d) of a new heavy resonance X . When multiple options are possible for quark flavor or for vector boson charge, these are left unspecified.

DY. For *Model C* this mode is vetoed, making production via VBF enhanced. This is the first time VBF production is considered in the VH analysis in ATLAS. The analysis aims at setting upper limits on the production cross-section of the new particles, which can be used to constraint the model parameters g_F and g_H . The search is performed in the mass range from 300 GeV to 5 TeV.

Specific to the VH analysis is the search for a new pseudoscalar scalar A , one of the heavier Higgs bosons predicted by the 2HDM model (see Sec. 3.4). The resonance can decay to a ZH final state. The search is performed in the mass range between 220 GeV and 2 TeV. Higher masses are excluded by the class of models targeted by this search, as they make the Higgs potential unstable. The search aims at setting limits on the production cross section, which is then used to constrain the model parameters $\tan(\beta)$ and $\cos(\beta - \alpha)$. The search targets both ggF and b -quarks associated production (bbA).

Specific to the VV analysis are two other signal interpretations. One is the radion (R), a new neutral scalar particle predicted by certain RS models [187, 188]. The other is a neutral spin-2 graviton (G_{KK}) [162, 189], the first Kaluza-Klein (KK) excitation in a bulk

RS model. Both are predicted to have dominant BRs to WW or ZZ final states and can be produced via ggF or VBF processes. This search is performed in the mass range between 300 GeV and 5 TeV.

A summary of the decay modes and channels of interest for each signal interpretation is given in Tab. 7.1.

Production	Process	Channels
HVT bosons		
DY	$pp \rightarrow Z' \rightarrow WW/ZH$	VV 1-lepton and VH 0/2-lepton
	$pp \rightarrow W' \rightarrow WZ/WH$	VV 0/1/2-lepton and VH 1-lepton
VBF	$pp \rightarrow Z'jj \rightarrow WW/ZHjj$	VV 1-lepton and VH 0/2-lepton
	$pp \rightarrow W'jj \rightarrow WZ/WHjj$	VV 0/1/2-lepton and VH 1-lepton
Pseudoscalar A		
ggF	$pp \rightarrow A \rightarrow ZH$	VH 0/2-lepton
bbA	$pp \rightarrow Abb \rightarrow ZHbb$	VH 0/2-lepton
Radion/Graviton		
DY	$pp \rightarrow R/G \rightarrow WW/ZZ$	VV 0/1/2-lepton
VBF	$pp \rightarrow R/Gjj \rightarrow WW/ZZjj$	VV 0/1/2-lepton

Table 7.1: Channels used in the searches for HVT bosons, pseudoscalar A , radion, and graviton.

Several SM processes can have similar final states as the signals and act therefore as background: W and Z boson production in association with jets (V +jets); top quark production, with top-quark pair production ($t\bar{t}$) as the primary contribution, but including also single-top-quark production; non-resonant diboson production (WW , WZ , or ZZ) with semi-leptonic decays; and multi-jet production. Other minor background processes for the VH topology are the production of $t\bar{t} + h$, $t\bar{t} + V$ ($V = W, Z$), and the irreducible SM background $V + h$.

All Monte Carlo (MC) samples are generated at the center-of-mass energy of $\sqrt{s} = 13$ TeV and are passed through the full GEANT4-based [131] ATLAS detector simulation. All samples include the simulation of in-time and out-of-time pileup by overlaying the simulated minimum bias events on the generated event, matching the pileup conditions of the different data taking periods. The MC production undergoes the same event reconstruction as data. A multiplicative factor to the event weight of the generated events is applied to correct for differences between data and MC. These include corrections of the jet energy scale and

resolution, of the triggering, reconstruction, and identification efficiency of leptons, and of the jet flavor-tagging efficiencies.

The HVT Z' and W' production via quark-antiquark annihilation was modelled at LO accuracy in QCD with `MadGraph5` (MG5) [190] generator, using the NNPDF2.3LO PDF set [191], interfaced with `PYTHIA8` [192] for modeling of the parton shower with the ATLAS A14 set of tuned parameters [193]. Different samples were generated assuming various W' and Z' masses ranging from 500 GeV to 5 TeV. For benchmark models A and B, only samples for model A were generated, as the differences in the final state kinematics are considered negligible once detector response effects are taken into account. Only the predicted production and decay rates differ, which are fixed at the moment of the statistical interpretation. The generated samples include decays of the Higgs boson to both b - and c -quarks, where the SM values of $\mathcal{B}(h \rightarrow b\bar{b}) = 0.569$, $\mathcal{B}(h \rightarrow c\bar{c}) = 0.0287$, and $m_H = 125$ GeV were assumed. Another set of samples is generated for model C for VBF production only.

The 2HDM ggA signal sample was generated with MG5 at LO accuracy in QCD with the narrow width approximation, using the `2HDM_GF_FeynRules` model [194], and the NNPDF2.3 LO PDF set. The 2HDM bbA process was generated using the four-flavor scheme at next-to-leading order (NLO) with massive b -quarks with `MadGraph5_aMC@NL02.2.3` and the NNPDF2.3NLO PDF set. Shower modeling was performed with `PYTHIA8` with A14 tuning. Resonance masses in the range between 220 GeV and 2 TeV were simulated for each signal process.

Signal samples for the RS graviton and radion were produced with MG5 interfaced to `PYTHIA8` using the NNPDF2.3LO PDF. For each interpretation, samples were produced for masses ranging from 300 GeV to 6 TeV.

The QCD multi-jet background is not well-modeled by MC and it is generally derived from data. In the context of the analysis, it would appear as a mis-modelling when comparing data and MC distributions. However, the event selection of the analysis is able to select a phase space with negligible multi-jet contamination. A summary of the MC generators used to produce the other background processes is given in Tab. 7.2.

Process	Generator
Vector boson + jets	
$W \rightarrow l\nu$	SHERPA2.2.1
$Z \rightarrow ll/\nu\nu$	SHERPA2.2.1
Top quark	
$t\bar{t}$	POWEHEG+PYTHIA8
single top	POWEHEG+PYTHIA8
Wt -channel	POWEHEG+PYTHIA8
$t\bar{t} + h$	MADGRAPH5_AMC@NLO + PYTHIA8
$t\bar{t} + V$	MADGRAPH5_AMC@NLO + PYTHIA8
Diboson	
$qg/q\bar{q} \rightarrow WW \rightarrow l\nu qq$	SHERPA2.2.1
$qg/q\bar{q} \rightarrow WZ \rightarrow llqq/\nu\nu qq/l\nu q\bar{q}$	SHERPA2.2.1
$qg/q\bar{q} \rightarrow ZZ \rightarrow llq\bar{q}/\nu\nu q\bar{q}$	SHERPA2.2.1
$gg \rightarrow WW \rightarrow l\nu qq$	SHERPA2.2.2
$gg \rightarrow ZZ \rightarrow llq\bar{q}/\nu\nu q\bar{q}$	SHERPA2.2.2
$qg/q\bar{q} \rightarrow ll\nu\nu$	SHERPA2.2.2
V + SM Higgs	
$qq \rightarrow Wh \rightarrow l\nu bb$	POWHEG+PYTHIA8
$qq \rightarrow Zh \rightarrow \nu\nu bb/llbb$	POWHEG+PYTHIA8
$gg \rightarrow Zh \rightarrow \nu\nu bb/llbb$	POWHEG+PYTHIA8

Table 7.2: Summary of the MC generators used to produce the various background processes. Adapted from analysis ATLAS internal note.

7.4 Data taking and trigger selection

The analysis uses pp collision data recorded by ATLAS during the 2015, 2016, 2017, and 2018 runs at a center of mass energy of $\sqrt{s} = 13$ TeV. Only events during which all ATLAS sub-detectors were fully operational are included. The resulting total integrated luminosity collected during this period is $139.0 \pm 2.4 \text{ fb}^{-1}$. The breakdown of the integrated luminosity per data taking period is shown in Tab. 7.3.

Year	$\mathcal{L}[\text{fb}^{-1}]$
2015+2016	36.2
2017	44.3
2018	58.5
Total	139

Table 7.3: Integrated luminosity for each data taking period.

The event selection relies on the lowest unrescaled single-lepton and E_T^{miss} (MET) triggers, according to the lepton channel. Different triggers were used according to the data taking period due to the evolving pileup conditions during Run 2. The full list of triggers is shown in Tab. 7.4.

The 0-lepton channel relies on different combinations of MET triggers, which rely on different *online* E_T^{miss} reconstructed at the High-Level Trigger, as well as different thresholds. In particular, the online MET reconstruction does not include muon information. The MET calculation of the `xe` trigger uses all noise suppressed cells from the LAr and Tile calorimeters. The `mht` trigger uses the jet based E_T^{miss} , where the MET is calculated using all the calorimeter jets reconstructed at the HLT, which have been energy-corrected for pileup contribution. The `pufit` trigger uses the `pufit` algorithm [98], which groups topoclusters into towers of size $\eta \times \phi \approx 0.71 \times 0.79$ that are subtracted with an event-dependent pileup correction. The latter reconstruction was found to be optimal as the pileup levels increased during Run 2. Since the E_T^{miss} triggers reach 100% efficiency at offline E_T^{miss} values of roughly 200 GeV, the 0-lepton channel only extends down to masses of 500 GeV (which corresponds roughly to $E_T^{\text{miss}} \sim 250$ GeV).

The 2-lepton channel uses single-electron and single-muon triggers, defined by different requirements on the E_T of the reconstructed HLT lepton, as well as lepton identification and isolation criteria. In most periods a logical `OR` of different settings is used, as at higher E_T values quality criteria can be relaxed to increase efficiencies, and vice versa.

The 1-lepton channel uses combinations of single-electron, single-muon, and MET triggers. Because the MET triggers do not include muons in the calculation, they will trigger on an event with a high p_T muon, hence compensating for single-muon trigger inefficiencies.

After passing the trigger requirements, the data events go through a cleaning procedure. Events are removed if they are deemed corrupted due to LAr noise burst and data corruption, or incomplete events. All events are also required to only have “clean” reconstructed jets. A procedure called “jet cleaning” identifies “bad jets” built from noisy calorimeter cells or non-collision background. Because jets affect the calculation of other objects in the event, such as E_T^{miss} , events with one or more unclean jet are removed.

Data-taking period	$evqq$ and $eeqq$ channels	$\mu\nu qq$ ($p_T(\mu\nu) < 150$ GeV) and $\mu\mu qq$ channels	$\mu\nu qq$ ($p_T(\mu\nu) > 150$ GeV) and $\nu\nu qq$ channels
2015	HLT_e24_lhmedium_L1EM20 OR HLT_e60_lhmedium OR HLT_e120_lhloose	HLT_mu20_iloose_L1MU15 OR HLT_mu50	HLT_xe70
2016a (run < 302919) ($L < 1.0 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$)	HLT_e26_lhtight_nod0_livarloose OR HLT_e60_lhmedium_nod0 OR HLT_e140_lhloose_nod0 HLT_e300_etcut	HLT_mu26_livarmedium OR HLT_mu50	HLT_xe90_mht_L1XE50
2016b (run \geq 302919) ($L < 1.7 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$)	same as above	same as above	HLT_xe110_mht_L1XE50
2017	same as above	same as above	HLT_xe110_pufit_L1XE55
2018	same as above	same as above	HLT_xe110_pufit_xe70_L1XE50

Table 7.4: List of triggers used in the analysis. Adapted from analysis ATLAS internal note.

7.5 Object selection

The same object reconstruction and selection procedure is performed on data and simulated MC samples. A detailed description of ATLAS event reconstruction is provided in Sec. 4.4, while the focus in this section will be in the quality criteria used to select well-reconstructed objects for the analysis.

Tracks

Tracks are reconstructed from hits in the ID using the Primary tracking algorithm with the **Tight** track quality selection and the **Tight** track-vertex association criteria. All tracks are required to have $|\eta| > 2.5$ and $p_T > 5$ GeV.

Electrons

Electrons are reconstructed from topological clusters in the electromagnetic calorimeter matched to tracks in the ID. All electron candidates are required to have $p_T > 7$ GeV and

$|\eta| < 2.47$, excluding the gap region between the barrel and the endcap LAr calorimeters $1.37 < |\eta| < 1.52$. Additionally, each electron track is required to have $|z_0 \sin \theta| < 0.5$ mm and transverse impact parameter significance $|d_0|/\sigma_{d_0} < 5$. Two identification criteria are used in the analysis. The **Tight** selection is used in the 1-lepton channel to select electrons from the decay $W \rightarrow e\nu$. This relies on tight identification and isolation criteria, and requires electrons with $p_T > 30$ GeV. The **Loose** criteria is used to select $Z \rightarrow e^+e^-$ electrons in the 2-lepton channel, with looser identification and isolation working points, and no isolation requirement for electrons with $p_T > 100$ GeV.

Muons

Muons are reconstructed from combined tracks using information from both the ID and the muon spectrometer. All muon candidates are required to have $p_T > 7$ GeV and $|\eta| < 2.5$, as well as $|z_0 \sin \theta| < 0.5$ mm and $|d_0|/\sigma_{d_0} < 3$. Similarly to electrons, two identification criteria are used. The **Tight** selection is used in the one-lepton channel to select muons from the decay $W \rightarrow \mu\nu$. This applies the medium identification working point and tight isolation criteria, and requires muons with $p_T > 30$ GeV. The **Loose** selection is used to select $Z \rightarrow \mu^+\mu^-$ decays in the 2-lepton channel, with looser identification and isolation working points, and no isolation requirement for electrons with $p_T > 100$ GeV.

Small-R jets

Small-R jets are reconstructed using the anti- k_t algorithm with radius parameter $R = 0.4$ and PFlow input objects. The topological clusters used to reconstruct the jet have been calibrated at the EM scale.

Different selection criteria are used for what will be referred to as signal and forward jets. Signal jets are reconstructed in the central η region ($|\eta| < 2.5$) and are required to have $p_T > 20$ GeV. Forward jets are reconstructed in the forward region ($2.5 < |\eta| < 4.5$) and are required to have $p_T > 30$ GeV. To reduce contamination from jets originating from pileup vertices, jets with $p_T < 60$ GeV and $|\eta| < 2.4$ are further required to pass a jet-vertex tagging [195] selection. Selected small-R jets are referred to as *signal jets*.

Large-R jets

The large-R jets are built with the anti- k_t algorithm with $R = 1.0$ using UFO input objects. As described in Sec. 5.3.2, the resulting UFO jets benefit from the optimal performance of PFlow jets at low p_T and of TCC jets at high p_T , and are expected to improve the performance of the analysis with respect to previous publications, which relied on the standard LCW calibrated topoclusters (see Sec. 4.4.4) as inputs.

Prior to jet reconstruction, the set of input objects is pre-processed with a combination of constituent-level pileup-suppression algorithms (see Sec. 5.5). The p_T of each constituent is first adjusted with the Constituent Subtraction (CS) method. The Soft Killer (SK) algorithm, with a grid granularity of $\Delta\eta \times \Delta\phi = 0.6 \times 0.6$, is then used to remove low p_T constituents. Further pileup suppression is obtained by applying the Soft-drop algorithm, with parameters $\beta = 1.0$ and $Z_{\text{cut}} = 0.1$, on the set of reconstructed jets, removing constituents associated with soft and wide-angle radiation.

Variable radius track jets

Variable radius (VR) track jets are used to identify b -tagged subjets in large-R jets (see Sec. 5.3.2). After reconstruction, they are assigned to the large-R jets in the event via ghost-association. VR track jets are built by running the anti- k_t algorithm on the tracks using a p_T -dependent radius parameter given by $R_{\text{eff}}(p_{T,i}) = \frac{\rho}{p_{T,i}}$, with $\rho = 30$ GeV and the upper and lower limit on the jet size set to $R_{\text{max}} = 0.4$ and $R_{\text{min}} = 0.02$. All VR jets are required to have $p_T > 10$ GeV, $|\eta| < 2.5$, and number of associated tracks $\text{nTrk} > 1$. Collinear track jets can occur and are problematic, as their interplay with the track-association step used by b -tagging algorithms is not well understood. In order not to expose b -tagging algorithms to these pathological cases, events with an overlap between track jets used for b -tagging ($p_T > 5$ GeV and $\text{nTrk} > 1$) and VR track jets selected by the analysis are removed.

Flavor-tagging

The DL1r flavor-tagging algorithm is used to tag signal jets and VR track jets. A cut on the DL1r score as defined in Eq. 4.9 is used to identify jets as b -tagged. The cut at the 70% b -tagging efficiency working point (WP) is used for signal jets, while the VR track jets are selected using the 85% WP. A higher efficiency WP corresponds to an increase in both signal and background acceptance.

Missing transverse momentum

Different metrics exist to evaluate the presence of a large momentum imbalance in the transverse plane, which are described in Sec. 4.4. The analysis makes use of the missing transverse momentum E_T^{miss} given by the sum of a hard term and a track-based soft term to reconstruct neutrinos in the event. Additional selection requirements make use of a track-based missing transverse momentum estimation p_T^{miss} , built from the negative vectorial sum of the transverse momenta of all the tracks associated to the primary vertex. In order to decrease contributions from background with large E_T^{miss} , which can arise from mis-measurements of leptons and jets energies, the E_T^{miss} significance is also used.

τ -leptons

Hadronically decaying τ -lepton candidates are used in the $\nu\nu b\bar{b}$ channel to reject background with real hadronic τ -leptons. Hadronic τ candidates [196] are reconstructed using $R = 0.4$ calorimeter jets. They are required to have one or three associated tracks, $p_T > 20$ GeV and $|\eta| < 2.5$. The τ identification is performed using a multivariate technique algorithm and the `Medium` working point is used in this analysis.

Overlap removal

As the different object collections are reconstructed and selected independently, it is possible that different objects are built from the same inputs. In order to avoid double counting of energy, an overlap removal procedure is implemented on the set of selected objects. First, a τ -lepton is removed if it overlaps with a muon within $\Delta R < 0.2$, unless the muon is not a combined muon and the τ has $p_T > 50$ GeV. Electrons are removed if they share an ID track with a muon. A small-R jet is removed if it is within $\Delta R = 0.2$ of an electron or a muon that passed isolation requirements. In order to retain jets originating from a true b -hadron decay that included muons in the decay chain, jets are only removed if they have fewer than three associated tracks, or if more than 70% of the transverse momentum sum of the associated tracks comes from the muon. Lastly, electrons and muons are removed if they overlap with any of the remaining jets within $0.2 < \Delta R < \min(0.4, 0.04 + 10 \text{ GeV}/p_T)$.

7.6 Event selection

After the final state objects have been reconstructed, and events with good data collection and object reconstruction quality have been selected, the next step is to define the final regions of the analysis. The event selection goes from general requirements common to all regions, to more specific selections targeting each decay channel.

The first step consists in reconstructing the leptonic decay. The events are separated into lepton channels, targeting the corresponding leptonic decays, according to the number of reconstructed charged leptons: $Z \rightarrow \nu\nu$ in the 0-lepton channel; $W \rightarrow e\nu$ or $W \rightarrow \mu\nu$ in the 1-lepton channel; and $Z \rightarrow ee$ or $Z \rightarrow \mu\mu$ in the 2-lepton channel.

The next step is to reconstruct the hadronic candidate. According to the signal hypothesis, this is either a Higgs boson (H), or a W or Z boson. In the former case, the region is said to be in the “VH analysis,” while in the latter it is said to be in the “VV analysis.” The hadronic reconstruction differs in terms of mass window and b -tagging requirements. According to whether the reconstruction of the hadronic decay uses small-R or large-R jets, the region is categorized as *resolved* or *merged*. This separation is not by itself orthogonal.

The two reconstruction strategies proceed in parallel, producing merged and resolved final regions, and only at the end the regions are made orthogonal by prioritizing one over the other. This is referred to as *prioritization* and it is performed to maximize the analysis sensitivity. The reconstruction of the leptonic decay is the same for VV and VH final regions, but can differ between resolved and merged, as a larger boost in the hadronic decay is generally accompanied by a more boosted leptonic system.

A last set of selection cuts is specific to each lepton channel, analysis, and kinematic regime and aims at reconstructing the full resonance decay from the reconstructed final state to obtain the invariant mass distribution to feed into the statistical fit.

7.6.1 Jet requirements

A set of kinematic cuts on the reconstructed jets is used to select events compatible with a H , W , or Z hadronic decay. The selections differ according to the reconstruction strategy.

Resolved regime

The $H/W/Z$ candidate is reconstructed by first selecting the small-R jets that are most compatible with the given decay hypothesis, and then summing their 4-vectors. Events are first required to have the leading small-R jet p_T above 45 GeV. The W/Z candidate is reconstructed from the two leading small-R jets in the event. The H candidate is reconstructed from the two leading b-tagged small-R jets or, in the case of only one b-tagged jet, from the b-tagged jet and the leading non-b-tagged jet. The use of b-tagging in the selection of the Higgs boson candidate allows to significantly reduce background contamination, which results in better sensitivity. Although the decay $Z \rightarrow b\bar{b}$ has a sizable BR, the use of b -tagging was not observed to bring a significant improvement in this case. The reconstructed dijet system (jj) is then required to have a reconstructed mass consistent with the H , W , or Z hypothesis. The following mass windows are used:

- Higgs boson: $110 < m_{jj} < 140$ GeV (0/1 lepton), $100 < m_{jj} < 145$ GeV (2-lepton)
- W boson: $62 < m_{jj} < 97$ GeV
- Z boson: $70 < m_{jj} < 105$ GeV

Merged regime

The $H/W/Z$ candidate is taken as the leading large-R jet (J) in the event. Events are required to have the leading large-R jet with $p_T > 250$ GeV ($p_T > 200$ GeV) in the VH (VV) analysis. In the VH analysis, the Higgs candidate is selected using a mass window

requirement of $75 \text{ GeV} < m_J < 145 \text{ GeV}$. The mass windows in the merged regime of the VV analysis are defined using the p_T -dependent W and Z mass cuts from the W/Z Tagger [185]. The W/Z tagger provides p_T -dependent two-dimensional cuts in the large-R jet mass and D_2 substructure variable to tag W or Z candidates against multi-jet background. The mass windows as a function of p_T are shown in Fig. 7.6. Events passing both the mass window and D_2 cuts are defined as **High-Purity (HP)**, while events that pass the mass cut, but fail the D_2 cut are classified as **Low-Purity (LP)**.

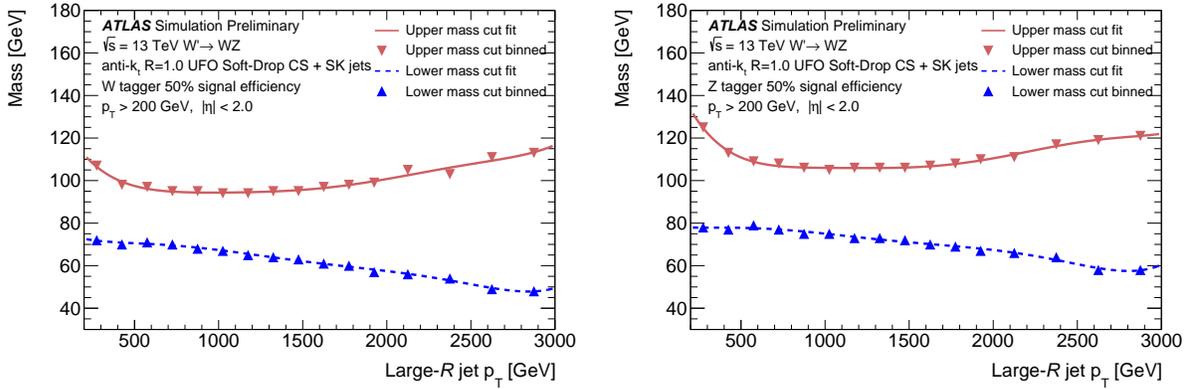


Figure 7.6: Upper and lower cut values on $m(J)$ for the cut-based W tagger (left) and Z tagger (right) in bins of large-R jet p_T [185].

7.6.2 0-lepton channel

The 0-lepton selection targets $Z \rightarrow \nu\nu$ decays. Events are required to have no Loose leptons and $E_T^{\text{miss}} > 250 \text{ GeV}$. The largest contamination in this region comes from the QCD-multijet background. Further cuts are applied specifically to suppress this contribution:

- The object-based MET significance, as defined in Sec. 4.4.5, is required to be $\mathcal{S} > 10$.
- The reconstructed $\mathbf{E}_T^{\text{miss}}$ is required to be isolated by requiring $\min[\Delta\phi(\text{jet}, \mathbf{E}_T^{\text{miss}})] > \frac{\pi}{9}$, where ϕ is the angle between $\mathbf{E}_T^{\text{miss}}$ and the nearest small-R jet.
- The track-based missing transverse momentum p_T^{miss} is required to be above 80 GeV.

The decay of the Z boson to two neutrinos does not allow the complete reconstruction of the $Z \rightarrow \nu\nu$ candidate, as the z -component of the four vector is not known. For this reason, the final discriminant is taken to be the transverse mass of the reconstructed ZH/ZZ/ZW candidate. This is obtained by summing the MET vector $(E_T^{\text{miss}}, E_x^{\text{miss}}, E_y^{\text{miss}}, 0)$ with the four-vector representing the H/W/Z candidate without the longitudinal components.

7.6.3 1-lepton channel

The 1-lepton channel targets $W \rightarrow l\nu$ decays. Events are required to have exactly one lepton satisfying the **Tight** criteria for either electrons or muons, and no additional **Loose** leptons. In order to select W decays, events are further required to have $E_T^{\text{miss}} > 60(100)$ GeV and the W candidate $p_T > 75(200)$ GeV in the resolved (merged) region.

The neutrino reconstruction only provides the transverse components of its four-vector. In order to reconstruct the full W four-vector, the z -component of the neutrino p_z^ν is obtained by imposing a W mass constraint on the lepton-neutrino system, via the relation $m_W^2 = (p_l + p_\nu)^2$. The result is given by,

$$p_{z,\nu} = \frac{1}{2p_{T,l}^2} \left[p_{z,l}A + E_l \sqrt{A^2 - 4(p_{T,l}^2 (E_T^{\text{miss}})^2)} \right], \quad (7.2)$$

with $A = m_W^2 + 2p_x^l E_x^{\text{miss}} + 2p_y^l E_y^{\text{miss}}$. In case of complex solutions, the real solution is taken. If both solutions are real, the smaller one is taken.

Additionally, a set of requirements are imposed to remove background contributions. The requirement

$$\frac{\min(p_T(W_{lep}), p_T(W/Z/H_{had}))}{m(VV/VH)} > 0.35 \quad (7.3)$$

is used to select events with an even p_T sharing between the hadronic and leptonic decay systems, to target signal-like two-body decays.

Contamination from $t\bar{t}$ is further suppressed by removing events with additional b -tagged signal jets not used to reconstruct the dijet system in the resolved regime, or with b -tagged VR jets outside the large-R jet in the merged regime.

In the resolved channel, several angular requirements are further applied to suppress QCD-multijet contributions. The cuts $\Delta\phi(l, E_T^{\text{miss}}) < 1.5$ and $\Delta\phi(j_1, j_2) < 1.5$ select events with well contained leptonic and hadronic decays, respectively. The cuts $\Delta\phi(l, j_1/j_2) > 1.0$ and $\Delta\phi(E_T^{\text{miss}}, j_1/j_2) > 1.0$ select events with well separated leptonic and hadronic decays. In addition, when the lepton is identified as an electron, the additional requirement $E_T^{\text{miss}}/p_T(W) > 0.2$ is applied in the resolved region. The contribution of QCD background in the merged region is less significant, and no specific anti-QCD cut is implemented.

7.6.4 2-lepton channel

The 2-lepton channel selects events compatible with a $Z \rightarrow ll$ decay. Events are required to have exactly two isolated `Loose` leptons of the same flavor (electrons or muons) and either of the two leptons has to be matched to the HLT lepton that fired the trigger and its E_T is required to be 5% above the trigger threshold to ensure full trigger efficiency. The leading lepton is required to have $p_T > 27$ GeV and the sub-leading lepton is required to have $p_T > 20(25)$ GeV in the resolved (merged) regime. In the case of two muons, these are required to have opposite charge. This requirement is not imposed on electrons because of their higher rate of charge misidentification due to possible converted photons from bremsstrahlung radiation.

The Z candidate is reconstructed as the four-vector sum of the two leptons. The invariant mass of the dilepton system is required to be consistent with the Z boson mass in order to suppress backgrounds without a resonant dilepton pair. Electron pairs are required to have $m_{ee} \in [83, 99]$ GeV, while a $p_{T,ll}$ -dependent cut is required for muon pairs to compensate for the di-muon mass resolution degradation at high Z transverse momentum. The following cut was optimized in the previous VV publication to maintain approximately a constant 95% selection efficiency across resonance masses:

$$(85.6 - 0.0117 \cdot p_{T,ll}) < m_{\mu\mu} < (94.0 + 0.0185 \cdot p_{T,ll}) \text{ GeV} \quad (7.4)$$

The same p_T balance requirement as in Eq. (7.3) is used in the region $m(VV/VH) < 320$ GeV to further suppress background:

$$\frac{\min(p_T(Z_{lep}), p_T(W/Z/H_{had}))}{m(VV/VH)} > 0.35 \quad (7.5)$$

At higher signal masses the background contamination is sufficiently low, so the cut is removed to recover signal efficiency.

The VH analysis has significant contributions from the $t\bar{t}$ background, which is often characterized by the presence of a neutrino in the final state. This contamination is suppressed by requiring the object-level MET significance to be $\mathcal{S} < 4$, removing events consistent with the presence of E_T^{miss} .

The signal resonance is obtained from the four-vector sum of the Z candidate from the dilepton system and the reconstructed hadronic decay. The final discriminant is given by the invariant mass of the reconstructed resonance.

7.7 Event categorization

The event selection just described selects events with signal-like topologies and is used to define the final regions that will be input to the statistical fit. However, at this point the regions are not orthogonal, neither between resolved and merged categories, nor between different production modes of a given signal hypothesis, and further criteria have to be imposed.

Final regions

In a given analysis, the searches targeting ggF and DY production processes have similar event topologies, and thus they share the same final regions. The bbA production is characterized by the presence of two extra b -quarks in the final state, while the VBF process has two extra quarks.

For the definition of VV/VH VBF-HVT final regions a recurrent neural network (RNN) is used, as mentioned earlier. The RNN was developed in the context of the previous VV search and was redeployed in this round for both VV and VH. The RNN takes as input up to two extra jets in the event and outputs the probability of the event as being VBF or not-VBF. Events are removed from any HVT final region and put in VBF-HVT final regions if they have an RNN score above 0.8.

In the VH analysis, in the searches for W' and Z' via DY or VBF production and in the search for A via the ggF process, only events with exactly one or two b-tagged jets are considered. Events with 0-btagged jets are discarded, as the background contamination is too large to add sensitivity to the search. The bbA signal interpretation is targeted by requiring at least one extra b -tagged jet. In the resolved region, events are required to have at least three b -tagged signal jets ($3 + b - tags$), while in the merged region only events with additional b -tagged VR track jets outside the large-R jet are considered ($2b - tag + 1 + add.$). For each lepton channel, the VH signal regions are classified according to the merged or resolved reconstruction of $H \rightarrow b\bar{b}$ decay and to the number of b -tagged jets.

In the VV analysis, for regions targeting $Z \rightarrow qq$ decays, events with zero or one b-tagged jet are combined in the same category (0/1), while events with 2 b-tags make up a different category to increase the significance by targeting $Z \rightarrow b\bar{b}$ decays. Search channels with a hadronically decaying W are analyzed in the inclusive region. As mentioned previously, in VV merged signal regions, events are sorted into **HighPurity** and **LowPurity** according to the quality of the reconstructed $V \rightarrow J$ candidate. In summary, for each lepton channel the VV final regions are classified according to: merged and resolved reconstruction of $V \rightarrow qq$ decay; high and low purity of merged V reconstruction; 0/1 or 2 b-tagged jets in $Z \rightarrow qq$

decays; and mass windows of $Z \rightarrow qq$ or $W \rightarrow qq$ decays.

Prioritization strategy

Events are separated in the resolved and merged category according to the reconstruction strategy. The wide range of masses targeted by the analysis results in a wide range of transverse momenta of the decay products. At low transverse momenta, the two quarks from a H/W/Z decay are well separated and can be reconstructed as individual small-R jets, called *resolved* reconstruction. However, as the boost of the mother particle increases, the daughter particles become increasingly collimated, until they cannot be resolved as two individual jets anymore. This causes a drop in signal efficiency, if one continues to rely on the small-R jets reconstruction strategy. The efficiency can be recovered by reconstructing the hadronic decay using a larger jet radius. This was indeed the motivation to introduce large-R jets and will be referred to as *merged* reconstruction. The interplay between resolved/merged reconstruction and signal efficiency is shown in Fig. 7.7 for the previous VH 2-lepton channel analysis. In practice, each event is reconstructed with both strategies. Most events are

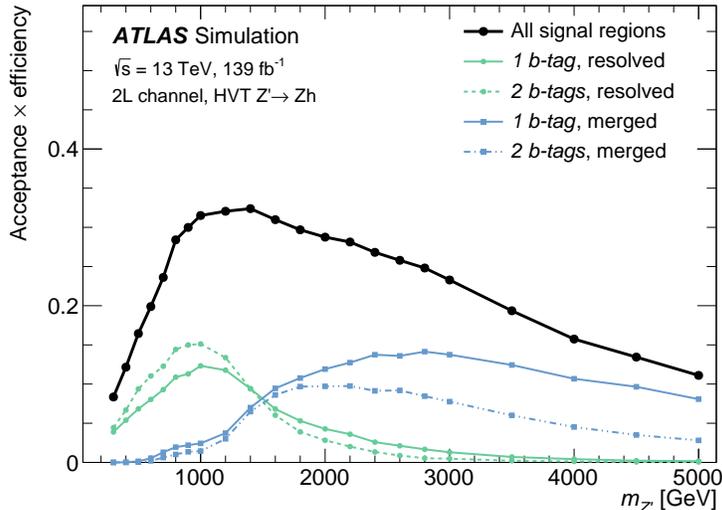


Figure 7.7: Acceptance \times efficiency in the VH 2-lepton analysis for $Z' \rightarrow Zh$ signal as a function of the resonance mass in latest VH publication [165].

better reconstructed with only one of the two strategies, so that when the alternative reconstruction is used, it produces a poorly reconstructed event that is rejected by the analysis. However, for a subset of events in the intermediate kinematic region, both strategies provide equivalent signal efficiencies. In this case, it is possible for an event to enter both the resolved and merged signal regions. In order to remove this overlap, a *prioritization strategy* is implemented.

It was found that the VH analysis reaches higher sensitivity by prioritizing the resolved region, while the VV analysis performs better by prioritizing the merged region. These strategies will be referred to as `PriorityResolved` and `PriorityMerged`, respectively. For instance, the VH `PriorityResolved` strategy is implemented as follows:

- If an event is in a Resolved SR, it is removed from any other Merged SR or CR.
- Else, if an event is in a Merged SR, it is removed from any other CR.
- Else if an event is in a Resolved CR, it is removed from any other Merged CR.

Similarly, in VV final regions, the order of selection is as follows: Merged HP SR \rightarrow Merged LP SR \rightarrow Resolved SR \rightarrow Merged HP CR \rightarrow Merged LP CR \rightarrow Resolved CR.

VV and VH Orthogonality

The VV and VH HVT signal regions are included in the same statistical fit for the HVT interpretation and are therefore required to be orthogonal. Because the jet mass windows overlap, this is generally not the case. The regions are made orthogonal using the Multi-Class Tagger (MCT). This will be discussed in detail in Sec. 7.10, but in general the MCT sorts the events between VV and VH final regions, according to whether the hadronic decay is deemed more likely to be coming from a W or Z boson, in the first case, or from a Higgs boson, in the second.

Control regions

Control regions (CRs) are used to constraint the normalization of the most dominant backgrounds in the final fit. All CRs are common to the VV and VH analyses. *Mass sideband* CRs are obtained by inverting the mass window requirement of the given signal region. More specifically, events are required to have mass values within $50 \text{ GeV} < m_{jj/J} < 200 \text{ GeV}$, but outside the VV and VH mass windows. A special control region targeting $t\bar{t}$ background and referred to as top-CR (TCR) is used in the 2-lepton channel. This is obtained by inverting the lepton flavor requirement, requiring leptons of opposite flavor.

7.8 Boosted jets Multi-Class Tagger (MCT)

The merged MCT is a deep neural network (DNN) large-R jet classifier trained to identify a jet as originating from a Higgs boson, a W boson, a Z boson, a top quark, or light quarks and gluons (q/g). This section presents the training and testing performance of the MCT, while the deployment in the context of the analysis is discussed in later sections.

7.8.1 Training

Jet reconstruction and selection

The DNN was trained using pileup-suppressed UFO $R = 1.0$ jets. The specific jet collection is the same as the one used in the analysis and described in Sec. 7.5. In order to focus on the kinematic region of interest for the analysis, only large-R jets with $m^{\text{reco}} \in [50, 200]$ GeV, $p_T^{\text{reco}} \in [200, 3500]$ GeV, and $|\eta|^{\text{reco}} < 2.0$ are selected. The VR track jets are reconstructed using the anti- k_t algorithm on tracks with $p_T > 10$ GeV and $|\eta| < 2.5$. The DL1r algorithm is applied on each VR track jet to provide the output probabilities for the jet to be coming from a b -quark $p(b)$, a c -quark $p(c)$, or light-quarks and gluons $p(qg)$.

MC samples and truth labeling

The training samples were Monte Carlo generated samples. Signal samples enriched in $W/Z/H$ /top-tagged jets are obtained from simulations of heavy BSM resonances decaying into boosted SM particles. The truth labeling relies on truth matching⁴ and further reconstruction quality criteria.

The QCD sample is obtained from multijet processes, where the jets are produced via the strong interaction, and represent light quark and gluon jets. To provide discrimination power over a wide range of large-R jet p_T , it is important that the training samples include a large number of events up to very high p_T regimes. For this reason, the di-jet samples are generated in bins of p_T , so that each bin is sufficiently populated. Similarly, the BSM samples are generated for different BSM resonance masses to span a wide range of p_T for the daughter particles. The effect of multiple pp interactions is also included in the simulation.

Jets from $H \rightarrow b\bar{b}$ decays are generated with $G \rightarrow HH$ processes, where G is a Randall-Sundrum graviton. The events are simulated using PYTHIA8 with the ATLAS A14 tune and the NNPDF2.3LO PDF. Only reconstructed Higgs candidates truth-matched to a true Higgs particle and with two ghost-associated b -hadrons are selected.

⁴All jets are truth matched by first dR matching the jet to a truth jet, and then dR matching the truth jet to a truth particle.

A sample of hadronically decaying top quarks is obtained from simulated $Z' \rightarrow t\bar{t}$ decays, generated with PYTHIA8 and NNPDF2.3LO PDF using the A14 tuning. The truth-labeling strategy to select well-reconstructed top-quark jets requires the two truth top quarks to be well separated with $\Delta R(t, \bar{t}) > 2.0$. Inclusive top decays are selected by truth-matching the reconstructed top-jet to the generator-level top-quark using $\Delta R < 0.75$. Only contained tops, with all the decay products contained in the large-R jet, are selected by requiring the truth jet to be truth matched also to the W boson with $\Delta R < 0.75$. The ungroomed jet mass is required to be $m > 140$ GeV and at least one b -quark is required to be ghost-associated to the ungroomed truth jet. The large-R jet is also required to have

$$\text{Split}_{23}/1.e3 > \exp[3.3 - 6.98e - 04 \cdot p_T/1.e3]. \quad (7.6)$$

W and Z boson jets are obtained from $W' \rightarrow WZ$ decays, where only hadronically decaying W/Z are considered. The samples are generated with PYTHIA8 and NNPDF2.3LO PDF using the A14 tuning. To select isolated jets, the W and Z bosons are required to have $\Delta(W, Z) > 2.0$. Reconstructed jets are required to be truth matched to the true W/Z bosons using $\Delta R < 0.75$. The ungroomed truth jet mass is required to be above 50 GeV and to pass the following cut on the d_{12} k_t -splitting scale:

$$\sqrt{d_{12}} > 55.25 \cdot \exp \left[\frac{-2.34 \cdot 10^{-3}}{\text{GeV}} p_T \right]. \quad (7.7)$$

Input variables

The DNN inputs include the kinematics of the large-R jet and the kinematics and b -tagging information of the associated VR track-jet, for up to three leading track jets. The flavor tagging information is provided by the raw scores of the DL1r algorithm, which represent the probability of the given track jet to be a b -, c -, or q/g -jet. The variables describing the kinematics of the large-R and track jets originally included the mass, transverse momentum, and (η, ϕ) coordinates, in order to allow the full four-vector reconstruction. In order to remove a potential artificial η -dependence between the true class of the jet and the boost and spin of the simulated resonance, the η variable was removed, without a significant decrease in performance. Having removed the ability to reconstruct the four-momentum of the jet, the ϕ coordinated was also removed.

This input information was supplemented with several substructure variables (see Sec. 5.4). These are referred to as *high-level* variables, as they provide pre-processed information of the *low-level* inputs, such as tracks or topoclusters. The expectation is that the high-level inputs provide already optimized discrimination power which can aid the classifier, while the

performance will improve with respect to using the individual variables thanks to the ability of the network to learn correlations between the large number of inputs. Classifiers using *high-level* inputs have already been shown to bring an increased performance in the context of *b*-tagging, with the DL1r algorithm being itself an example.

The inputs were selected among an original set of 100 variables, of which the majority were substructure observables useful for one-, two-, and three-prong identification, including N-subjettiness variables and ratios, energy correlation functions (ECFs) with different β values, and their corresponding ratios. In particular, all variables that had already been seen to perform well in the context of top-tagging or *W/Z* tagging were included [184]. The original set was down-selected by removing variables which did not affect the network performance. This removed redundant information and allowed to keep the size of the training samples manageable. The final reduced list of inputs used for the training is shown in Tab. 7.5. Examples of input variables are shown in Figs. 7.8, ??, ??, and ??. As expected, the most powerful observable is the jet mass, which for a large-R jet originating from a heavy particle has a scale associated to the mass of the particle.

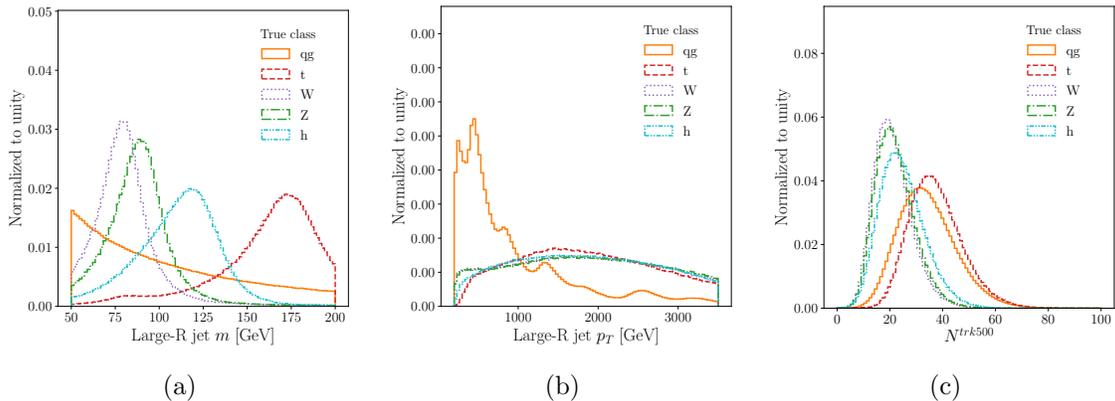


Figure 7.8: Subset of input variables to the Merged MCT describing the large-R jet.

Pre-processing

The list of inputs passed to the NN is fixed by definition. For cases where less than three VR track jets are associated to the large-R jet, the corresponding variables are set to 0.

As the input variables have widely different scales and units, standardization was a necessary step to obtain a satisfactory model performance. For each feature x , the mean μ and standard deviation σ were obtained only from the samples in the training dataset. Then, before being passed through the network, every event had its features standardized

Type	Observable	Definition
Large radius jet	p_T	Transverse momentum
	m	Mass
	N^{Const}	Constituents multiplicity
	N^{trk500}	Tracks multiplicities $p_T > 500$ MeV
3 leading track jets	p_T	Transverse momentum
	DL1r pb	Bottom quark probability
	DL1r pc	Charm quark probability
	DL1r pu	Light quark probability
	τ_2	N-subjettiness
	τ_3	
	τ_{21}	
	τ_{32}	
Substructure observables	C_2	Energy correlation functions
	D_2	
	$ECF(n = 1, \beta = 1)$ $ECF(n = 3, \beta = 1)$	
	Angularity	Other
	Aplanarity	
	PlanarFlow	
	FoxWolfram20	
	ZCut ₁₂	
	Split ₁₂	
	Split ₂₃	

Table 7.5: Merged MCT input variables

by applying the corresponding transformation

$$x' = (x - \mu)/\sigma. \tag{7.8}$$

Jets corresponding to different classes show different p_T distributions. In general, it is desirable to make the network’s decision independent of the p_T of the jet. In particular, QCD jets tend to have a lower transverse momentum, which would bias the network to believe that a high p_T jet is most likely not a qg -jet and a low p_T signal jet to be most likely a qg -jet.

In order to remove the dependence of the NN decision on the large-R jet p_T , the training samples were reweighted to obtain a flat large-R jet p_T distribution. The reweighting was performed for each class separately. In order to have the most accurate reweighting, the p_T

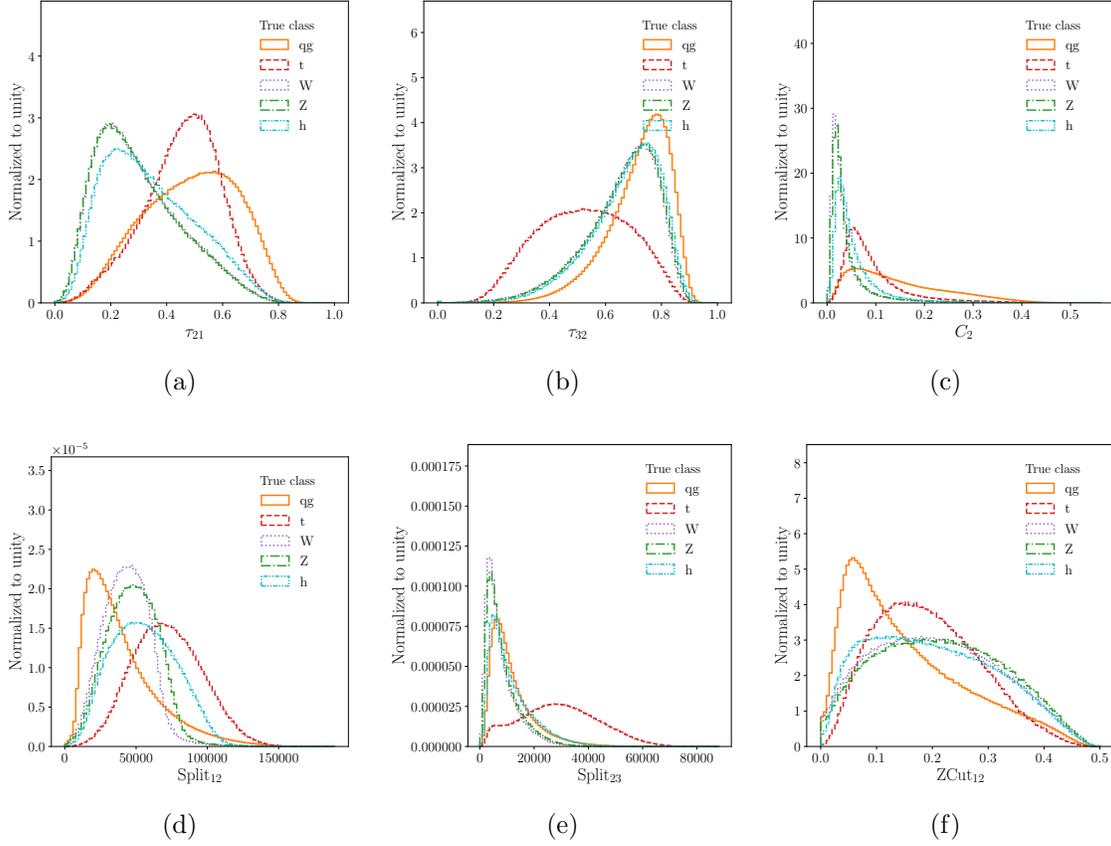


Figure 7.9: Subset of input variables to the Merged MCT describing the large- R jet sub-structure.

distribution was re-binned using the finest binning that allows to retain a statistical error below 5% in each bin. The p_T density was then made flat by assigning to each event in a bin i of width b and containing n events the weight $w_i^{\text{Flat } p_T} = 1/(n \cdot b)$. Lastly, in order to have a balanced class representation, the samples were reweighted to have an equal class normalization. This adds a constant multiplicative factor w^j to the weight of every event belonging to class j . After the reweighting procedure, each jet-event i true labeled, for example, as a Higgs boson is assigned a weight $w_i = w_i^{\text{Flat } p_T} \cdot w^{\text{Higgs}}$.

Hyperparameters and training

The model was trained using Keras with TensorFlow [197] backend. The number of samples used for training and validation was 9M and 4M, respectively. A dataset of 5.5M was holdout for testing. The architecture used is a fully connected DNN. A dropout layer was inserted between every hidden layer for regularization. The network was trained with the maximum number of epochs set to 500, but early stopping was implemented to interrupt

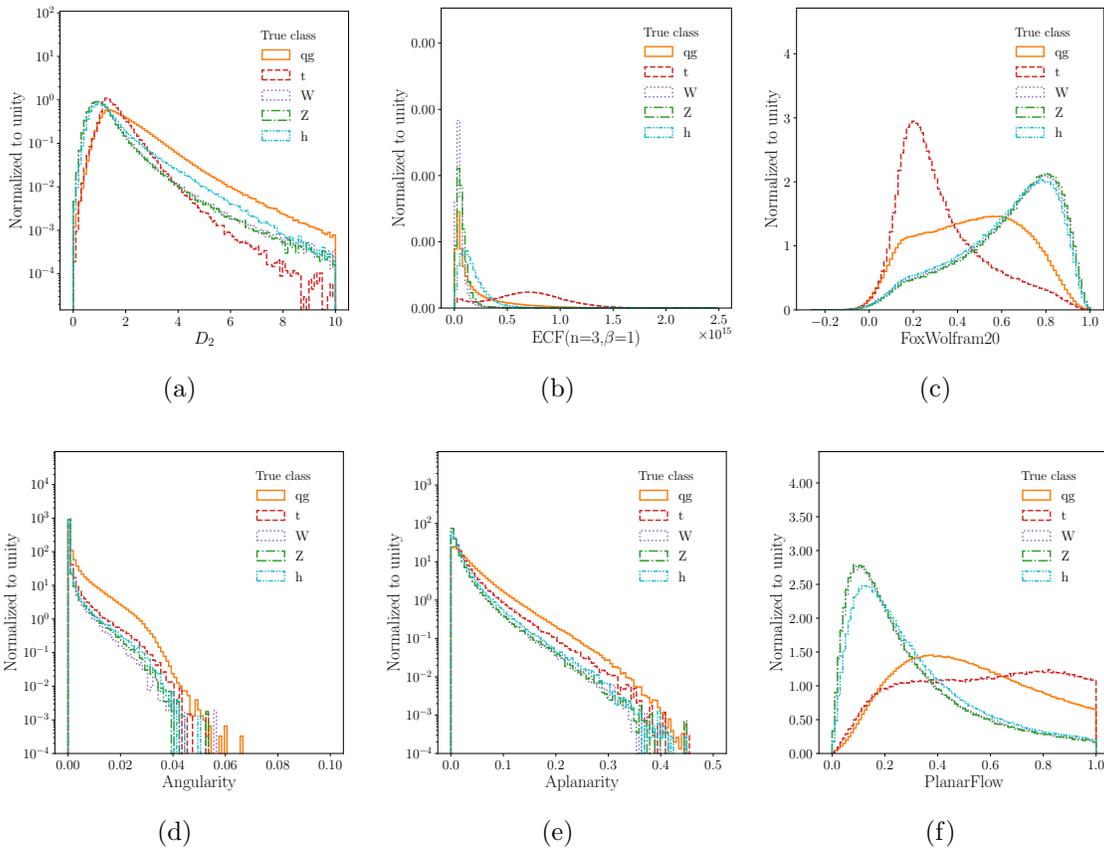


Figure 7.10: Subset of input variables to the Merged MCT describing the large-R jet substructure.

the training when no further reduction in the loss was observed for more than 40 epochs. The hyperparameters were optimized using a grid search and the final choice is shown in Tab. 7.6. The training and validation accuracy of the model was found to be 0.74.

Batch size	1000
Learning rate	0.0001
Dropout probability	0.1
Hidden layers	3
Nodes per hidden layer	200

Table 7.6: Merged MCT final choice of hyperparameters.

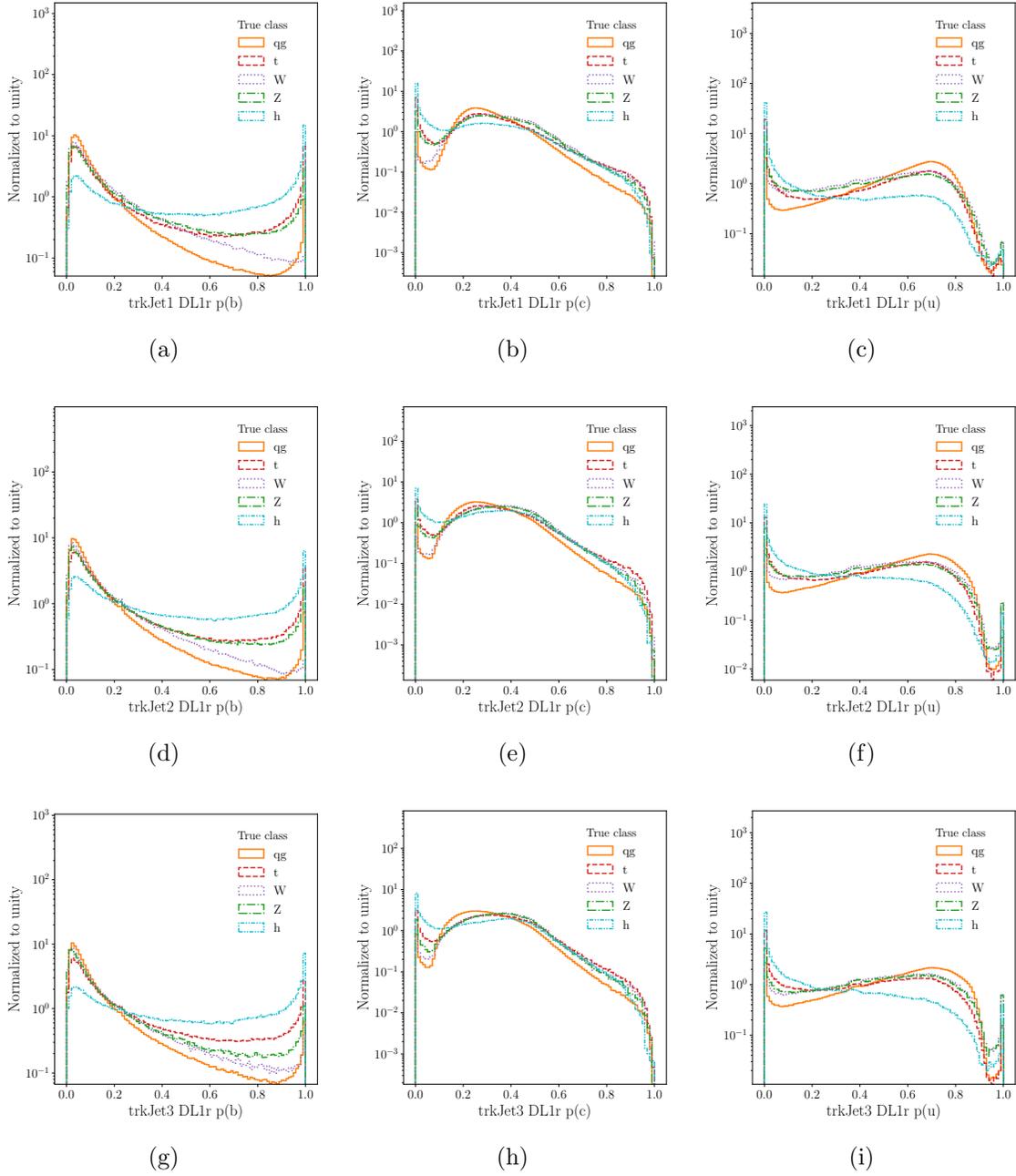


Figure 7.11: Subset of input variables to the Merged MCT describing the b -tagging DL1r scores of the three leading track jets inside the large-R jet.

7.8.2 Testing performance

Fig. 7.12 shows the output probabilities for all the events in the testing dataset separated by their true class label. The score of the given true class peaks at 1, while the corresponding background classes peak at 0, indicating the MCT is performing as expected. Similarly, the

plots in Fig. 7.13 show the log-likelihood ratios for different true class combinations, showing significant discrimination power.

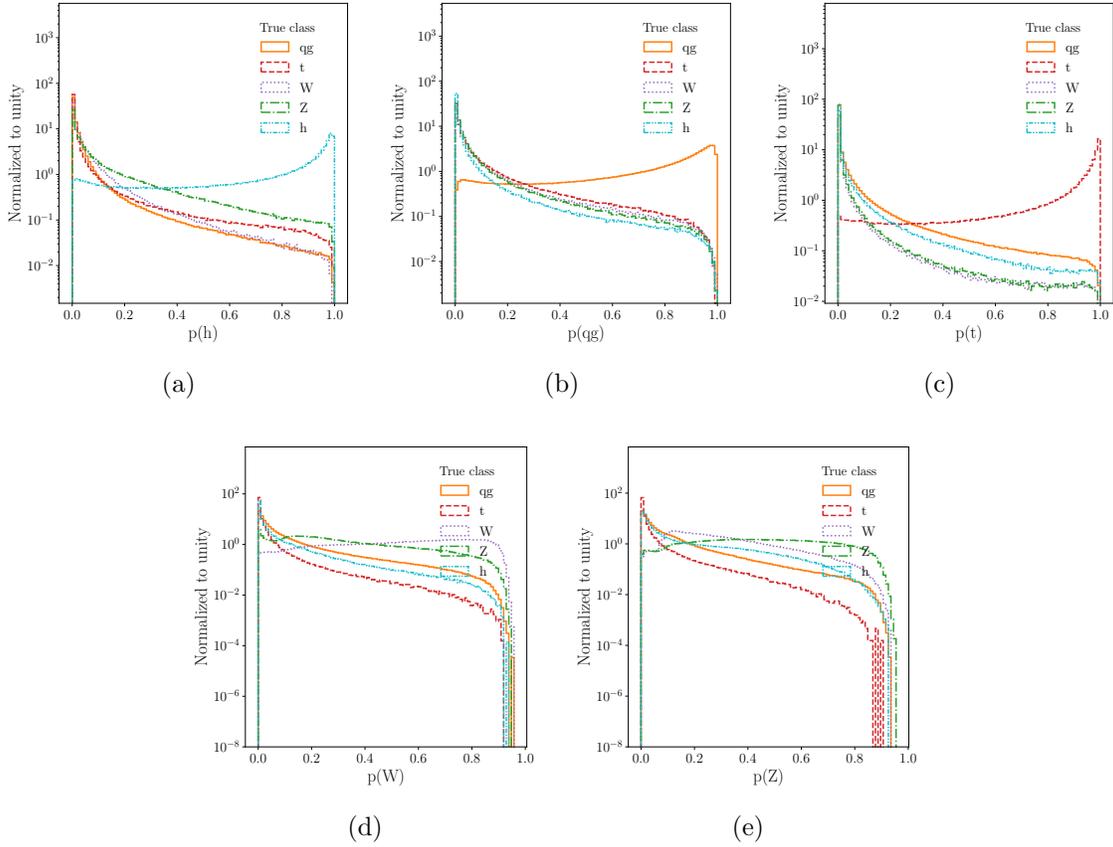


Figure 7.12: Merged MCT output probabilities tagging on true class label.

The confusion matrices for the testing dataset are shown in Fig. 7.14. The events are separated in bins of large- R jet p_T . The matrices are highly diagonal with little p_T dependence, except a small decrease in performance in the lowest p_T bin of $[200, 250]$ GeV. The W vs. Z discrimination is the task that causes the most confusion. However, for the purpose of this analysis, only the “vector boson” (V) class, taken as the maximum score between the W and Z scores, was considered, resulting in an average accuracy in the network predictions above 75% for all classes.

The ROC (Receiver Operating Characteristic) curves, representing the signal efficiency vs. background rejection, were produced for all signal-background combinations: for a Higgs signal in Fig. 7.15, for a q/g signal in Fig. 7.16, for a top-quark signal in Fig. 7.17, for a W signal in Fig. 7.18, and for a Z signal in Fig. 7.19. Each ROC was built using the output score distribution for the specified signal class of the given true signal and true background events. For each signal-background pair, the ROC is built using events in different p_T bins. Similarly

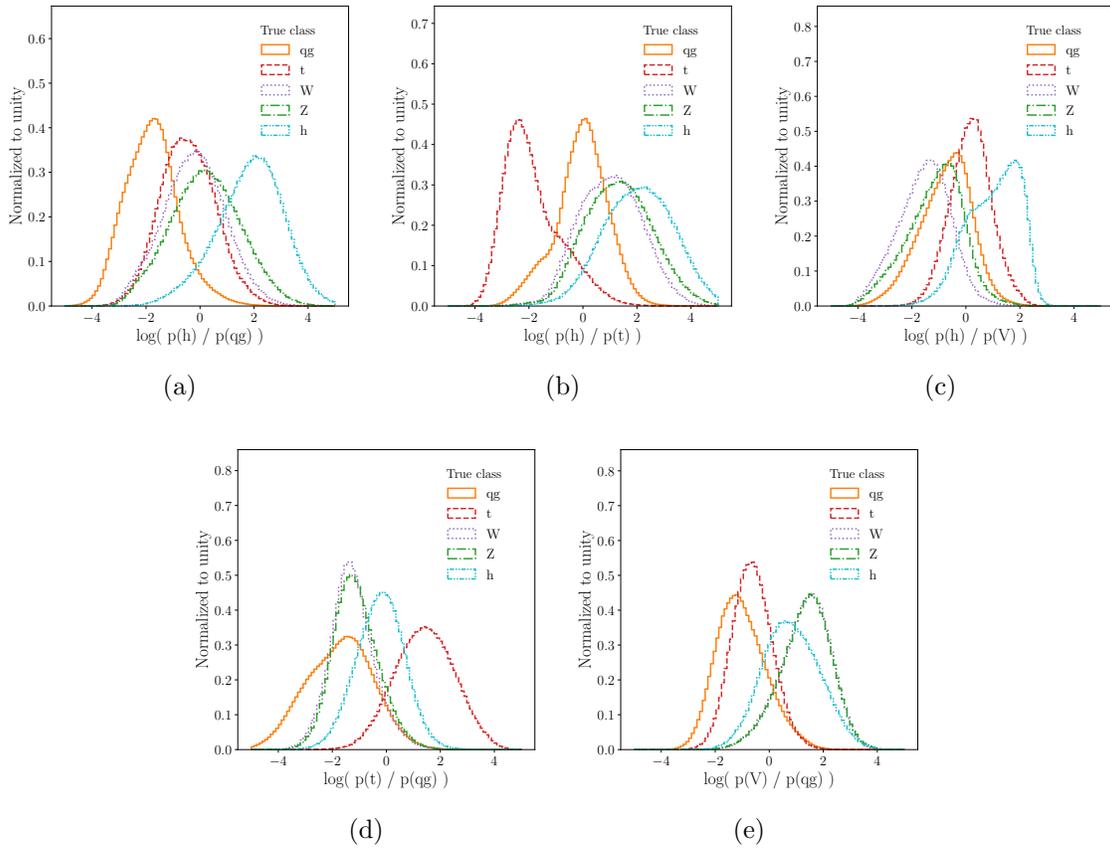


Figure 7.13: Merged MCT output log-likelihood ratios tagging on true class label.

to the confusion matrices, only a minor p_T dependence is observed, with degradation only in the lowest p_T bin. The ROC can also be built using likelihood ratios between the signal and background class. As discussed in the introduction, this results in a better discrimination power with respect to using the raw scores, as shown in Fig. 7.20.

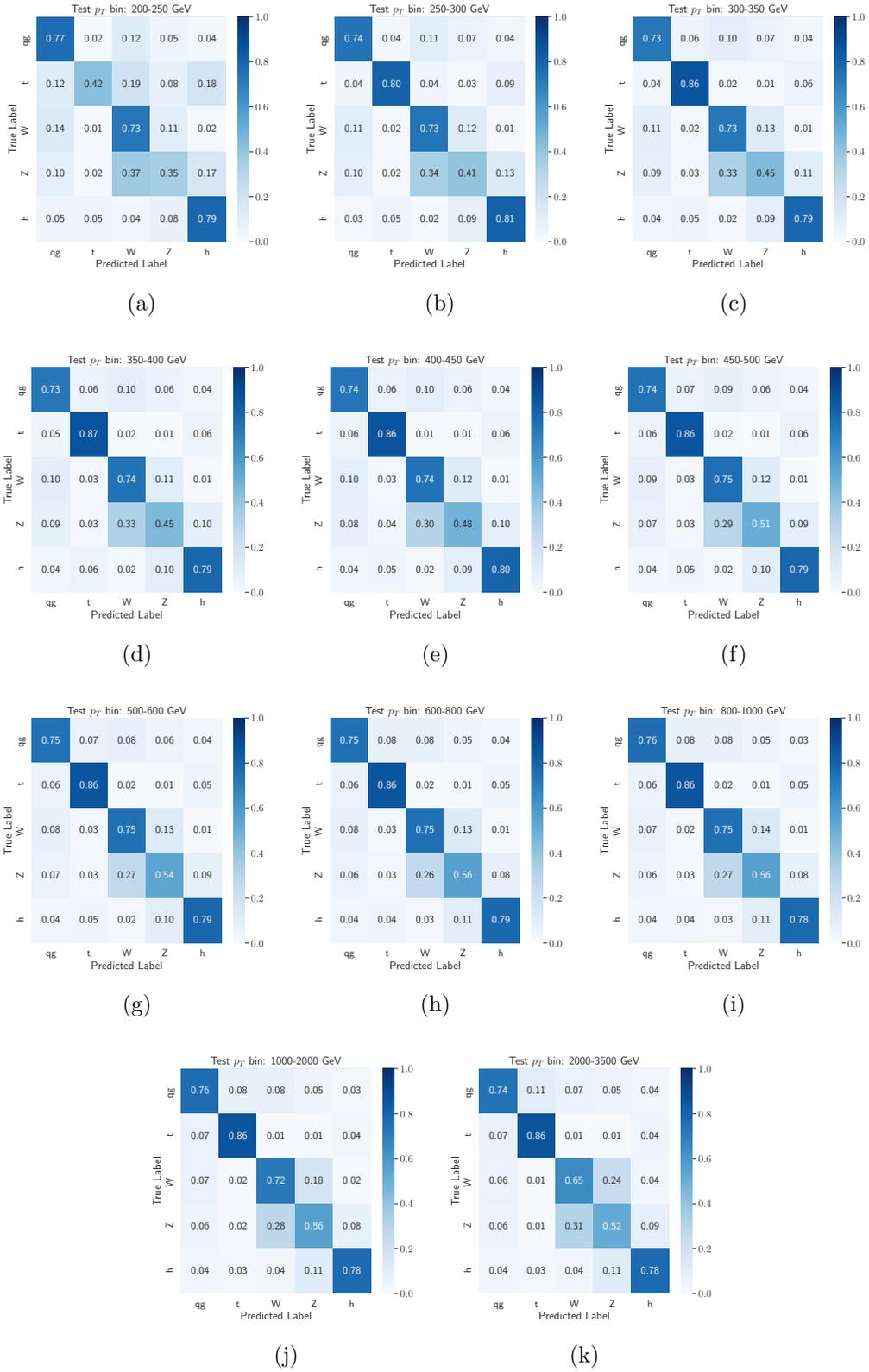


Figure 7.14: Merged MCT confusion matrices as a function of p_T of the large-R jet.

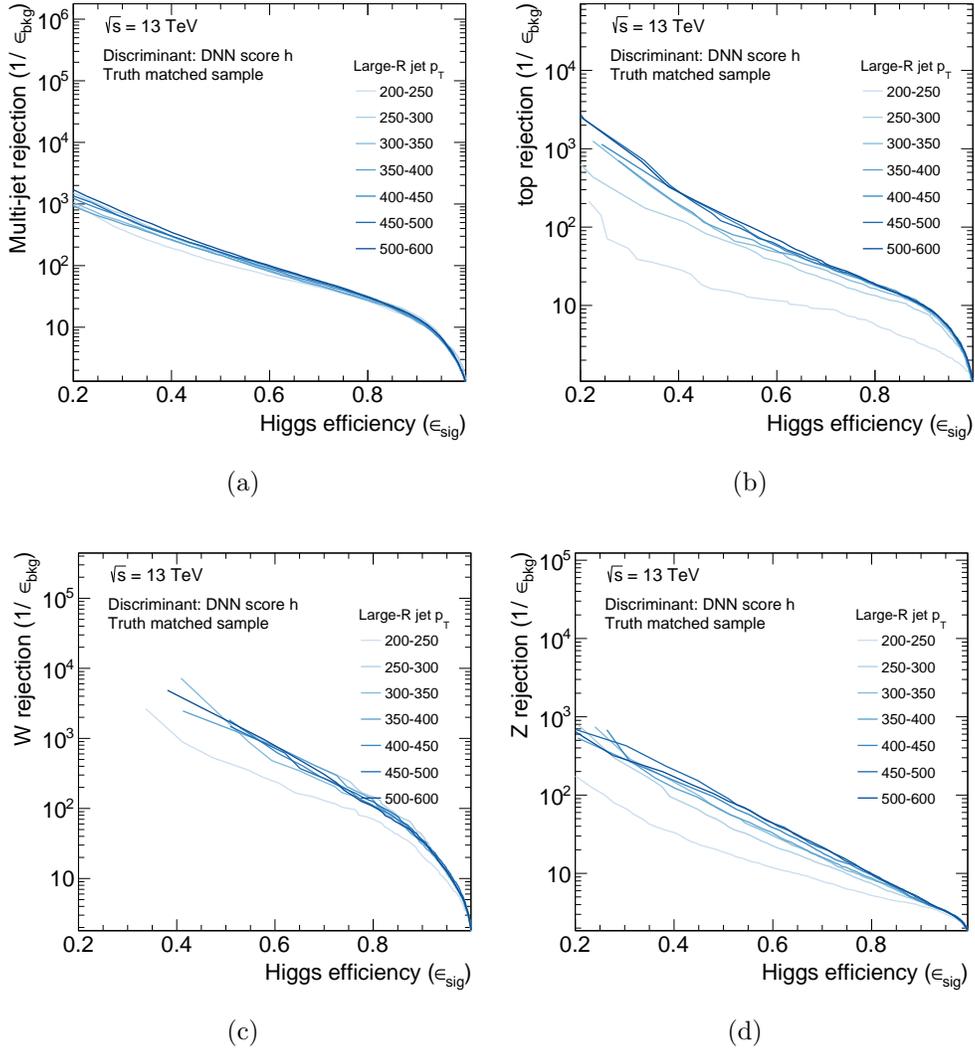
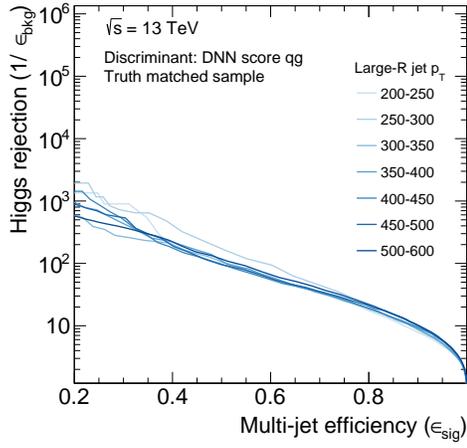
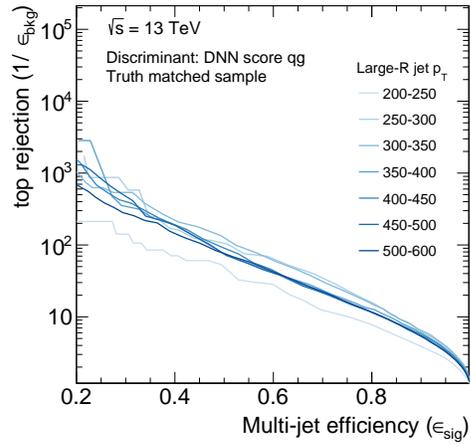


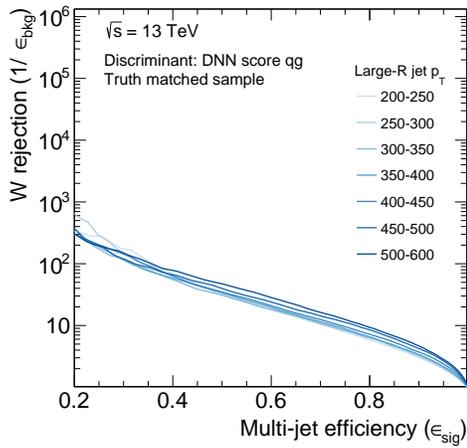
Figure 7.15: Merged MCT ROC curves built using the output score $p(h)$, taking the Higgs class as signal and each of the remaining classes as background. The ROCs are shown as a function of p_T of the large-R jet.



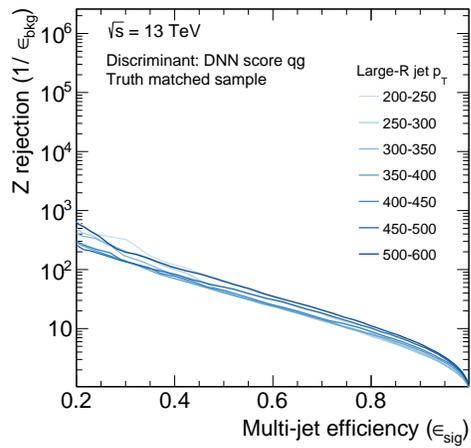
(a)



(b)

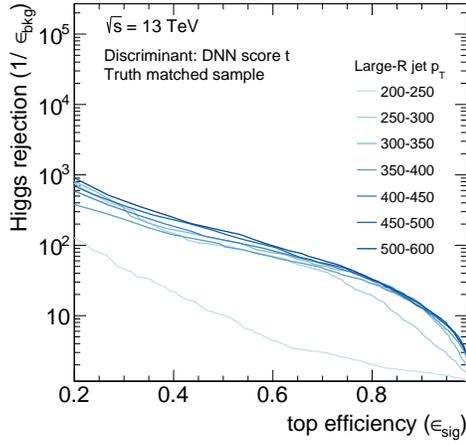


(c)

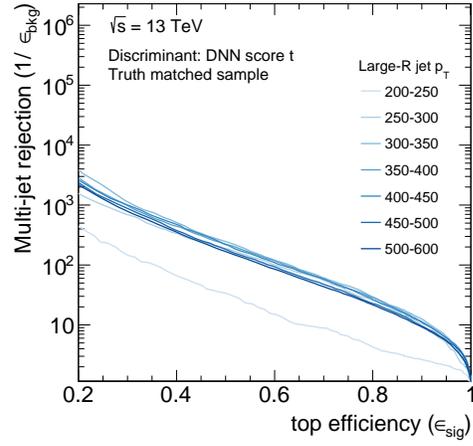


(d)

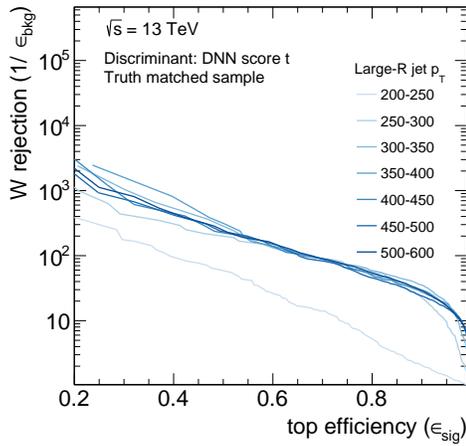
Figure 7.16: Merged MCT ROC curves built using the output score $p(q/g)$, taking the q/g class as signal and each of the remaining classes as background. The ROCs are shown as a function of p_T of the large-R jet.



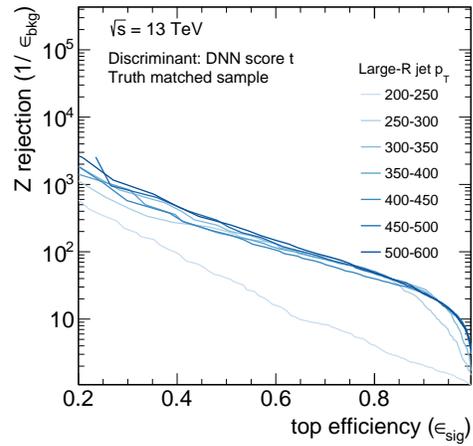
(a)



(b)

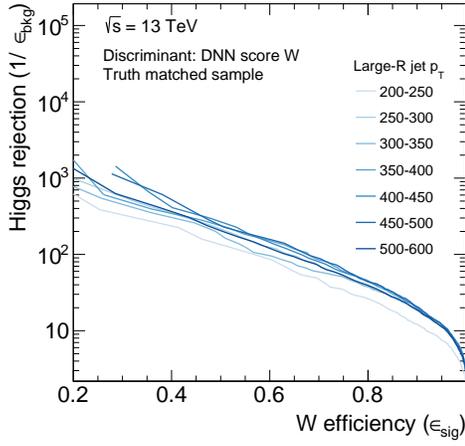


(c)

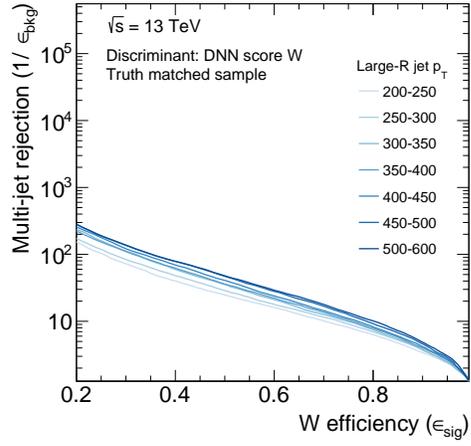


(d)

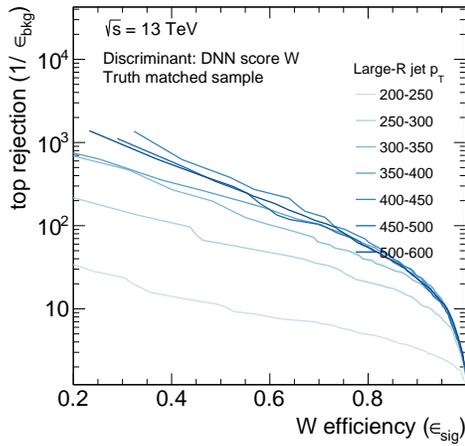
Figure 7.17: Merged MCT ROC curves built using the output score $p(t)$, taking the top class as signal and each of the remaining classes as background. The ROCs are shown as a function of p_T of the large-R jet.



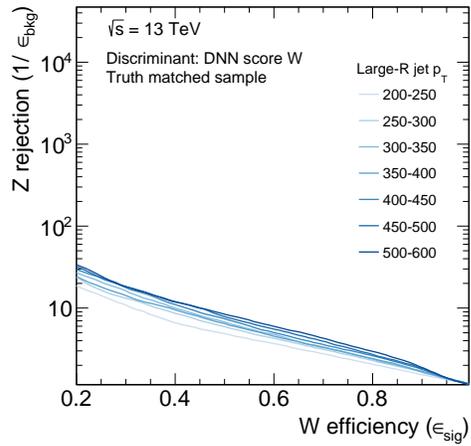
(a)



(b)

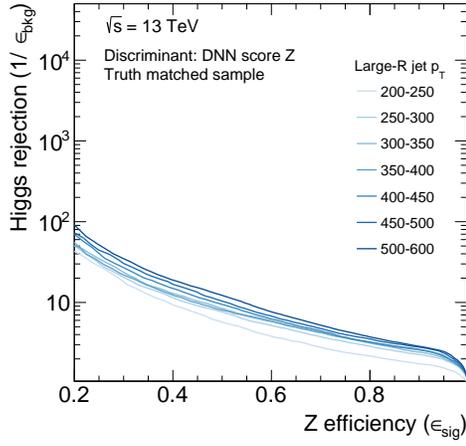


(c)

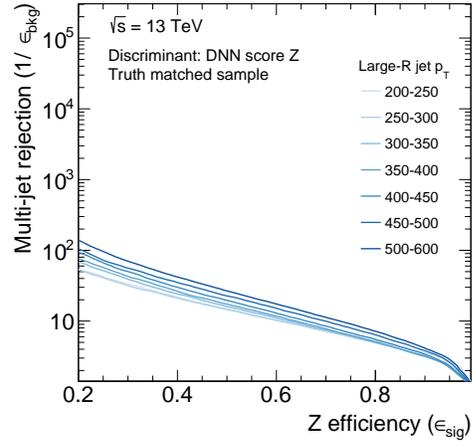


(d)

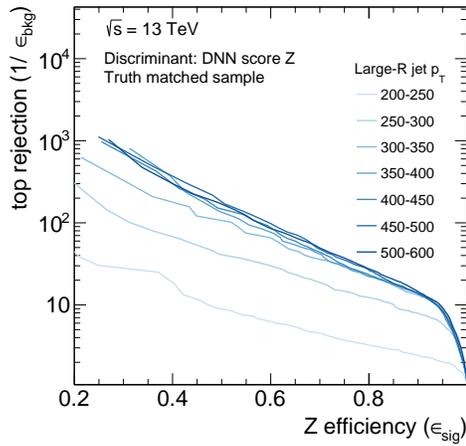
Figure 7.18: Merged MCT ROC curves built using the output score $p(W)$, taking the W class as signal and each of the remaining classes as background. The ROCs are shown as a function of p_T of the large-R jet.



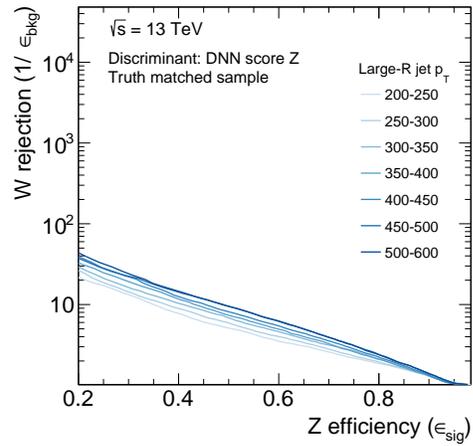
(a)



(b)

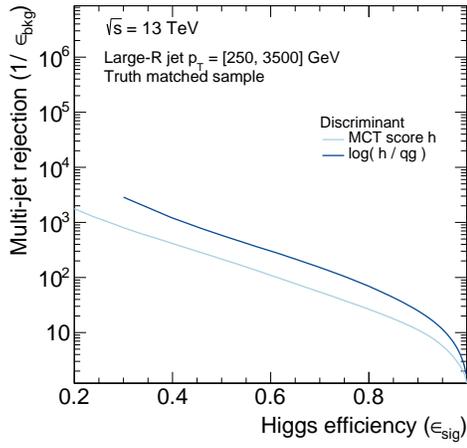


(c)

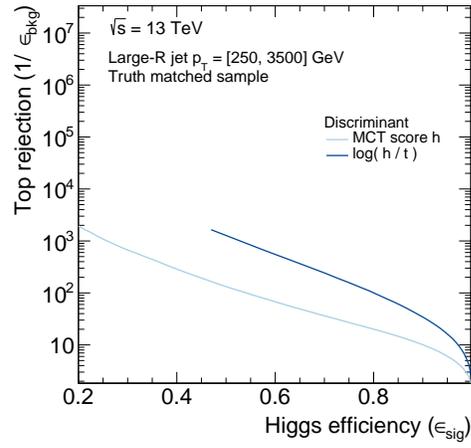


(d)

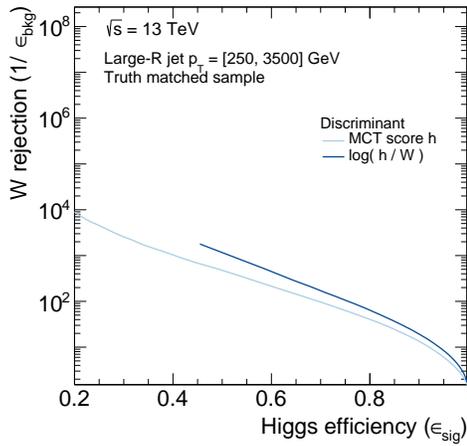
Figure 7.19: Merged MCT ROC curves built using the output score $p(Z)$, taking the Z class as signal and each of the remaining classes as background. The ROCs are shown as a function of p_T of the large-R jet.



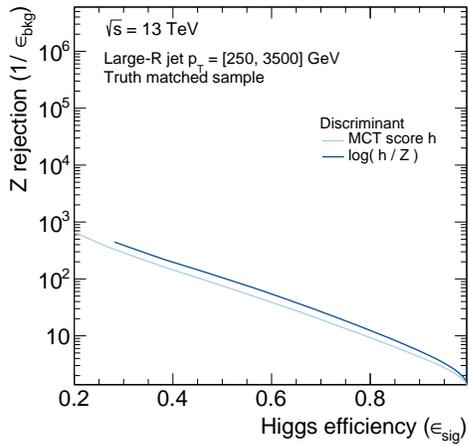
(a)



(b)



(c)



(d)

Figure 7.20: Merged MCT ROC curves for Higgs signal class. Comparison of performance when using as discriminant the output score of the signal class vs. using the log-likelihood ratio of signal and background score.

7.9 Resolved jets MCT

The resolved MCT was developed to classify the hadronic decay in the resolved regime, where the relevant objects are small-R jets. The development followed closely what was done for the merged MCT.

7.9.1 Training

Jet reconstruction and selection

The small-R jet collection is the same as the one used in the analysis: $R = 0.4$ jets built by running anti- k_t on PFlow inputs. All jets are required to be in the region $|\eta| < 2.5$ and to have $p_T > 20$ GeV, with the leading jet required to have $p_T > 45$ GeV.

MC samples and truth labeling

The resolved event topology is more dependent on the generating process. For this reason, the samples used for training were a subset of the Monte Carlo samples used in the analysis, as shown in Tab. 7.7. The signal classes were obtained from the respective signal samples: HVT-VH and ggA were used as sources of Higgs boson decays; HVT-VV, radion, and graviton as sources for the W and Z boson classes. The $t\bar{t}$ events were used as sources of top-quark decays, and the V+jets samples as sources of events with light quarks and gluons.

Class	Process
Light quarks and gluons	V+jets
Top quark	$t\bar{t}$
W boson	HVT $V' \rightarrow WW/WZ$ ($W \rightarrow qq$), Graviton-WW, Radion-WW
Z boson	HVT $V' \rightarrow WZ$ ($Z \rightarrow qq$), Graviton-ZZ, Radion-ZZ
Higgs boson	HVT $V' \rightarrow VH$, ggA

Table 7.7: MC samples used to select the training events for the Resolved MCT.

Except for the QCD sample, all training samples were truth matched by first matching the truth particle to the closest truth jet, and then matching the truth jet to the closest reconstructed small-R jet, using $dR = 0.35$. In the case of the Higgs, W, and Z signal samples, the training events were required to have both daughters of the truth boson to be truth matched to two of the three leading jets. This removed events where one of the truth quarks was outside the η acceptance region, as well as cases of “super merged” boson decays, where the truth quarks overlap. For events with top quarks, the three leading jets

were required to be truth matched to the three decay products from the top quark. This was done to remove possible noise from a partial reconstruction of the top decay, similarly to what was done for the merged MCT by accepting only contained tops.

In order to train on events similar to the ones passing the analysis pre-selection, the training samples were further required to pass the trigger selection for the given lepton channel, as well as to have at least two signal jets. The leading jet was required to have $p_T > 45$ GeV and $\eta < 2.5$, while the second leading jet was required to have $p_T > 20$ GeV and $\eta < 2.5$. In addition, if the third leading jet was found to have $p_T < 20$ GeV or $\eta > 2.5$, since this jet would not pass the analysis selection for small-R jets, the corresponding input variables were set to 0. Lastly, in order to focus on the kinematic region of interest for the resolved regime, only events where the sum of the p_T of the two leading jets was below 500 GeV was considered for training.

Input variables

In the resolved regime, true Higgs, W, and Z boson decays are reconstructed using two small-R jets, which usually are the leading jets in the event. Top decays produce instead three small-R jets. For this reason, the inputs to the DNN were chosen to describe the three leading small-R jets in the event: kinematic and b-tagging information, as well as some reconstructed variables, as shown in Tab. 7.8. In particular, the inputs include the dR between any possible pair of the three leading jets and the mass of the reconstructed object from any two pairs of jets and from all three jets. The distributions of the input variables for the signal samples were shown to be lepton-channel independent, which motivated the choice to train lepton-channel agnostically. Examples of distributions of input variables for the five true classes are shown in Figs. 7.22 and 7.21.

Pre-processing

Because of the much larger number of V+jets and $t\bar{t}$ Monte Carlo events with respect to signal events, the background classes were down-sampled to have a similar number of events as the signal classes per p_T bin. Then, similarly to the Merged MCT, the training events were re-weighted to have a per-class flat distribution of the sum of the p_T of the two leading jets. The samples were then re-weighted to have an equal class normalization.

Hyperparameters and training

The model was trained using Keras with TensorFlow [197] backend. The number of samples used for training and validation was 1.6 M and 411 K, respectively. A dataset of 686 K was holdout for testing. The architecture used was a fully connected DNN. A dropout layer

Type	Observable	Definition
3 leading small-R jets	p_T	Transverse momentum
	η	Pseudorapidity
	ϕ	Azimuthal angle
	m	Mass
	DL1r pb	Bottom quark probability
	DL1r pc	Charm quark probability
	DL1r pu	Light quark probability
Masses of reconstructed objects	$m_{J_1+J_2}$	
	$m_{J_1+J_3}$	
	$m_{J_2+J_3}$	
	$m_{J_1+J_2+J_3}$	
Angular separations	$\Delta R_{J_1,J_2}$	
	$\Delta R_{J_1,J_3}$	
	$\Delta R_{J_2,J_3}$	

Table 7.8: Resolved MCT input variables describing the three leading small-R jets.

was inserted between every hidden layer for regularization. The network was trained with the maximum number of epochs set to 1000, with early stopping implemented to stop the training when no further reduction in the loss was observed for more than 40 epochs. The hyperparameters were optimized using a grid search and the final choice is shown in Tab. 7.9. The training and validation accuracy of the model was found to be 0.72.

Batch size	1000
Learning rate	0.0001
Dropout probability	0.3
Hidden layers	3
Nodes per hidden layer	300

Table 7.9: Resolved MCT final choice of hyperparameters.

7.9.2 Testing performance

The model performance was evaluated on the testing dataset. Fig. 7.23 shows the output probabilities for all the events in the testing dataset separated by their true class label. The score of the given true class peaks at 1, while the other classes peak at 0, as desired. Similarly, Fig. 7.24 shows the log-likelihood ratios for different class combinations, where the

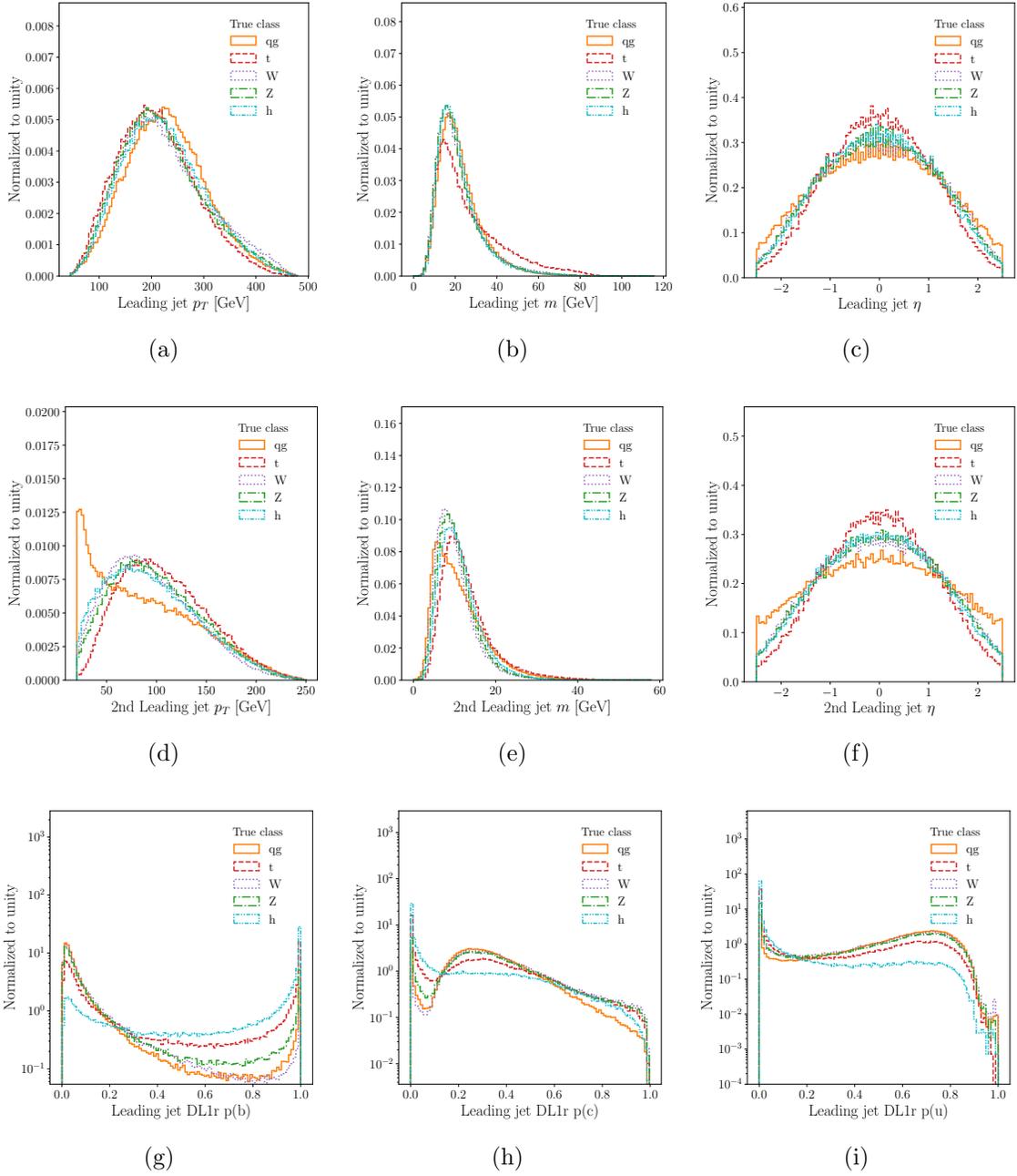


Figure 7.21: Subset of input variables for the Resolved MCT tagged according to the true class representing the kinematic variables of the leading (top) and 2nd leading (middle) jets, and the b-tagging scores of the first leading jet (bottom).

$p(V)$ score is taken as $\max(p(W), p(Z))$. The output probabilities show some uncertainty in the W and Z predictions. However, this is mostly resolved when looking at the likelihood ratios, when uncertainty about the other classes is taken into account. A confusion remains for a subset of events between the Higgs and V classification, which will be explained in the

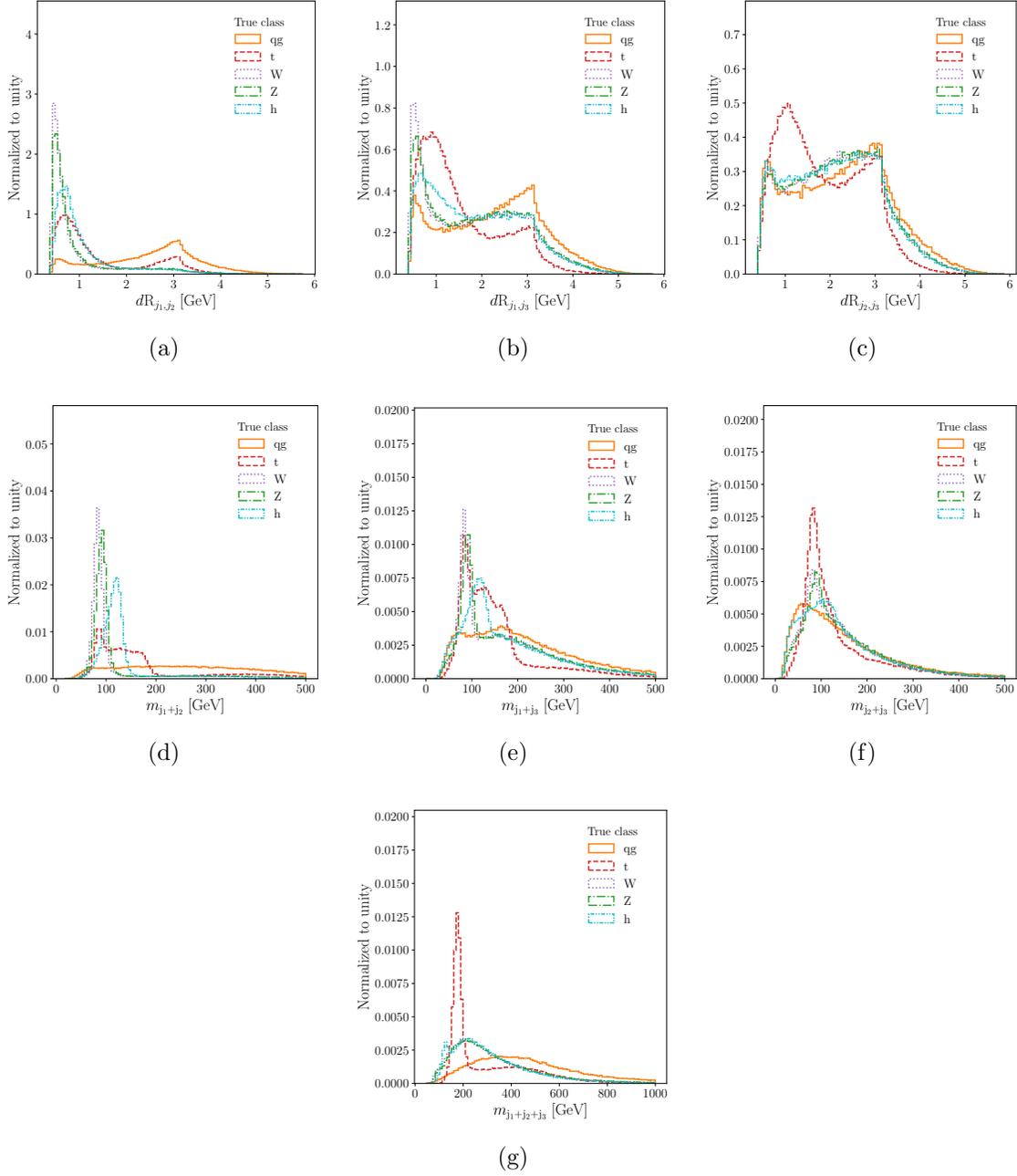


Figure 7.22: Subset of input variables for the Resolved MCT tagged according to the true class representing the mass and dR distances of reconstructed objects.

following.

The confusion matrices are shown in Fig. 7.25. The events are separated in bins of the sum of the p_T of the two leading small- R jets. The matrices are highly diagonal with little p_T dependence, except a small decrease in performance in the two lowest p_T bins. The Z class is the one that suffers the most from the confusion with both the W and the Higgs

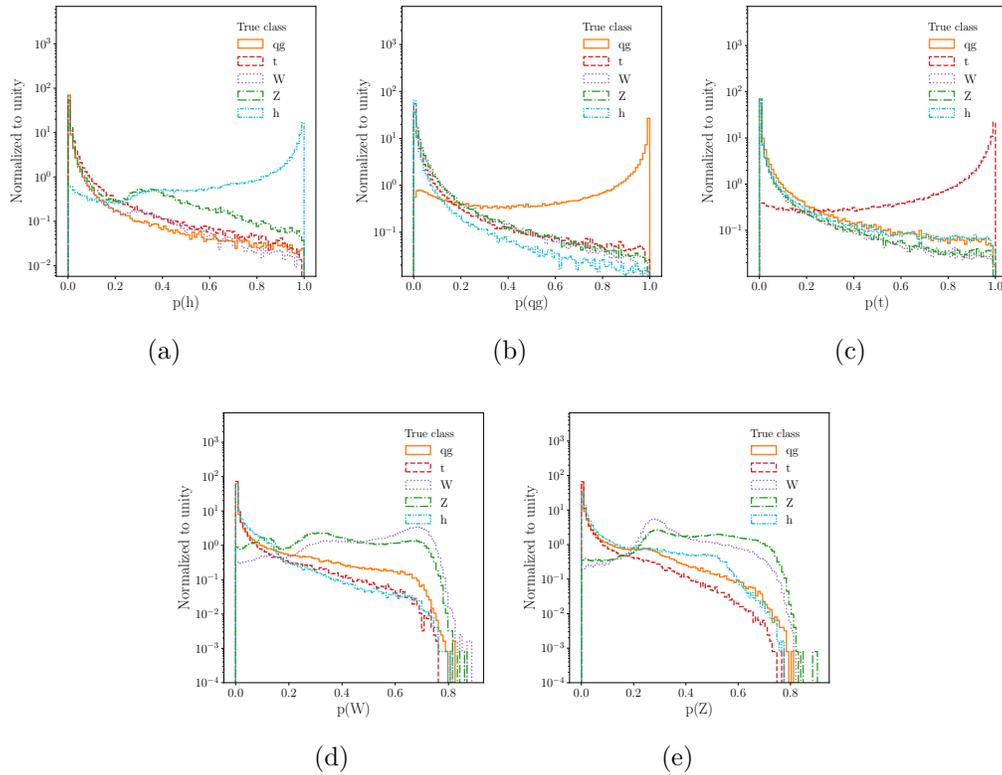


Figure 7.23: Resolved MCT output probabilities tagging on true class label.

class. However, part of this confusion is resolved by considering only the “vector boson” (V) class, taken as the maximum score between the W and Z scores. The resulting average accuracy in the network predictions is above 75% for all classes.

The confusion matrix for true Higgs, W, and Z events was further investigated by separating the events by true sample of origin, by p_T bin and by number of b-tagged three leading small-R jets, as shown in Fig. 7.26. A dependence on the number of b-tags is observed and is expected, as a W boson should in principle only populate the 0 b-tag region, a Z boson hadronically decays to two b-quarks 15% of the times, and a Higgs boson hadronically decays to two b-quarks 95% of the times.

The ROC curves, representing the signal efficiency vs. background rejection, were produced for all signal-background combinations: for a Higgs signal in Fig. 7.27, for a q/g signal in Fig. 7.28, for a top-quark signal in Fig. 7.29, for a W signal in Fig. 7.30, and for a Z signal in Fig. 7.31. Each ROC was built using the output score distribution for the specified signal class of the given true signal and true background events. For each signal-background pair, the ROC is built using events in different p_T bins. The greatest p_T dependence is observed for the q/g class and the greatest degradation on performance occurs for the lowest p_T bin.

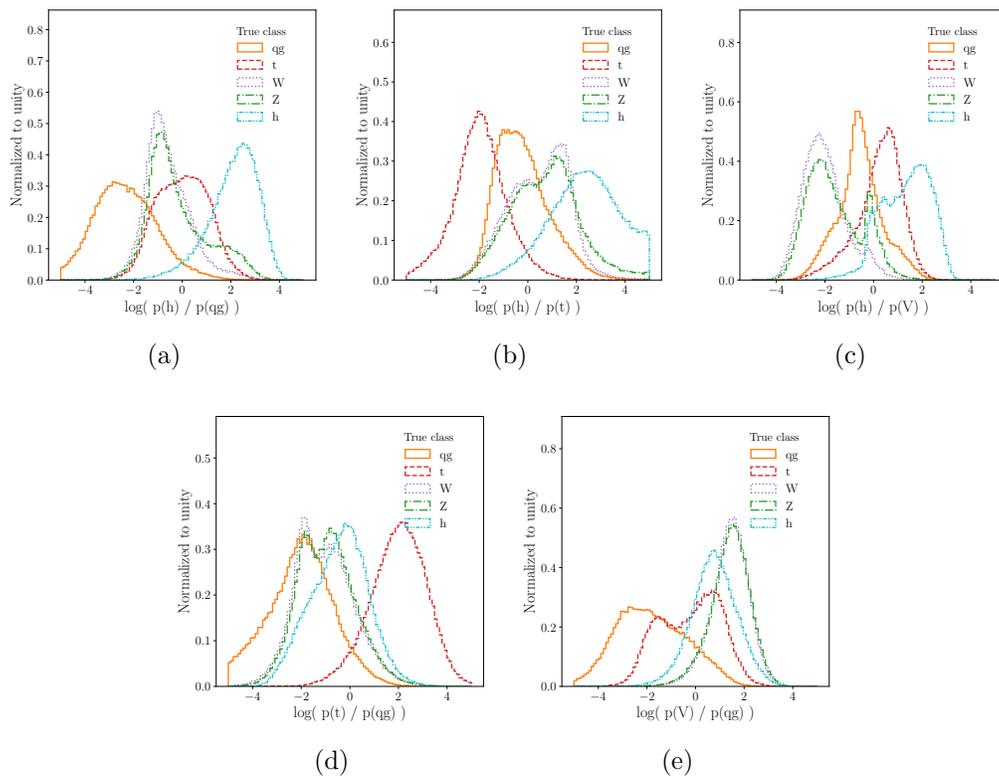


Figure 7.24: Resolved MCT output log-likelihood ratios tagging on true class label.

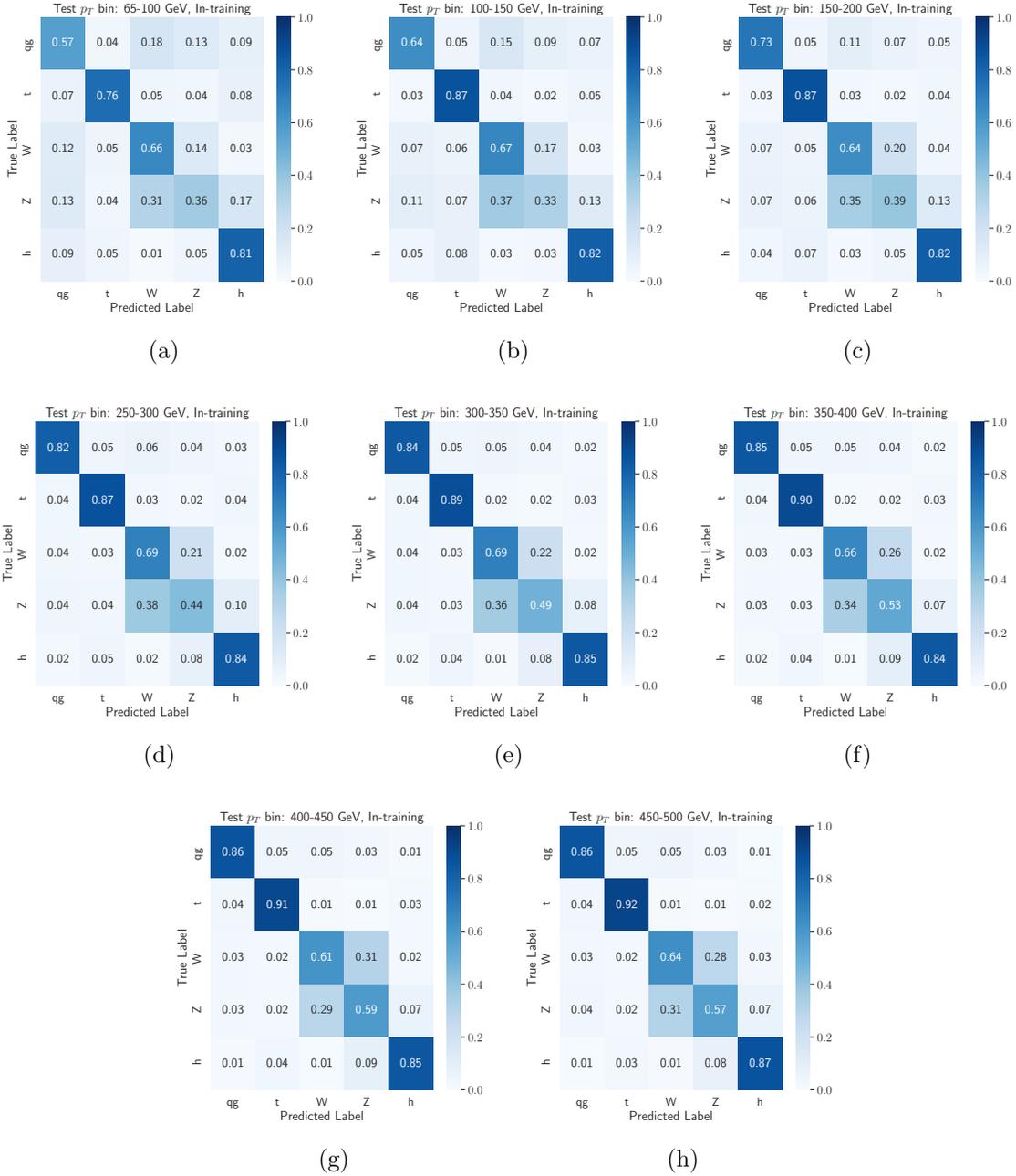


Figure 7.25: Resolved MCT confusion matrices as a function of the sum of the p_T of the two leading small-R jets.

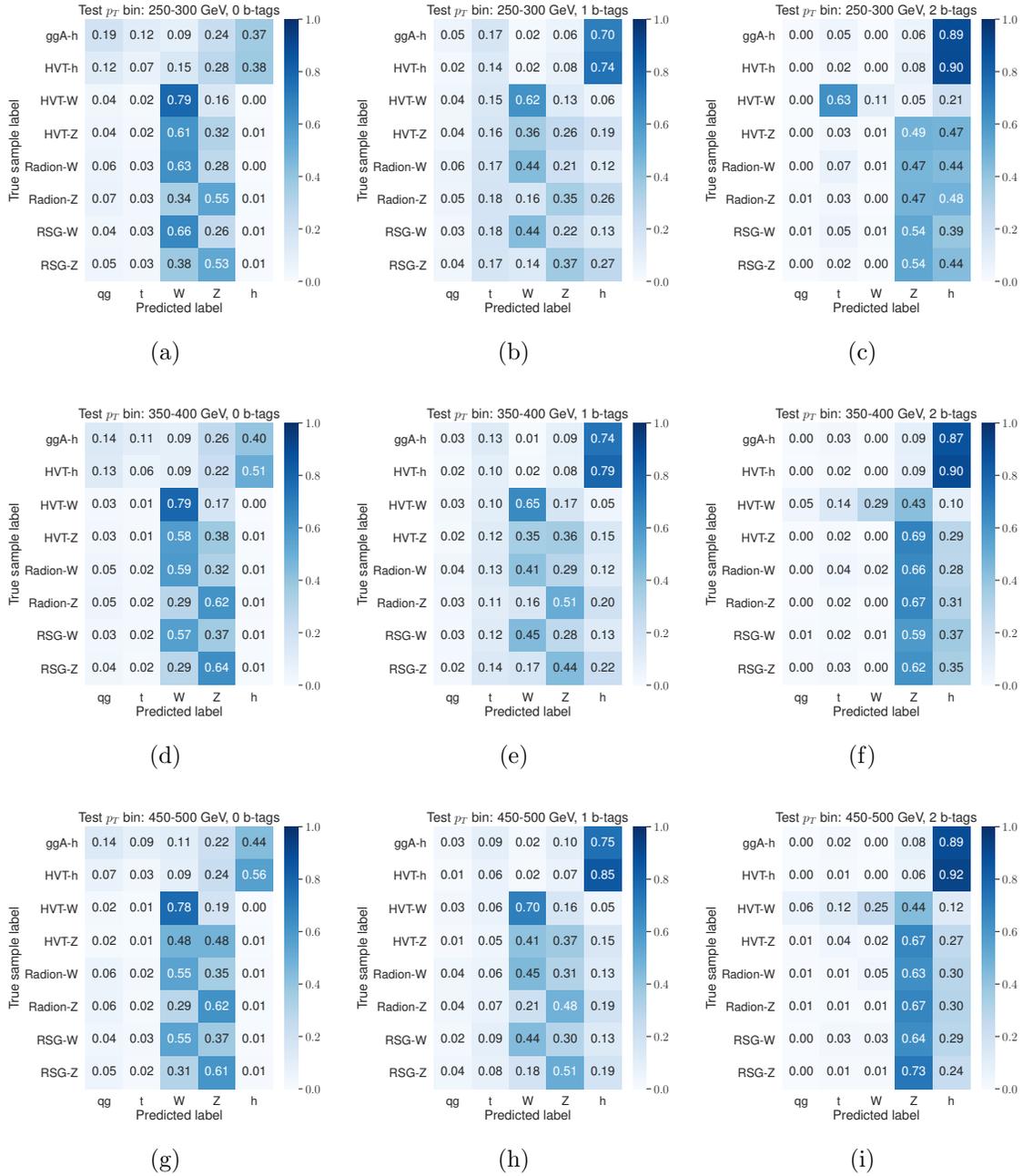


Figure 7.26: Resolved MCT confusion matrices separated by true sample of origin and number of b -tags among the three leading small- R jets, in different bins of the sum of the p_T of the two leading small- R jets.

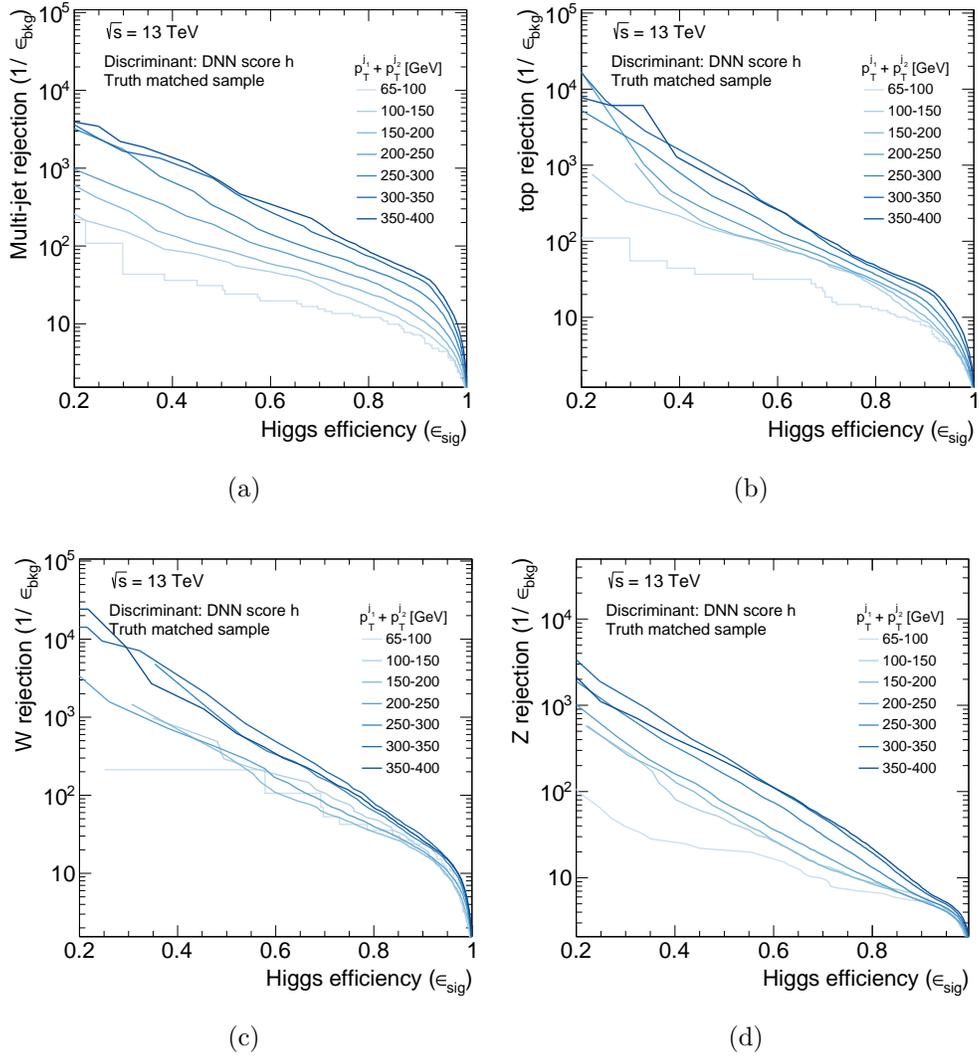
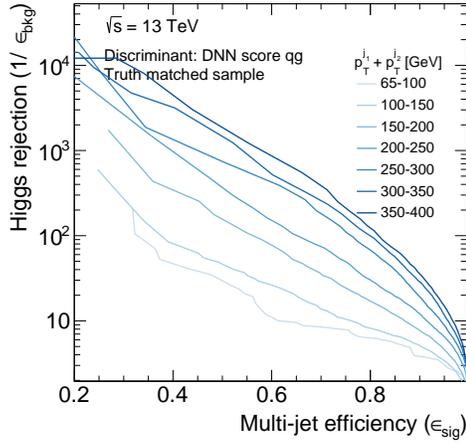
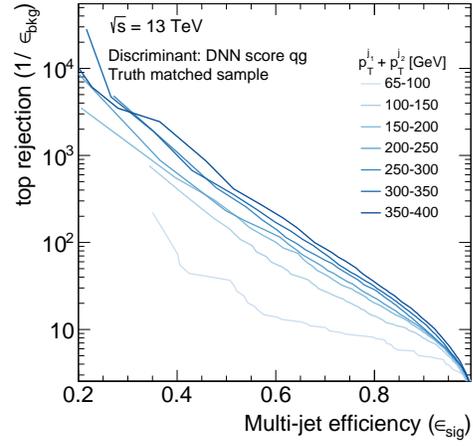


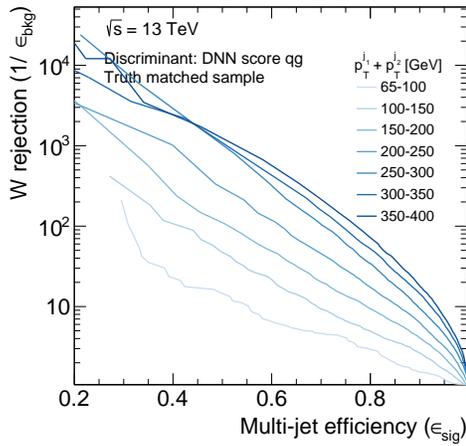
Figure 7.27: Resolved MCT ROC curves built using the output score $p(h)$, taking the h class as signal and each of the remaining classes as background. The ROCs are shown as a function of the sum of the p_T of the two leading small-R jets.



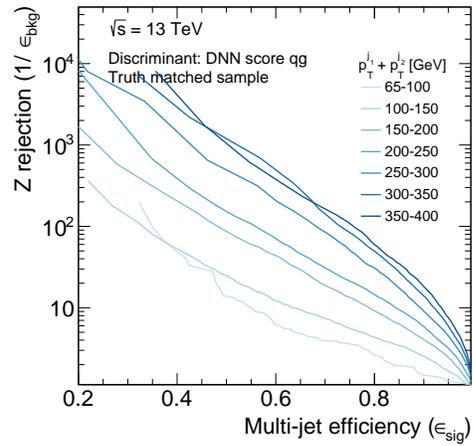
(a)



(b)

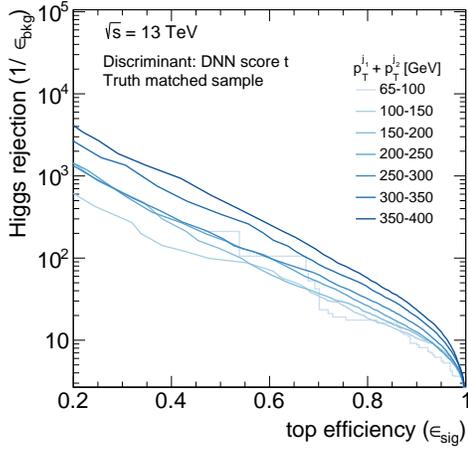


(c)

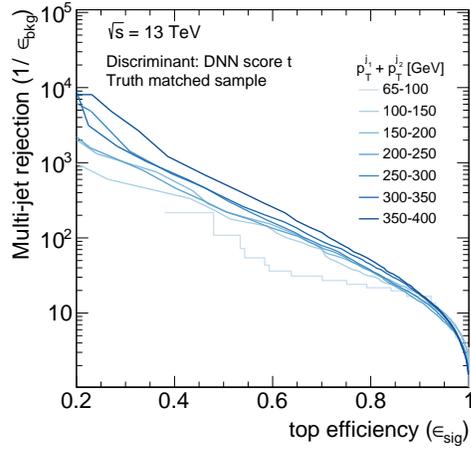


(d)

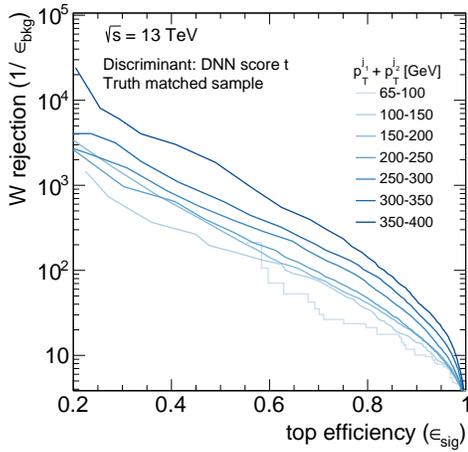
Figure 7.28: Resolved MCT ROC curves built using the output score $p(q/g)$, taking the q/g class as signal and each of the remaining classes as background. The ROCs are shown as a function of the sum of the p_T of the two leading small-R jets.



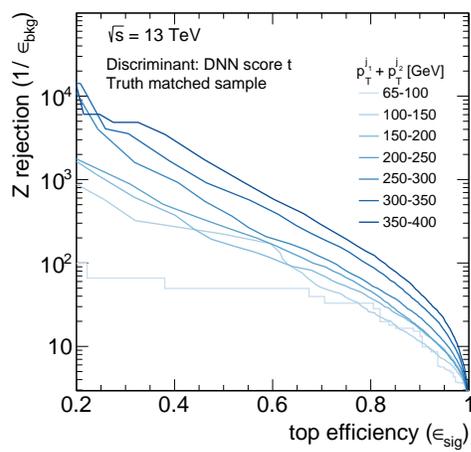
(a)



(b)

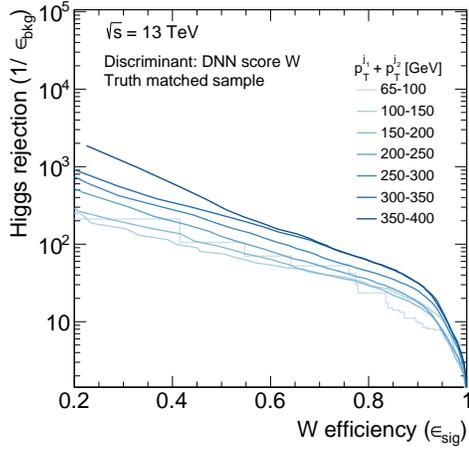


(c)

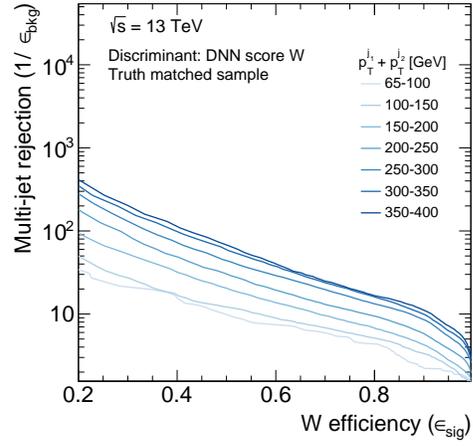


(d)

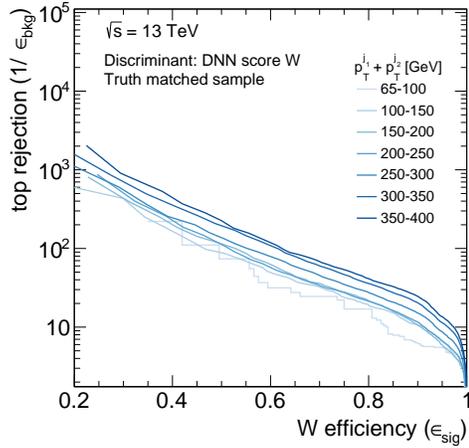
Figure 7.29: Resolved MCT ROC curves built using the output score $p(t)$, taking the top class as signal and each of the remaining classes as background. The ROCs are shown as a function of the sum of the p_T of the two leading small-R jets.



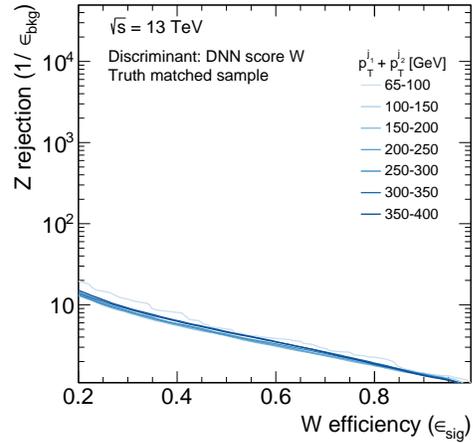
(a)



(b)

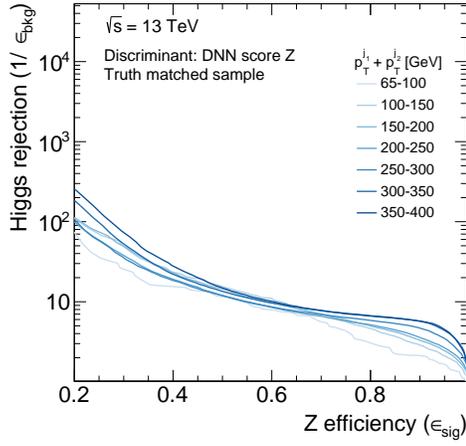


(c)

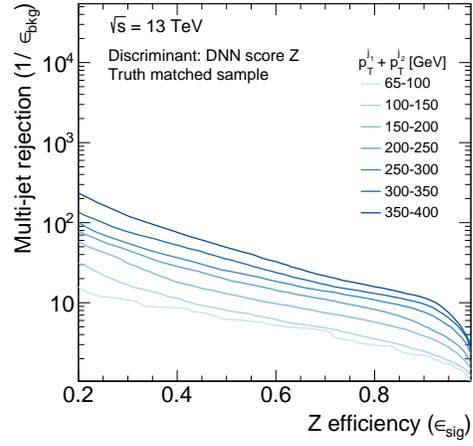


(d)

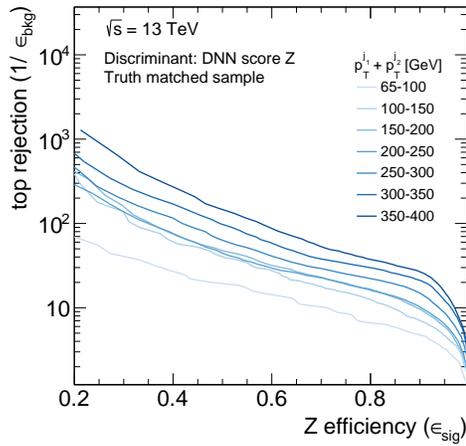
Figure 7.30: Resolved MCT ROC curves built using the output score $p(W)$, taking the W class as signal and each of the remaining classes as background. The ROCs are shown as a function of the sum of the p_T of the two leading small-R jets.



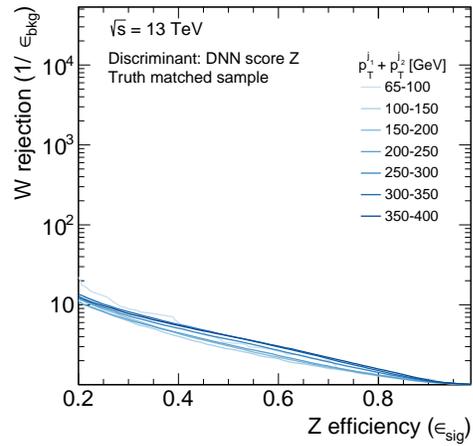
(a)



(b)



(c)



(d)

Figure 7.31: Resolved MCT ROC curves built using the output score $p(Z)$, taking the Z class as signal and each of the remaining classes as background. The ROCs are shown as a function of the sum of the p_T of the two leading small-R jets.

7.10 MCT deployment in the analysis

The VV and VH final regions have to be made orthogonal in order to perform a combined statistical fit for the HVT interpretation. The events are sorted in orthogonal regions using the two Multi-Class Tagger neural networks described in the previous sections. This strategy differs from what was done in previous combination efforts and ultimately results in a higher search sensitivity. This section discusses the motivation, development, and results of this new deep-learning-based strategy.

7.10.1 Motivation

The search for a new spin-1 HVT boson (V') is performed in both the VV and VH channels, as a new W' can decay both to Wh and WZ , and a new Z' can decay to Zh and WW . In order to exploit the complementarity of the different searches, analyses assuming the same underlying model can be combined to provide more stringent limits on the model parameters and increase the statistical power of the search. However, before performing the statistical combination, one has to ensure orthogonality of the signal regions going into the fit.

The main categories used to define the final regions are given by the lepton channel and by the reconstruction strategy of the hadronic decay. While in a given analysis different lepton channels are orthogonal by construction, the resolved and merged categories are not. As explained in Sec. 7.7, the resolved and merged reconstruction strategies are used to maximize the efficiency in low- and high-boost scenarios, where the hadronic decay is better reconstructed as two resolved small-R jets (jj) or as a single large-R jet (J), respectively. However, for a subset of events, both reconstruction strategies provide equivalent efficiencies and, within a given analysis, it is possible for an event to end up in both resolved and merged final regions. In these cases, a choice has to be made on which final region the event should go into, a decision called *prioritization*. As already discussed, these are the `PriorityResolved` and `PriorityMerged` strategies for the VH and VV analysis, respectively. After prioritization is enforced, the final regions within a given analysis are fully orthogonal.

In a given lepton channel, the VV and VH SRs are not orthogonal a priori, because the jet mass windows overlap. The definitions of the mass windows were discussed in Sec. 7.6.1 and are summarized in Tab. 7.10. Recall that the VV analysis uses the p_T -dependent WZTagger mass cut, where the upper Z (W) mass-cut is approximately in the range [94, 115] GeV ([106, 130] GeV). This overlap can be understood schematically as shown in Fig. 7.32. The x -axis shows the resolved mass window selections on $m(jj)$, while the y -axis shows the selections on the large-R jet mass $m(J)$. Note that this is approximate, as the jj di-jet

system is not necessarily identical between the W/Z candidate and the H candidate. All the events in the shaded regions (A , B , and C) can potentially enter both VV and VH SRs. The grey shaded regions (C) correspond to regions where the overlap is in the same kinematic region for VV and VH, while the red-saded regions (A and B) represent regions of possible mixed-overlap, where an event enters either both VH-SR-Res and VV-SR-Merg, or both VH-SR-Merg and VV-SR-Res.

Analysis	Channel	Resolved	Merged
VH	0-lep	$75 \leq m(jj) \leq 145$	$75 \leq m(J) \leq 145$
	1-lep		
	2-lep		
VV	W	$68 \leq m(jj) \leq 98$	Pass_WZTagger_WMassCut
	Z	$78 \leq m(jj) \leq 106$	Pass_WZTagger_ZMassCut

Table 7.10: Mass window definitions in VV and VH signal regions. All numbers are in units of GeV. In the VH analysis the cuts are assigned per lepton channel, while in VV they are given according to whether the final region is looking for a hadronically decaying W or Z boson. The VV $m(J)$ window is defined using a p_T -dependent cut provided by the WZTagger and shown in Fig. 7.6.

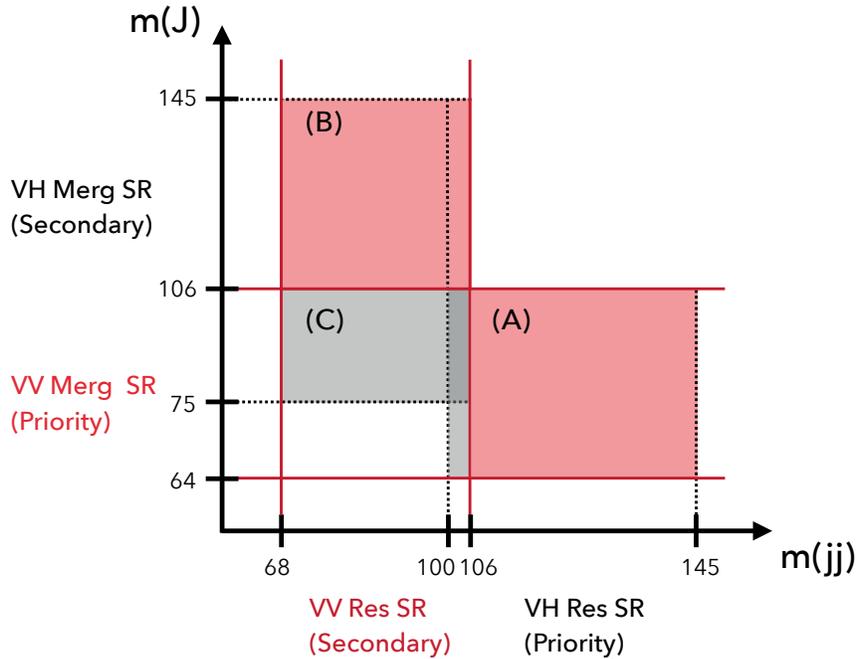


Figure 7.32: Schematic visualization of VV and VH resolved (x -axis) and merged (y -axis) mass windows overlap. See text for explanation.

The VV and VH semi-leptonic analyses were part of previous combination efforts [175, 176] interpreted in the context of the HVT framework. In these publications, the analyses

were combined after having been optimized as standalone searches. Therefore, the orthogonality condition between the SRs had to be imposed a posteriori. The best compromise was found to be rejecting the VV mass window in VH. This meant retaining 100% of the events in the VV SRs, at the expense of losing some events in the VH SRs. The idea being that most of the events lost in VH would end up in the VV SRs and therefore would still contribute to the combination. While this was found to be the best strategy, it is not fully efficient and resulted in a loss of sensitivity in VH at high mass, as will be shown in the following studies. Moreover, because the orthogonality cuts are applied a posteriori, some events can still be lost: an event that does not enter the VV SRs, but is within the VV mass window, will be removed from the VH analysis. Additionally, events that end up in *mixed* regions, do not get sorted exclusively.

Considering only events in the overlap region and referring to Fig. 7.32, the effect of the mass cut used in the previous combination is the following:

- Events in the shaded gray area (C) always migrate from VH to VV.
- The events in the shaded red regions (A, B) remain shared between VH and VV.

The latter case was tolerated because deemed negligible. However, it is still not desirable, particularly for region A where events are in the priority regions of both analyses. The number of events in the inclusive overlap region and mixed-overlap regions only, for the 36.1fb^{-1} mc16a dataset is shown in Tab. 7.11 for the 2-lepton channel.

	data	HVT-WZ	HVT-ZH	ttbar	Wjets	Zjets
Total	8546693	326257	287777	24496926	764373	39371912
Any SR	–	145928	102940	57031	64	1376956
VH SR's	–	7856	97235	54612	26	814163
VV SR's	–	145057	12892	2639	40	581859
Overlap VH & VV SR	412	6985	7187	220	2	19066
% of total	0.01	2.14	2.50	0.00	0.00	0.05
% of any SR	–	4.79	7.00	0.39	3.12	1.38
Res-VH & Merg-VV	74	150	1057	50	0	3396
% of overlap	18.00	0.91	14.71	22.73	0.00	17.81
Merg-VH & Res-VV	29	64	414	1	0	1689
% of overlap	7.04	2.15	5.76	0.45	0.00	8.86

Table 7.11: Number of events in the 2-lepton channel for data, HVT signals, and the most important background processes. The data counts in the signal regions are not shown because the analysis was still blinded. The number of events that enter the VV and VH signal regions can be compared to the event counts in the inclusive overlap region, as well as in the mixed-overlap regions.

In this publication for the first time the VV and VH processes were considered in the same analysis, making it possible to implement a recycling strategy so that no event is lost and the mixed-regions can be taken care of properly. Moreover, the harmonization of the two analyses into a single effort opened up the possibility for a more efficient event categorization into the respective VV and VH signal regions. In this analysis a new orthogonalization strategy that makes use of the output scores of the MCTs was proposed. In fact, it was shown that the loss in sensitivity from the mass cut strategy was mostly due to events in the gray shaded regions, so that most of the recovering of the sensitivity will be due to the MCT.

7.10.2 Studies overview

The studies presented here use the 2016 dataset with 36.1 fb^{-1} (referred to as `mc16a`) to reduce the processing time, but it can be considered representative of the full Run 2 dataset. For these optimization studies only the HVT signal samples with DY production was used, as shown in Tab. 7.12, with the corresponding signal regions, shown in Tab. 7.13. Note that the 0-lepton channel of the VV analysis does not have a resolved signal region.

	0-lep	1-lep	2-lep
VH	HVT $Z' \rightarrow ZH$	HVT $W' \rightarrow WH$	HVT $Z' \rightarrow ZH$
VV	HVT $W' \rightarrow WZ$	HVT $W' \rightarrow WZ$	HVT $W' \rightarrow WZ$

Table 7.12: Signal samples used in the orthogonality studies.

	VH Res	VH Merg	VV Res	VV Merg
0 Lepton	Res SR 1b Res SR 2b	Merg SR 1b0add Merg SR 2b0add	-	Merg HP GGF WZ SR Merg LP GGF WZ SR
1 Lepton	Res SR 1b Res SR 2b	Merg SR 1b0add Merg SR 2b0add	Res GGF WZ SR 01b Res GGF WZ SR 2b	Merg HP GGF WZ SR 01b Merg HP GGF WZ SR 2b Merg LP GGF WZ SR 01b Merg LP GGF WZ SR 2b
2 Lepton	Res SR 1b Res SR 2b	Merg SR 1b0add Merg SR 2b0add	Res GGF WZ SR	Merg HP GGF WZ SR Merg LP GGF WZ SR

Table 7.13: VV and VH HVT signal regions (SRs) used in the orthogonality studies. An event is in a VV or VH SR, if it enters any SR in the corresponding column.

The following region definitions will also be used throughout the studies:

- *VH-SR*: an event that enters any HVT-VH SR (columns “VH Res” and “VH Merg” in Tab. 7.13).
- *VV-SR*: an event that enters any HVT-VV SR (columns “VV Res” and “VV Merg” in Tab. 7.13).

- *Overlap region*: an event that enters both VV-SR and VH-SR.
- *Mixed-Overlap region*: an event that enters the overlap region with VH-SR-Res and VV-SR-Merg, or VH-SR-Merg and VV-SR-Res.

7.10.3 MCT strategy

The new proposed orthogonalization strategy, which will be referred to as *MCT strategy*, uses the $p(h)$ and $p(V)$ scores of the resolved and merged MCTs to categorize the events into the VV and VH signal regions. This can be done in two ways: 1) using the MCT scores directly, before any prioritization is enforced; 2) first applying each analysis' own resolved versus merged prioritization strategy, and then use the MCT to choose the final region. Both options were studied. In particular, for the latter case all possible combinations of prioritization strategies were reconsidered. The best option was found to be retaining the current prioritization strategy and then apply the MCT selection. In practice, the procedure to orthogonalize is the following:

- Run analysis event selection to find active signal regions
- Apply prioritization strategy (`PriorityResolved` for VH and `PriorityMerged` for VV)
- Set MCT $p(h)$ score to 0 if event is not in *VH-SR*, according to Tab. 7.13
- Set MCT $p(V)$ score to 0 if event is not in *VV-SR*, according to Tab. 7.13
- Take $p_{\max} = \max(p(h), p(V))$.
 - If $p_{\max} = p(h)$, turn the VV signal region off.
 - If $p_{\max} = p(V)$, turn the VH signal region off.

The orthogonalization strategy is optimized with the goal of deviating as little as possible from the sensitivity of the baseline analyses, where no orthogonality is yet imposed. The following strategies will be compared:

- *Baseline*: no orthogonality.
- *MCT*: use MCT strategy.
- *MassCut*: reject the VV mass window in VH.

Note that the MassCut strategy is applied only to events that are in the overlap region. This removes possible inefficiencies due to a loss of events from applying the cut a posteriori. The effect of the MCT and MassCut strategy is compared in Tab. 7.14, which shows the percentage of events in the overlap region that are assigned to the VH and VV analysis using the MCT or the Mass sorting strategy. Note that, with the MCT strategy the sorting is always an OR between the VV and VH final regions, so the percentages always add up to 100%. On the other hand, the mass cut strategy cannot orthogonalize cases of *Mixed-Overlap*. Since an event that enters a VV SR is always sorted into VV by construction, the VH percentage gives the number of events that are still shared.

In the following sections the two strategies will be compared in terms of signal efficiency as a function of jet transverse momentum 7.10.4, signal significance as a function of heavy resonance mass 7.10.5, and expected limit sensitivity 7.10.6.

0 Lepton channel

		data	HVT-WZ	HVT-WW	HVT-ZH	HVT-WH	ttbar	Wjets	Zjets
MCT	VH(%)	65.52	21.55	–	76.95	–	55.01	75.35	80.34
	VV(%)	34.48	78.45	–	23.05	–	44.99	24.65	19.66
Mass	VH(%)	10.87	1.30	–	4.28	–	18.91	7.07	5.73
	VV(%)	100.	100.	–	100.	–	100.	100.	100.

1 Lepton channel

		data	HVT-WZ	HVT-WW	HVT-ZH	HVT-WH	ttbar	Wjets	Zjets
MCT	VH(%)	83.69	57.49	22.00	–	89.17	84.55	83.13	86.02
	VV(%)	16.31	42.51	78.00	–	10.83	15.45	16.87	13.98
Mass	VH(%)	19.04	1.68	1.78	–	10.44	18.58	13.47	14.85
	VV(%)	100.	100.	100.	–	100.	–	100.	100.

2 Lepton channel

		data	HVT-WZ	HVT-WW	HVT-ZH	HVT-WH	ttbar	Wjets	Zjets
MCT	VH(%)	67.23	21.89	–	82.43	–	74.55	100.	76.30
	VV(%)	32.77	78.11	–	17.57	–	25.45	0.00	23.70
Mass	VH(%)	21.12	2.49	–	17.87	–	17.27	0.00	20.72
	VV(%)	100.	100.	–	100.	–	100.	100.	100.

Table 7.14: Percentage of events in overlap region that end up in VH or VV analysis using the MCT or Mass sorting strategy.

7.10.4 Signal efficiency

The effect of the orthogonality cut on the signal efficiency is compared for the MCT and MassCut strategies. This is shown for the 2-lepton channel in Figs. 7.33 and 7.34, for

the merged and resolved regions respectively. The 0- and 1-lepton channels showed similar behaviors.

The events are grouped into one of four categories – resolved or merged, and VV or VH – depending on which signal region they enter, according to Tab. 7.13. The signal efficiency ϵ_S of a signal S in the region \mathcal{R} is given by the ratio of the number of signal events in \mathcal{R} that pass the orthogonality cut to the total number of signal events in \mathcal{R} . The efficiency is evaluated in bins of p_T of the reconstructed hadronic decay to separate the events into decays that are kinematically similar. In the merged regions the p_T corresponds to the p_T of the large-R jet, while in the resolved regions is taken as the scalar sum of the p_T of the two leading small-R jets. In blue are the efficiencies of the HVT-VH signal, in red of the HVT-VV signal. The dark hue represents the MCT strategy, while the light hue follows the MassCut strategy. Note that for the former, the histograms are exclusive, e.g. an event cannot enter more than one histogram, while this is not necessarily true for the MassCut strategy due to the mixed-overlap-region. The ideal cut would leave 100% of VH signal in VH SRs, and remove 100% VH signal from VV SRs, and vice versa. In the resolved SR the two strategies remove the same amount of “wrong” signal and leave most of the “correct” signal. In the merged SRs, the MassCut strategy removes most of the VV signal from VH SRs, but also a significant portion of VH signal. On the other hand, the MCT strategy does a more efficient sorting, leaving most of the correct signal in both VV and VH signal regions. This behavior is observed for all lepton channels, making the MCT strategy a more efficient event categorization strategy.

7.10.5 Signal significance

Selection criteria are often optimized with respect to the expected significance S of a given signal hypothesis. The concept of the significance was discussed in Sec.6.3. In these studies, an approximation is used, where S is estimated as the number of standard deviations of the background distribution to which the signal corresponds. Consider n events where $n = n_b + n_s$. Here n_b is the total number of MC events from known SM processes and n_s is the number of MC generated signal events. The quantity n_b is assumed to be known with an uncertainty σ_n , which in the following is just the statistical uncertainty in a given bin. The random variable n is assumed to be Poisson distributed, with Poisson error given by \sqrt{n} . The significance S is then calculated as,

$$\mathcal{S}_i = \frac{n_s}{\sqrt{n_b + \sigma_b^2}}. \quad (7.9)$$

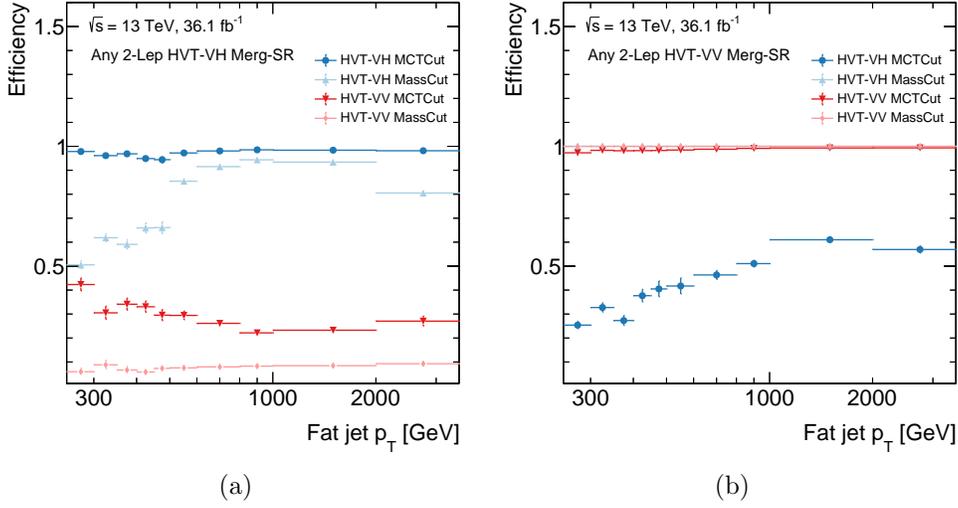


Figure 7.33: Efficiency of HVT-VH and HVT-VV signals as a function of p_T of the large-R jet in merged signal regions (SRs). The events are grouped according to whether they pass a given selection, according to Tab. 7.13, into HVT-VH Merg SR (left) and HVT-VV Merg SR (right).

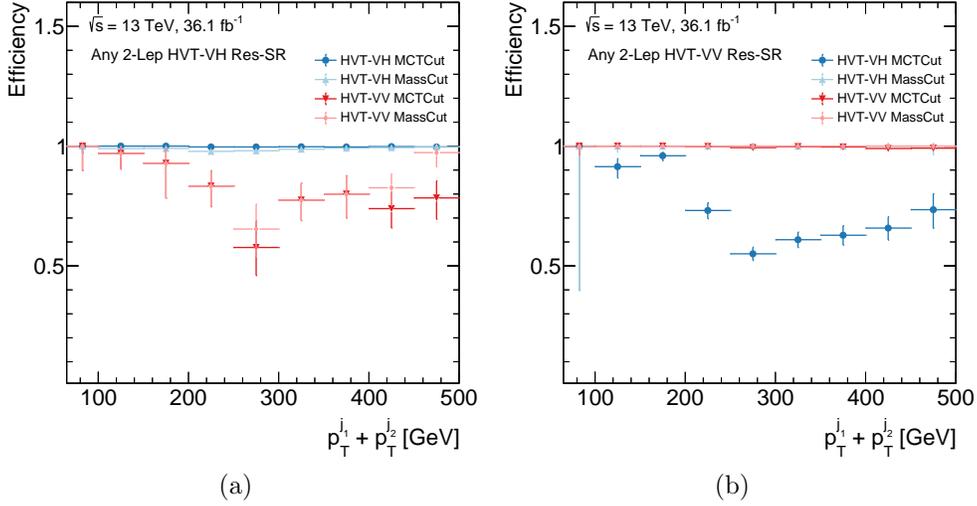


Figure 7.34: Efficiency of HVT-VH and HVT-VV signals as a function of the scalar sum of the p_T of the two leading small-R jets in resolved SRs. The events are grouped according to whether they pass a given selection, according to Tab. 7.13, into HVT-VH Res SR (left) and HVT-VV Res SR (right).

When the data is binned, each bin i contains n_i events, and a significance S_i can be calculated for each bin. The total binned significance of a histogram is then obtained as,

$$\mathcal{S} = \sqrt{\sum_i S_i^2}. \quad (7.10)$$

In the following, the histogram used to calculate the total binned significance is that of the final discriminant of the analysis, i.e. the invariant mass distribution. In a given signal region, the total significance is calculated for different signal mass points, and is shown in a significance scan as a function of the signal mass.

The expected HVT signal significance as a function of V' resonance mass is shown in Figs. 7.35 and 7.36 for VH signal regions, and in Figs. 7.37 and 7.38 for VV signal regions. This is shown for the 2-lepton channel, with similar performance having been observed for the 0- and 1-lepton channels. The MassCut strategy causes a reduction in VH signal significance in the Merged VH signal regions, while the MCT strategy does not affect the significance. In the VV signal regions, the MassCut strategy gives the same result as the *Baseline* analysis by construction, while the MCT strategy is not observed to bring any decrease in performance.

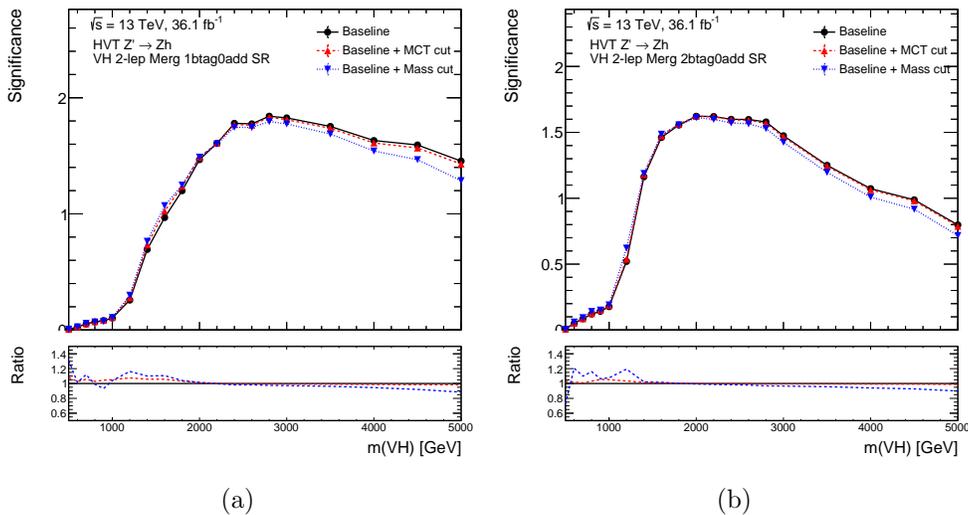


Figure 7.35: Significance scans as a function of Z' resonance mass in VH merged signal regions. The baseline analysis is compared to the MCT and MassCut strategies for orthogonalization.

7.10.6 Expected limit sensitivity

The expected limit sensitivities when using the MCT and MassCut orthogonalization strategies are compared. The limits are calculated following the procedure described in Sec. 6.3 without including systematics. The likelihood is built using Asimov data, as the analysis has not been unblinded yet. Figs. 7.39, 7.40, and 7.41 show the limit for Z' or W' signal interpretations calculated for the mass points [300 GeV, 500 GeV, 1 TeV, 2 TeV, 3 TeV, 4 TeV, 5 TeV] for 0-, 1-, and 2-lepton channels, respectively. As expected from the previous studies, the MCT strategy does not cause any loss in performance with respect to the *Baseline* analyses.

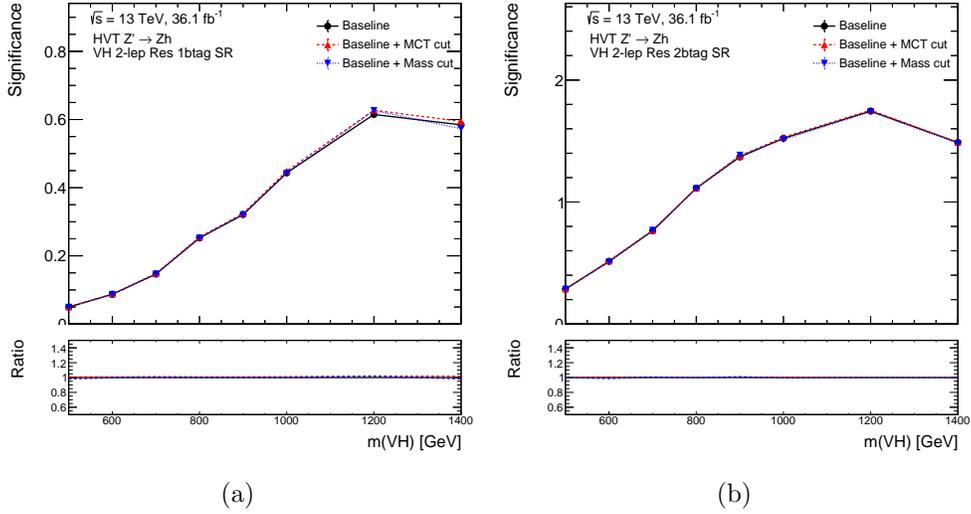


Figure 7.36: Significance scans as a function of Z' resonance mass in VH resolved signal regions. The baseline analysis is compared to the MCT and MassCut strategies for orthogonalization.

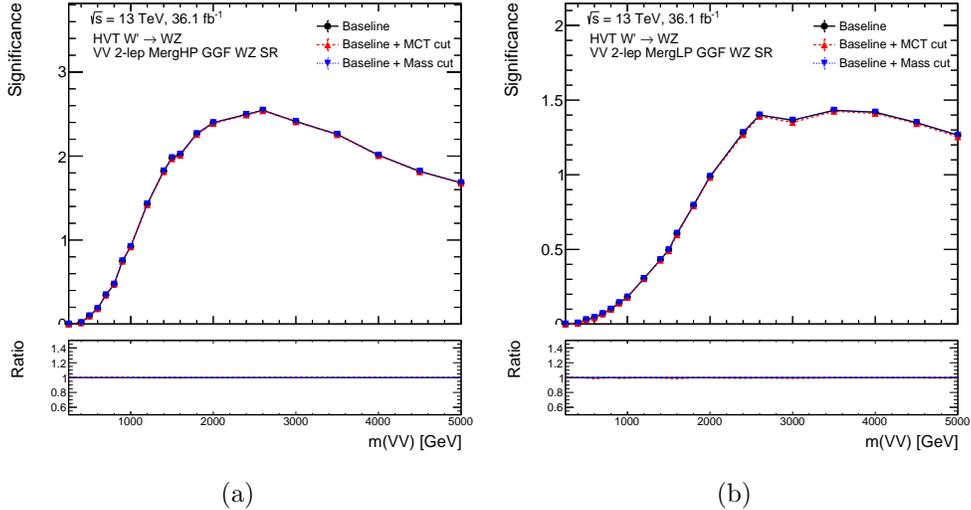
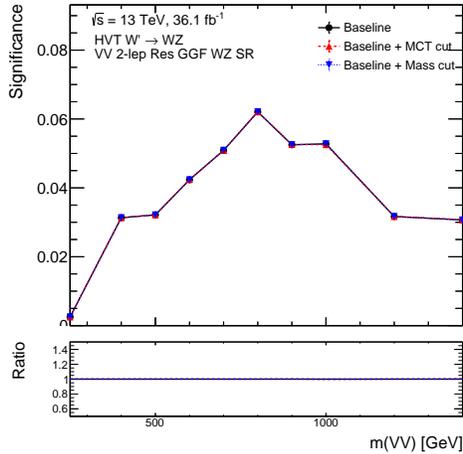


Figure 7.37: Significance scans as a function of Z' resonance mass in VV merged signal regions. The baseline analysis is compared to the MCT and MassCut strategies for orthogonalization.

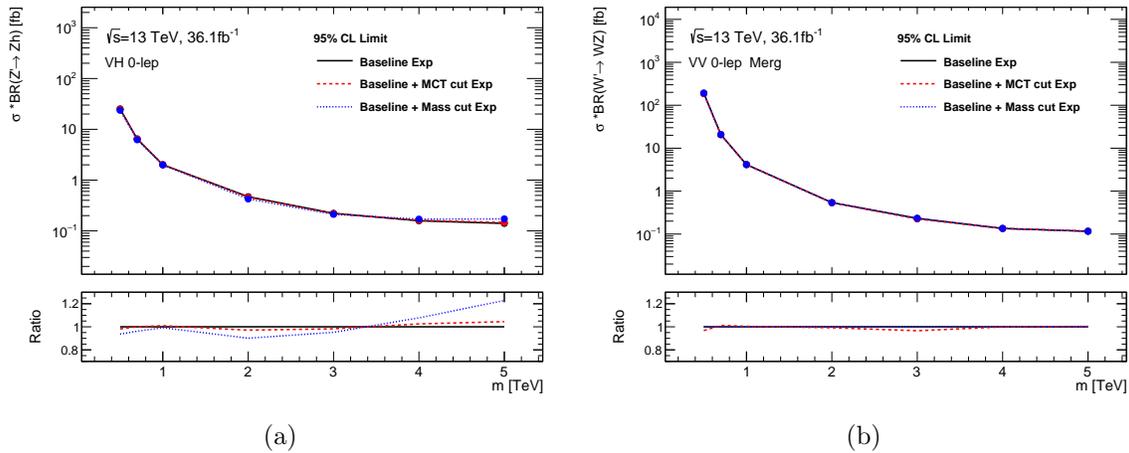
On the other hand, the *Mass cut* strategy causes a loss of sensitivity at high mass in the 0- and 2-lepton channels. A similar loss was observed in previous combination efforts. This is due to the higher mass region being more dependent on the merged signal regions, where the cut on the mass of the large-R jet was observed to decrease the VH signal efficiency and significance.



(a)

Figure 7.38: Significance scans as a function of Z' resonance mass in VV resolved signal regions. The baseline analysis is compared to the MCT and MassCut strategies for orthogonalization.

In conclusion, the MCT strategy does not cause any loss in sensitivity in the limits, allows to recover up to 20% loss in sensitivity at high resonance mass with respect to the *Mass cut* strategy, and allows to simplify the combined search for new heavy resonances.



(a)

(b)

Figure 7.39: Expected limits in VH and VV 0-lepton channel, shown in the inclusive regions. The baseline analysis is compared to the MCT and MassCut strategies for orthogonalization.

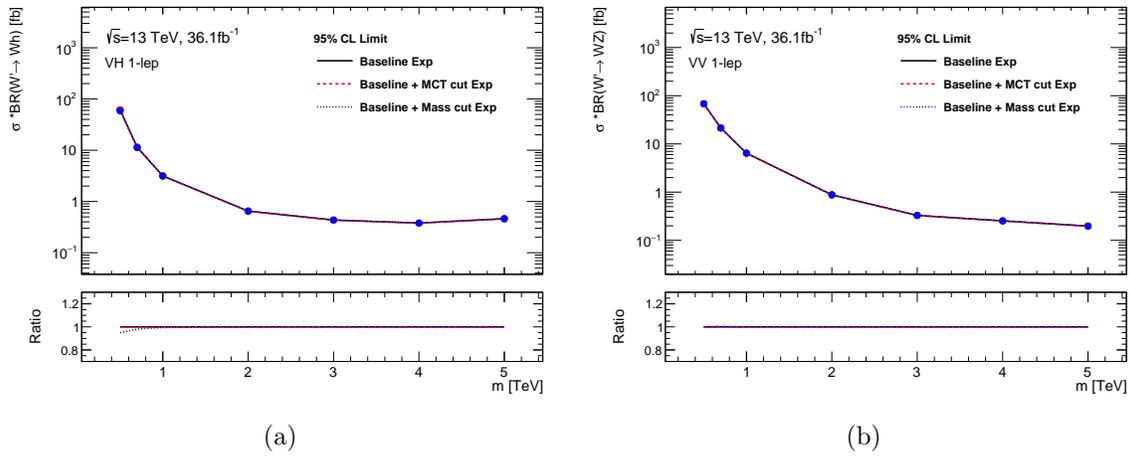


Figure 7.40: Expected limits in VH and VV 1-lepton channel, shown in the inclusive regions. The baseline analysis is compared to the MCT and MassCut strategies for orthogonalization.

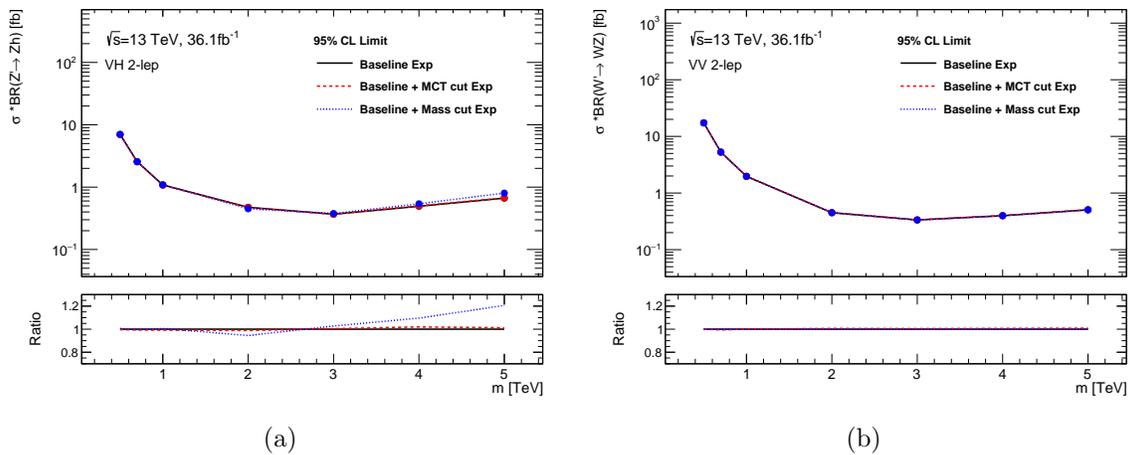


Figure 7.41: Expected limits in VH and VV 2-lepton channel, shown in the inclusive regions. The baseline analysis is compared to the MCT and MassCut strategies for orthogonalization.

7.11 MCT Modeling

In this section, the modeling of the Multi-Class Tagger (MCT) scores is studied in the context of the analysis. As described in the previous section, the MCT is used to orthogonalize the VV and VH signal regions, with a recycling strategy that ensures that no event is cut away. Most importantly, this means that the cut on the MCT is applied only to the small subset of events that ends up in both VV and VH signal regions. As long as the MCT scores are well modeled and cuts on the MCT scores do not produce or exacerbate mis-modeling, a full calibration of the classifier is deemed not necessary.

Being the analysis blinded, the modeling is studied in the pre-selection (Sec. 7.11.2) and control regions (Sec. 7.11.3). The pre-selection regions allow to have access to larger statistics and to include all the events that will end up in the signal regions, while the modeling in a top-enriched control region is studied as representative of the procedure that would be used to calibrate a top or W -tagger.

In order to disentangle true mis-modeling coming from the MCT from that induced by differences in background contributions between data and Monte Carlo (MC), preliminary normalization scale factors (SFs) are derived for the background samples by fitting the reconstructed mass of the hadronic decay in the given region. This is discussed in Sec. 7.11.1. After fixing the normalizations of the backgrounds, the MCT scores look well-behaved and the SFs are close to one. One might argue that even in the case of scale factors close to unity, the scale factor itself comes with an uncertainty that has to be evaluated and included in the fit. However, because the MCT is evaluated for every event that enters the analysis, as the systematic variations are applied, the score distributions will vary accordingly. It is therefore argued that the possible sources of uncertainties are already taken into account in the way the statistical fit is performed. Nonetheless, a study is also included in Sec. 7.11.4 of the effect of an artificial systematic uncertainty in the MCT scores on the final analysis sensitivity, and no effect is observed.

7.11.1 Derivation of background normalization scale factors

The number of events from a background process b in a given region is given by $N_b = \sigma^b \cdot \epsilon_0^b \cdot L$, where σ^b is the theory background cross section, ϵ_0^b is the nominal experimental efficiency, and L is the luminosity. The difference between observed and expected background normalization can be corrected by deriving a scale factor of the form $\tau = N^{\text{data}}/N^{\text{MC}}$.

The normalization scale factors of the background processes can be obtained via a binned maximum likelihood fit (see Sec. 6.3 and, specifically, Eq. (6.14)) of the predicted observable

of interest to the observed data. The normalization scale factor for each background sample can be included in the fit model as nuisance parameters that allow to vary the normalization of the given process. The fit outputs the maximum likelihood estimator of the scale factors.

The nuisance parameters can be introduced in the fit as unconstrained nuisance parameters, which are allowed to take on any value, or as constrained parameters with a Gaussian prior. In the first case, the fit is fully data-driven and outputs directly the maximum-likelihood-estimator of the scale factor $\hat{\tau}$. Nuisance parameters of this type introduce significant freedom in the optimization procedure, so they should only be used when necessary to avoid overfitting. In the case of constrained nuisance parameters, the efficiency is assumed to be sampled from a Gaussian with mean ϵ_0^b and standard deviation δ . The efficiency can then be parametrized as $\epsilon^b(\alpha) = \epsilon_0^b \cdot (1 + \delta \cdot \alpha)$, where the prior for α is a normal distribution with mean 0 and standard deviation 1. The best fit finds $\hat{\epsilon}^b(\alpha) = \epsilon_0^b \times (1 + \delta \cdot \hat{\alpha})$. The corresponding scale factor is given by $\hat{\tau} = 1 + \delta \cdot \hat{\alpha}$, with uncertainty $\delta \cdot \hat{\alpha}$.

An uncertainty associated to the luminosity is also included in the fit as a constrained nuisance parameter with a Gaussian prior and it is applied to all non-data-driven normalization coefficients. The Gaussian prior uncertainty is obtained from the auxiliary measurement of the total integrated luminosity from Run 2 of $(139.0 \pm 2.4) \text{ fb}^{-1}$, which corresponds to an uncertainty of $\delta_L = 1.7\%$.

In these studies, the normalization scale factors are derived from fitting the mass distribution of the hadronic decay, corresponding to the large-R jet mass in merged regions and the di-jet mass in resolved regions. The fits are performed only in unblinded regions, either pre-selection or control regions. Shape systematic uncertainties are not included, so the fits are limited by the irreducible contributions of background shape mismodeling. Nonetheless, the residual mismodelings are found to be small.

7.11.2 Modeling in pre-selection regions

The modeling of the MCT scores was studied in the pre-selection regions, which provide a large statistical sample and contain the important events that will end up in the signal regions. Events in the pre-selection regions were required to pass trigger, lepton selection, and anti-QCD cuts specific to each lepton channel, as well as the MCT training selection in Tab. 7.15. The results are shown in the 2-lepton pre-selection regions, while the 0- and 1-lepton channels are provided in the Appendix 9.

The V +jets ($V = W, Z$) MC background was separated into sub-samples according to the truth flavor of the jets - the two leading signal jets or the two leading track jets in the large-R jet, according to the region of interest. As the MCT scores are sensitive to the

Merged MCT	Resolved MCT
$m^J \in [50, 200]$ GeV	$p_T^{j1} > 45$ GeV and $ \eta^{j1} < 2.5$
$ \eta^J < 2$	$p_T^{j2} > 20$ GeV and $ \eta^{j2} < 2.5$
$p_T^J \in [200, 3500]$ GeV	$p_T^{j1} + p_T^{j2} < 500$ GeV

Table 7.15: MCT training cuts applied to the pre-selection region definition.

different flavor contributions, this was necessary to disentangle mismodeling originating from incorrect background normalizations from mismodeling induced by the MCT. The following V +jets sub-samples were defined according to the truth flavor of the jets:

- $V + bb, V + cc, V + bc$: The two jets are truth tagged as b/c -quarks.
- $V + bl, V + cl$: One jet is truth tagged as a b/c -quark, while the other jet is tagged as a light-quark.
- $Z + l$: No jet is truth-tagged as a heavy quark, and one or two jets is truth tagged as a light-quark.
- $V + c, V + b$: Only one jet, truth tagged as a b/c -quark.
- V : No signal jets.

In the following studies, the last two categories are grouped into the V, Vc, Vb sub-sample, as they bring a negligible contribution⁵.

When considered inclusively, the different sub-samples have a similar shape. In order to provide the fit with a handle on the different flavor contributions, each fitted region was separated into three sub-regions according to the number of b -tagged jets – either zero, one, or two. The three regions are provided to the fit model and one normalization scale factor is output for each V +jets background sub-sample, as well as for the other SM backgrounds.

The following data/MC comparison plots show the MCT scores $p(h)$ and $p(V)$ for the Merged (Resolved) MCT in the merged (resolved) pre-selection regions, both inclusively in Figs. 7.42 and 7.44, and in the zero, one, and two b -tagged regions in Figs. 7.43 and 7.45. The latter distributions show the different dominant contributions from the V +jets components: $Z + l$ in the zero b -tagged region; $Z + bl, Z + cl$ in the one b -tagged region; and $Z + bb, Z + bc, Z + cc$ in the two b -tagged region. The corresponding normalization uncertainties were left unconstrained in the fit for the given region. The jet mass distribution used for

⁵The event selection in VH regions and VV resolved regions always requires at least two jets, so the last two categories should be empty. In VV merged regions, there is no minimal requirement on the number of track jets associated to the large-R jet, so it is possible for these regions to be populated.

the fit is also shown to gauge the presence of residual shape mismodeling. The uncertainties in the plots are given by the combined statistical uncertainties and the uncertainties for the normalization coefficients derived from the fit. The only noticeable mismodeling appears in the resolved $p(h)$ score in Fig. 7.45b. However, part of the mismodeling is likely due to the residual shape mismodeling in the jet mass distribution 7.45a. In addition, once a full treatment of the systematic uncertainties is included, the fits are expected to improve and the uncertainties to increase. Overall, the MCT is therefore observed to be well-modeled in the pre-selection regions.

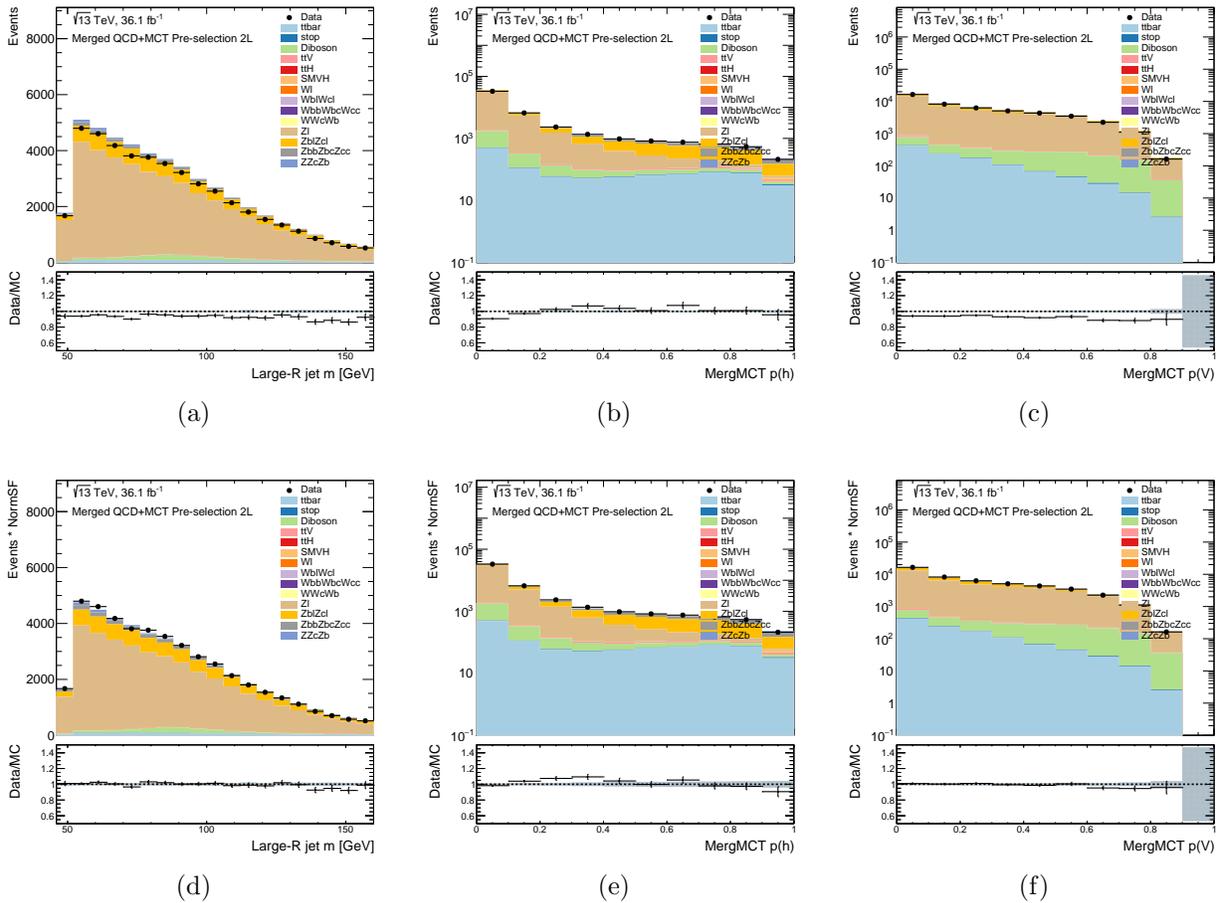


Figure 7.42: Data and MC comparison in the inclusive merged pre-selection region in the 2-lepton channel. Distributions of the large-R jet mass and the raw merged MCT scores $p(h)$ and $p(V)$ are shown before (top) and after (bottom) applying the derived normalization SFs.

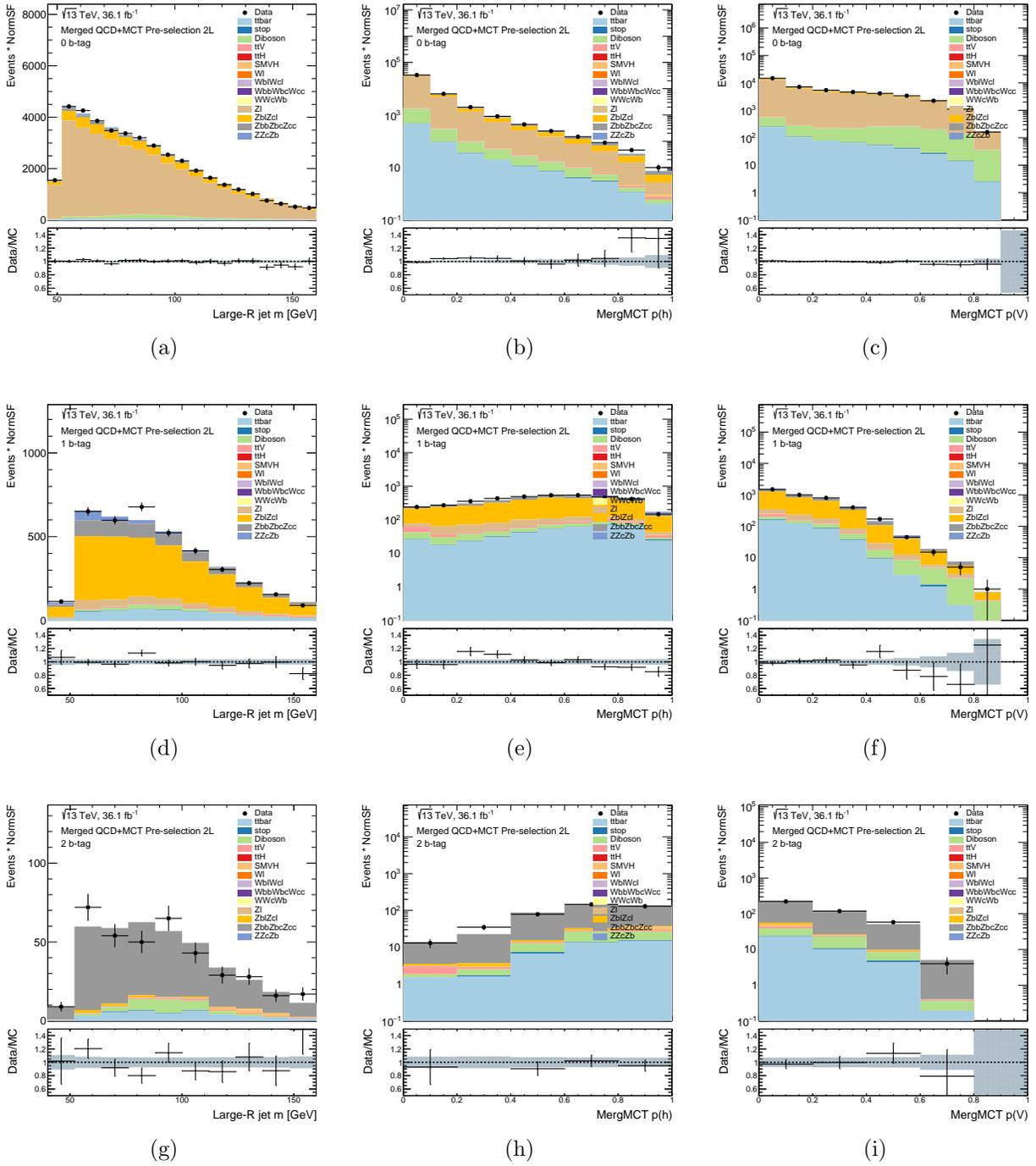


Figure 7.43: Data and MC comparison in the merged pre-selection region separated by the number of b -tagged jets in the 2-lepton channel. Distributions of the large- R jet mass and the raw merged MCT scores $p(h)$ and $p(V)$ are shown after applying the normalization SFs in the 0 b -tag (top), 1 b -tag (middle), and 2 b -tag (bottom) regions.

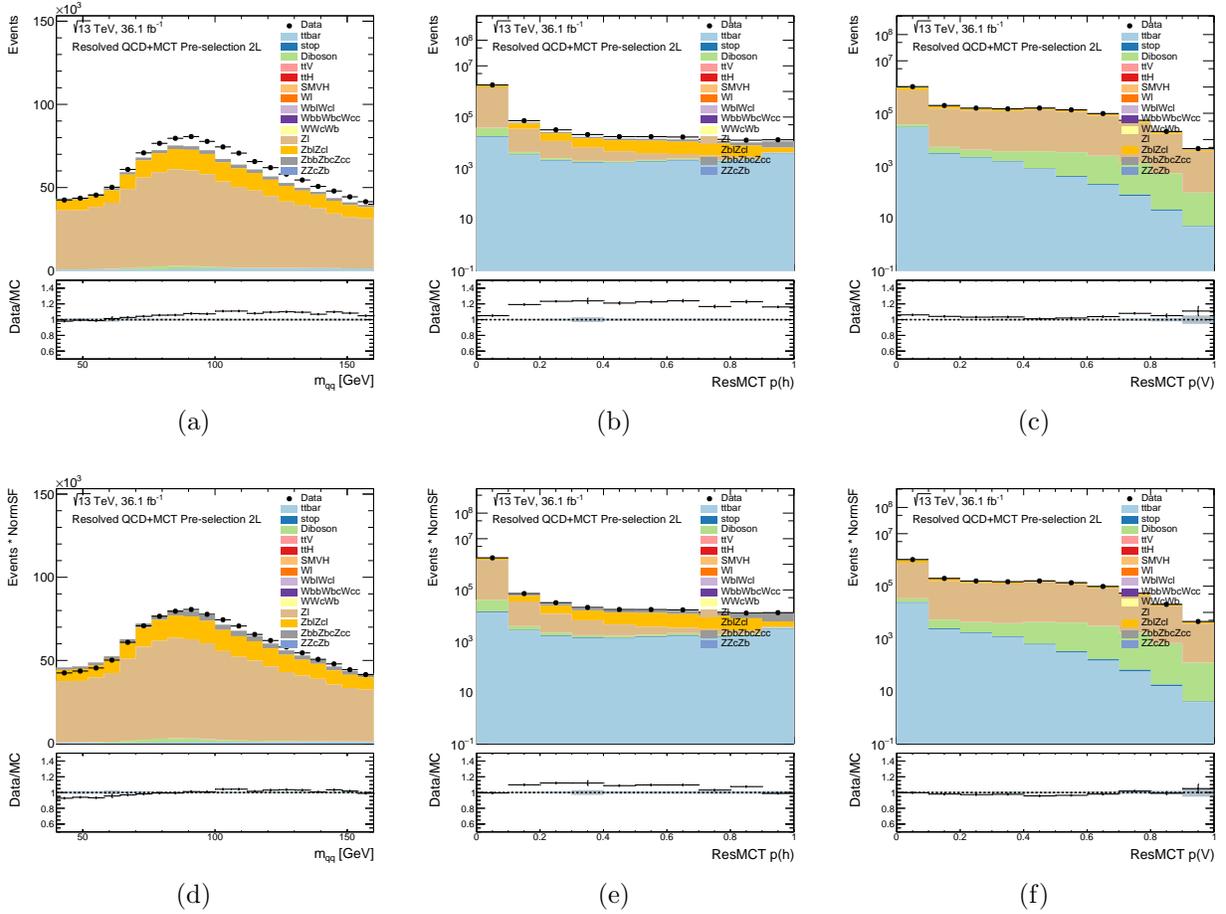


Figure 7.44: Data and MC comparison in the inclusive merged pre-selection region in the 2-lepton channel. Distributions of the large- R jet mass and the raw merged MCT scores $p(h)$ and $p(V)$ are shown before (top) and after (bottom) applying the normalization SFs.

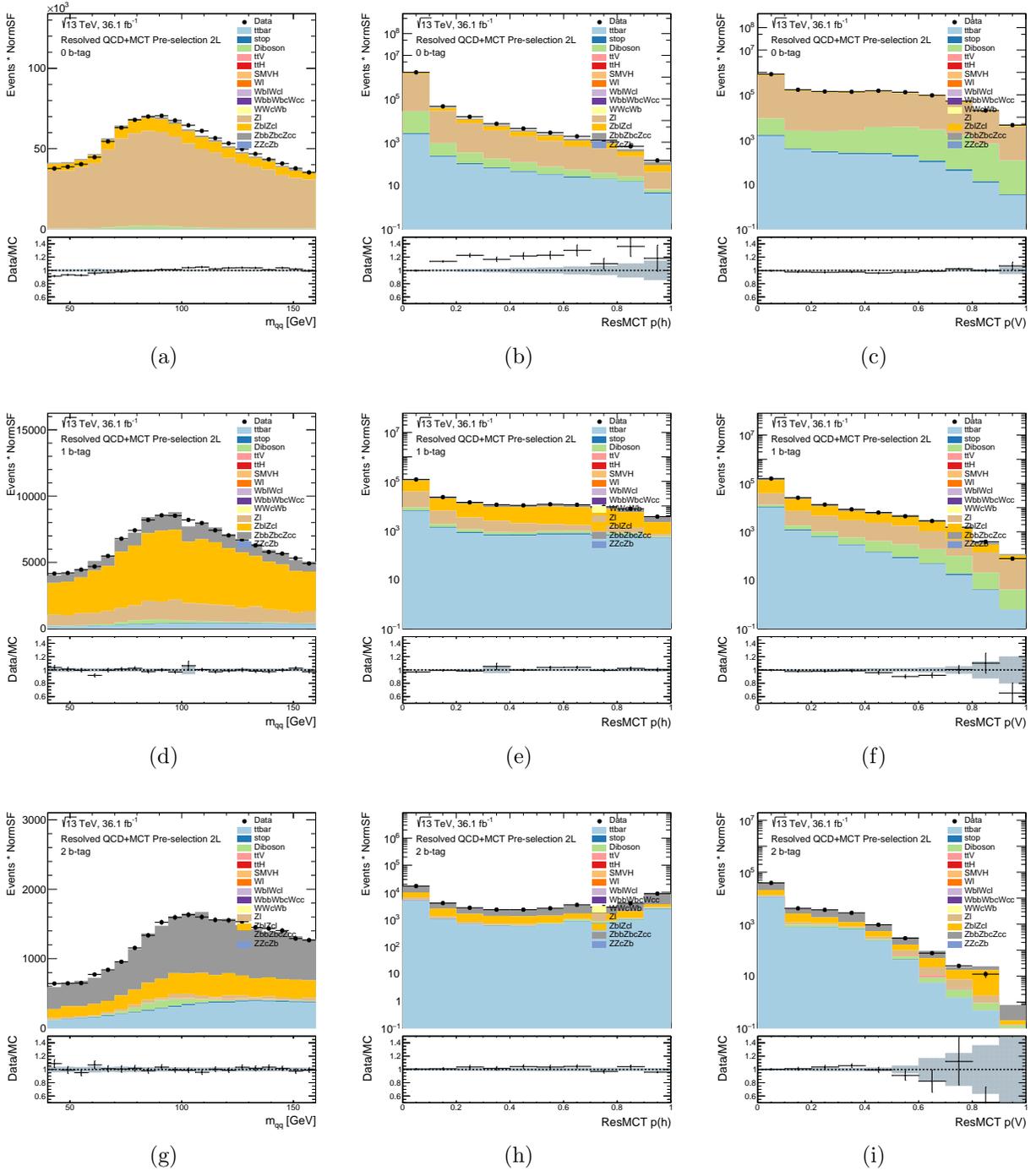


Figure 7.45: Data and MC comparison in the resolved pre-selection regions separated by the number of b -tagged jets in the 2-lepton channel. Distributions of the di-jet mass and the raw resolved MCT scores $p(h)$ and $p(V)$ are shown after applying the normalization SFs in the 0 b -tag (top), 1 b -tag (middle), and 2 b -tag (bottom) regions.

7.11.3 Modeling in top-enriched control region

A similar study as discussed in the previous section was carried out in the top-enriched CR `VV1Lep_MergHP_GGF_WZ_01bttag_TCR`. This was done to emulate what would be the procedure if one were to calibrate the MCT, as a top CR provides a subset of events rich in true W jets coming from non-contained top decays. In this case, there was no need to separate the V +jets background into the flavor components, as the only dominant background is $t\bar{t}$. The region was fit inclusively, with only the $t\bar{t}$ normalization uncertainty left unconstrained. The results are shown in Fig. 7.46. As observed in the pre-selection regions, after applying the normalization SFs no differences are observed between data and MC.

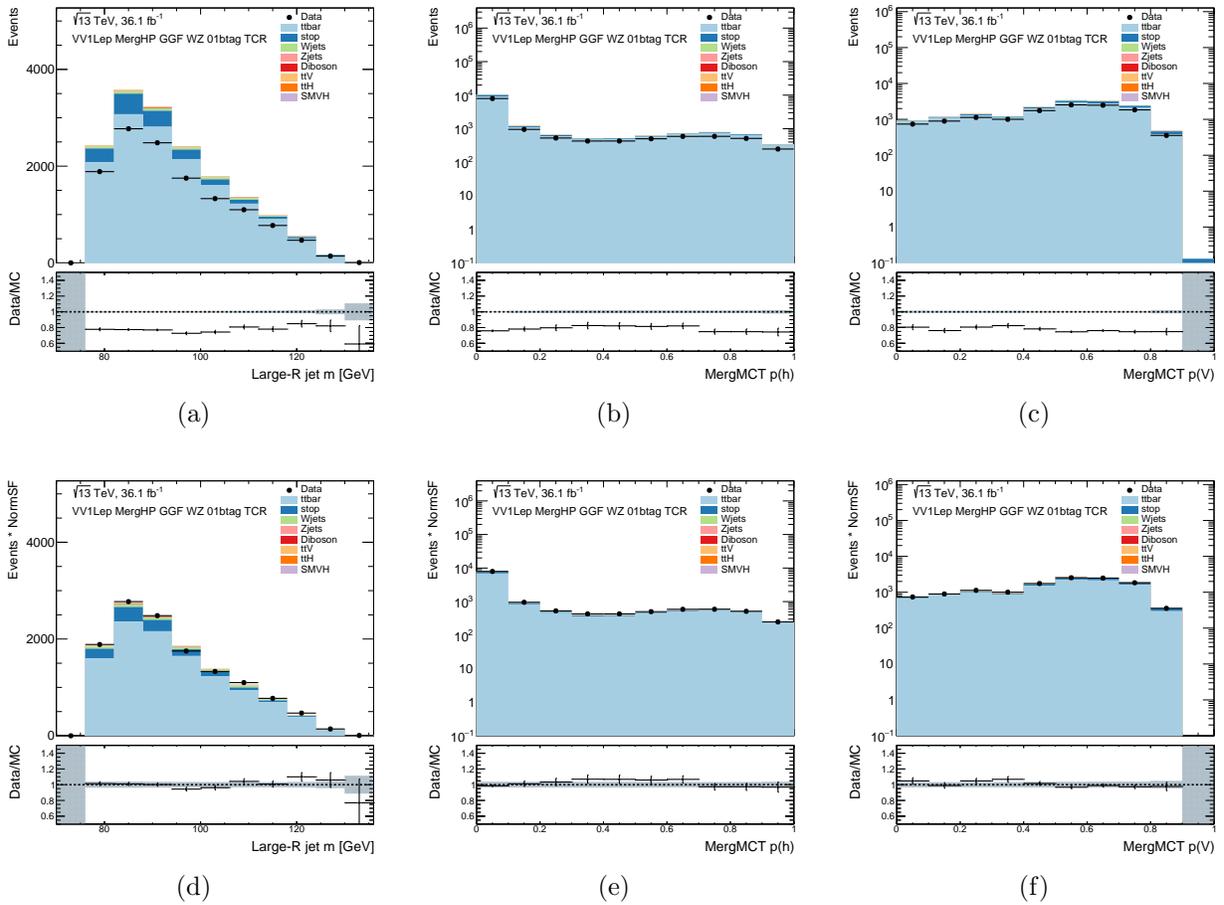


Figure 7.46: Data and MC comparison in top-enriched CR `VV1Lep_MergHP_GGF_WZ_01bttag_TCR`. Distributions of the large- R jet mass and the merged MCT scores $p(h)$ and $p(V)$ are shown before (top) and after (bottom) applying the normalization SFs.

7.11.4 Sensitivity to systematic variations of MCT scores

To conclude the MCT modeling studies, it was investigated whether an artificial upward or downward variation on the MCT scores would affect the analysis sensitivity. The expectation is that the limits would not be affected anyway because of the small subset of events to which the MCT is applied. The mismodeling was simulated as a $\pm 10\%$ systematic effect on the resolved (merged) $p(h)$ score, which was propagated to the resolved (merged) $p(V)$ score as $p(V') = p(V) + (p(h) - p(h'))$ and with the scores min-maxed-out at 0 and 1. Note that this is an unrealistic extreme scenario, as the error would most likely be propagated more evenly among all the remaining four scores.

0 Lepton channel

		data	HVT-WZ	HVT-WW	HVT-ZH	HVT-WH	ttbar	Wjets	Zjets
Nominal	VH	65.52	21.55	–	76.95	–	55.01	75.35	80.34
	VV	34.48	78.45	–	23.05	–	44.99	24.65	19.66
p(h)*1.1	VH	69.03	25.88	–	81.91	–	57.33	79.52	84.31
	VV	30.97	74.12	–	18.09	–	42.67	20.48	15.69
p(h)*0.9	VH	60.01	16.63	–	70.21	–	51.54	68.59	73.69
	VV	39.99	83.37	–	29.79	–	48.46	31.41	26.31

1 Lepton channel

		data	HVT-WZ	HVT-WW	HVT-ZH	HVT-WH	ttbar	Wjets	Zjets
Nominal	VH	83.69	57.49	22.00	–	89.17	84.55	83.13	86.02
	VV	16.31	42.51	78.00	–	10.83	15.45	16.87	13.98
p(h)*1.1	VH	86.42	63.91	26.33	–	91.59	87.47	86.90	89.10
	VV	13.58	36.09	73.67	–	8.41	12.53	13.10	10.90
p(h)*0.9	VH	78.55	49.22	17.18	–	85.27	80.25	77.02	79.87
	VV	21.45	50.78	82.82	–	14.73	19.75	22.98	20.13

2 Lepton channel

		data	HVT-WZ	HVT-WW	HVT-ZH	HVT-WH	ttbar	Wjets	Zjets
Nominal	VH	67.23	21.89	–	82.43	–	74.55	100.	76.30
	VV	32.77	78.11	–	17.57	–	25.45	0.00	23.70
p(h)*1.1	VH	72.09	25.87	–	85.95	–	77.73	100.	79.68
	VV	27.91	74.13	–	14.05	–	22.27	0.00	20.32
p(h)*0.9	VH	61.17	16.66	–	77.19	–	70.00	100.	70.57
	VV	38.83	83.34	–	22.81	–	30.00	0.00	29.43

Table 7.16: Percentage of events in overlap region that end up in VH or VV analysis using the MCT strategy, comparing the nominal scenario, and the effect of an up and down variation on the scores (see text). All values are given as percentages of the number of events in the overlap region for the given data or MC sample.

Tab. 7.16 shows the migration of events after the MCT sorting in the nominal, up-, and down-variation scenarios. Figs. 7.47 and 7.48 show the inclusive signal efficiencies as a function of p_T , similarly to the studies presented in Sec. 7.10.4, for the three scenarios. The only significant difference is observed in the background efficiencies (defined as the efficiency of a signal in the incorrect SR). Indeed, this resulted in no difference in the limit sensitivities (not shown) and similar results were observed in the other lepton channels. The conclusion therefore is that the analysis is not sensitive to possible systematic variations of the MCT scores.

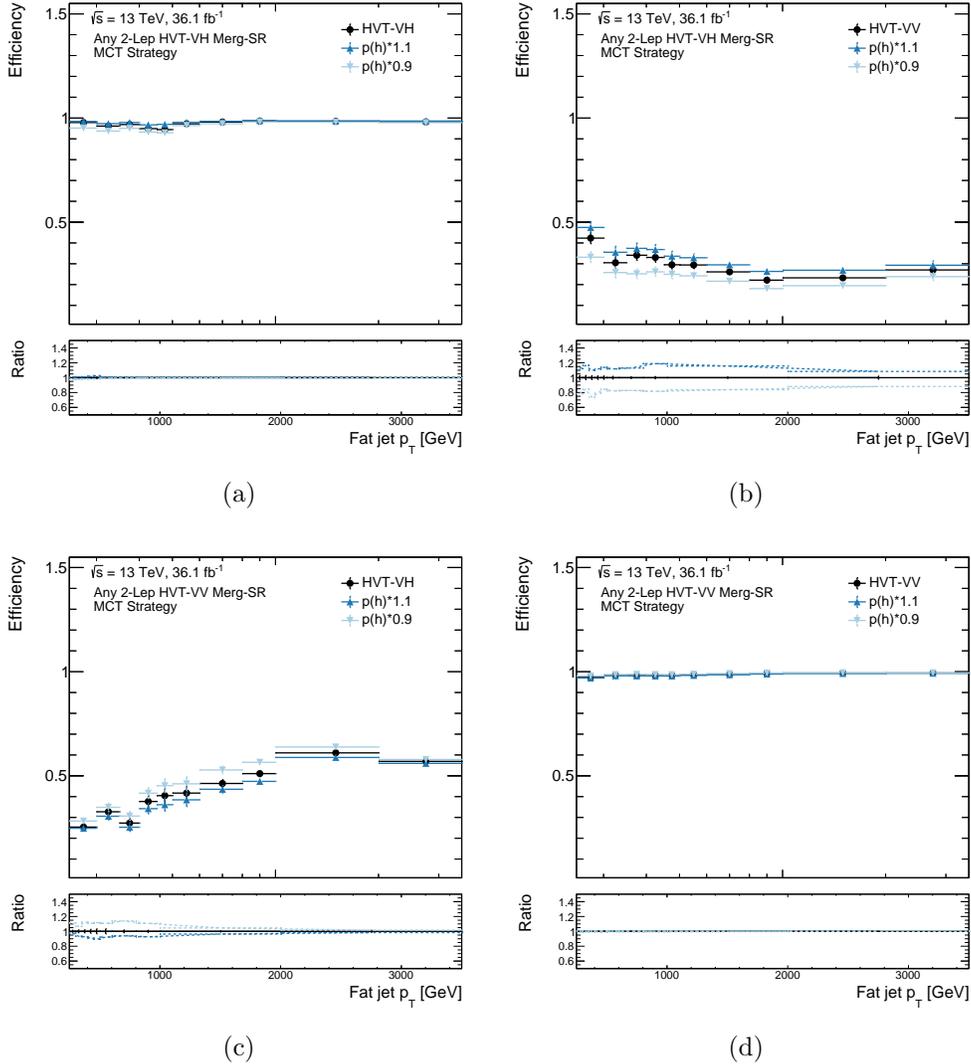


Figure 7.47: Signal efficiency as a function of p_T in the 2-lepton channel for VV and VH merged signal regions, after orthogonalization with the MCT strategy. The nominal result is compared with the effect of an artificial up- and down-variation on the resolved and merged $p(h)$ scores (see text for explanation).

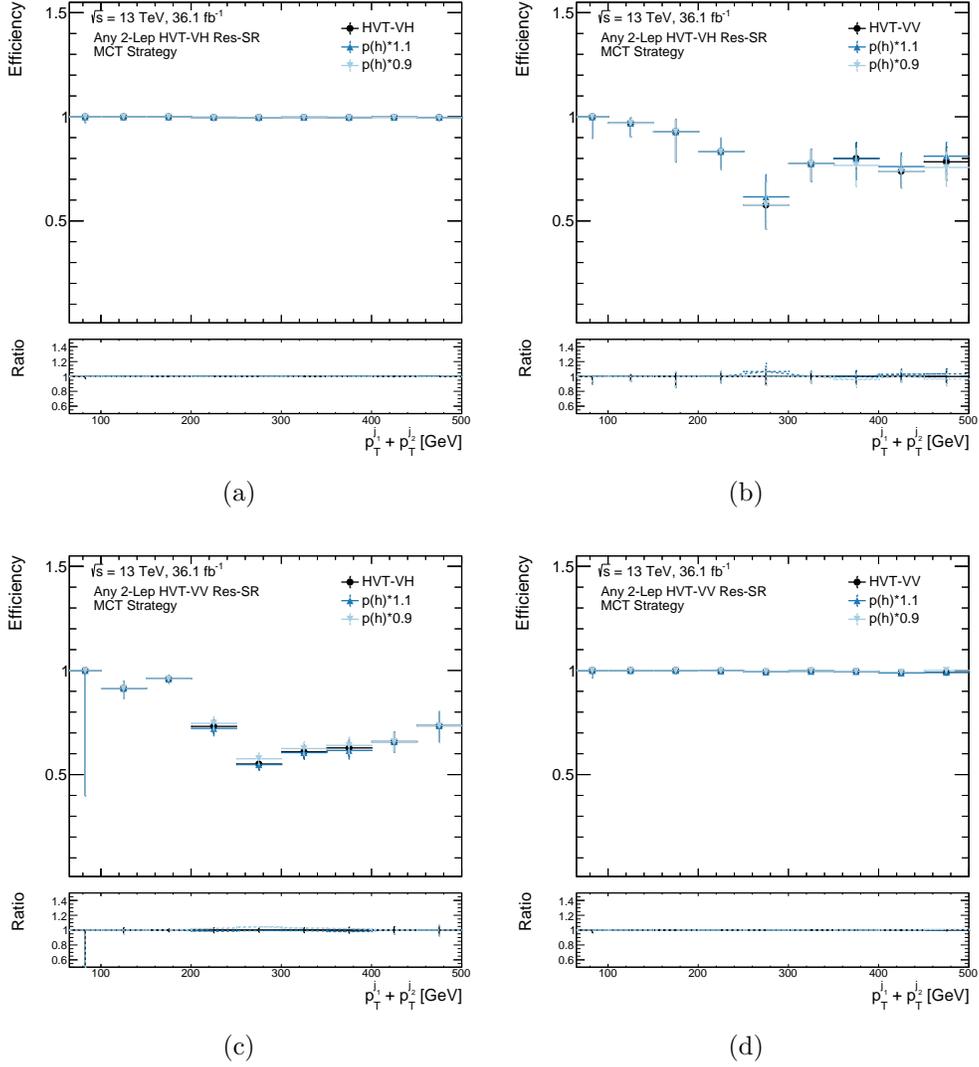


Figure 7.48: Signal efficiency as a function of p_T in the 2-lepton channel VV and VH resolved signal regions, after orthogonalization with the MCT strategy. The nominal result is compared with the effect of an artificial up- and down-variation on the resolved and merged $p(h)$ scores (see text for explanation).

Chapter 8

Firmware algorithm development for the HL-LHC Global Trigger upgrade

The Global Trigger (GT) will be a major addition to the ATLAS Level-0 trigger system to be installed during the Phase II upgrades in preparation for the High-Luminosity LHC (HL-LHC¹). In order to handle the larger event size and unprecedented pileup levels, the GT will provide a platform to run complex algorithms at the first stage of the trigger chain and bring the event rate from 40 MHz down to 1 MHz. It is important to note that there will be no legacy triggers left as back-up at the beginning of Run 4. The successful operation of the GT will be necessary for ATLAS to take data.

The trigger installed during the Phase II upgrades is expected to run for more than ten years, during which physics objectives might change, possibly due to new discoveries. One of the design principles of the GT is therefore for it to be sufficiently adaptable to allow ATLAS to react as quickly as possible to such changes. For this reason, the different functions that execute the trigger algorithms will be implemented in firmware, which provides more flexibility than standard hardware triggers. The firmware will then be executed on a common hardware platform based on Field Programmable Gate Arrays (FPGAs), which in turn reduces the hardware complexity. It follows that the GT is primarily a firmware project, a very different paradigm from what historically has been the hardware-based trigger in ATLAS. Most of the work within the GT upgrade project goes into the software and firmware co-development of the new trigger algorithms. A significant contribution of this thesis was the development of the software simulation framework for GT firmware algorithm development, as well as the development of a new jet reconstruction and triggering strategy.

Every trigger is designed to target a specific physics signature. For instance, events including hadronic decays of heavy particles are often characterized by the presence of at least one energetic small-R jet. In order to record this type of events, a one-jet trigger selects events where the leading small-R jet has a transverse momentum (p_T) above a given threshold. In practice, of the forty million events that the Level-0 trigger receives per second,

¹See Sec. 4.3.5 for an overview of the Phase II trigger upgrade.

only one million of them can be accepted. This 1 MHz event rate has to be shared between the different Level-0 triggers according to ATLAS physics goals and priorities. The number of events that a jet trigger can accept – i.e. how low the p_T threshold can be – is thus determined by the fraction of the 1 MHz event rate allocated to it. At the same time, the majority of the events seen by the detector are minimum bias events (see Sec. 4.3.3), while interesting collisions are orders of magnitude rarer. The ability to correctly discriminate between signal and background – which, for a jet trigger, can go from correctly reconstructing the jet energy, to identifying and discarding pileup-induced jets – is essential for maximizing the retention of rare signal events. In the noisy environment of the HL-LHC, this task will be significantly more difficult, to the point that, if the hardware trigger system was kept unchanged, the current algorithms would be unable to retain the physics performance required by the experiment. The deployment of fast algorithms able to perform a more sophisticated signal to background discrimination at Level-0 is necessary to maintain ATLAS physics reach.

A brief introduction to the GT system is given in Sec. 8.1. The remaining sections of this chapter describe the contributions to the GT project of this thesis. These include the development of the software simulation framework for trigger algorithms (Sec. 8.2.1), the design and validation of a new jet reconstruction and triggering strategy (Sec. 8.2 and 8.3), and studies for a pileup-jet suppression method using deep neural networks to improve multi-jet triggers (Sec. 8.4).

8.1 The Global Trigger (GT)

The GT² consists of three sub-systems: the Multiplexers, the Event Processors, and the Demultiplexers. Each sub-system consists of a farm of large FPGAs, with each FPGA corresponding to a *node*, a common hardware unit on which the same firmware is deployed. The serial data arriving at 40 MHz from the calorimeter detector subsystems, the Phase I FEXs (see Sec. 4.3.4), and the Muon Central Trigger Processor Interface (MUCTPI) is deserialized by Multiplexer Processor (MUX) nodes for pipelined data processing. The MUX aggregates the full event data from a specific bunch crossing (BC) on a single event processor node, called Global Event Processor (GEP). At this point the events are decoupled from the BC rate and can be processed in parallel on the 48 GEP nodes, allowing the implementation of asynchronous complex algorithms. The event data from each BC is distributed to the GEP

²The GT system is currently under development. The description presented here is mostly based on the one proposed in the Technical Design Report [10]. As this picture has been continuously evolving, certain timing and resource utilization estimates might have changed with time.

nodes in a round-robin fashion, with each node receiving new data every 48 BCs, increasing the latency between the arrival of two events to a GEP node from 25 ns to 1.2 μ s. This process is shown schematically in Fig. 8.1.

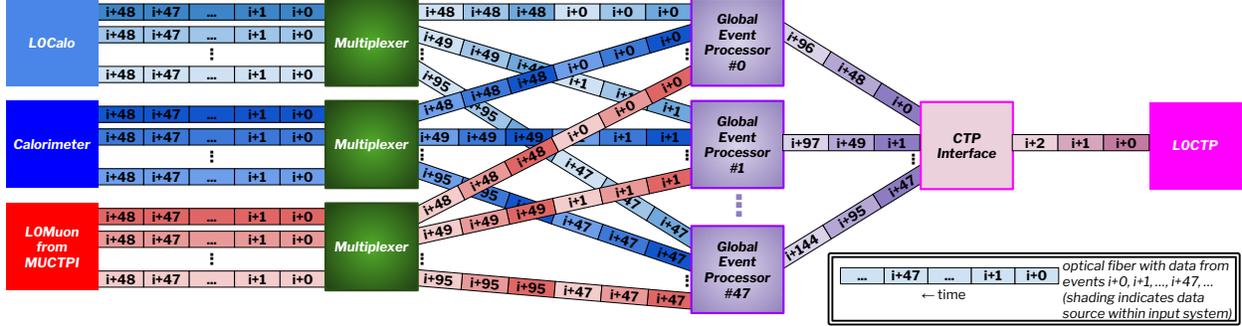


Figure 8.1: Time-multiplexing of incoming synchronous data at 40 MHz by the MUX nodes. Each MUX receives data for every bunch crossing (BC). The data is processed, organized, and dispatched to the GEP nodes. Each GEP node receives the complete event data for one BC and analyzes the data asynchronously. Results from the GEP nodes are demultiplexed by the CTPi and sent to the CTP [10].

The same set of firmware functions is executed on each GEP node to build the TOBs and produce trigger hypotheses based on object multiplicities, energy thresholds, and topological relationships, as sketched in Fig 8.2. As the data is being received, algorithms that do not require the full event data can start, fully exploiting the data transmission time. Ordering data geometrically also favors the pipelining of the steps of non-iterative algorithms on the FPGA board. For instance, if the data arrives ordered in η , a sliding window algorithm can start processing the detector plane in full slices over ϕ . Pipelined data processing and parallel execution of different algorithms, allows to drastically reduce the FPGA resource utilization and extend the Level-0 latency up to $\sim 6 \mu$ s.

The output of a GEP node is called the Trigger Input (TIP), containing flag bits of which trigger requirements have been satisfied and multiplicities of reconstructed objects. A Global-to-CTP Interface (gCTPi) demultiplexes the data, re-builds the event with the correct BC number, transmits the trigger bits to the CTP, and sends the data to the readout system on request. The CTP combines trigger inputs from the GT and MUCTPI, as well as from the forward detectors and other detector calibration sub-systems, and makes the final Level-0 trigger decision. It also applies deadtime and prescales. The Level-0 accept rate is of 1 MHz.

Inputs

At every LHC BC, the GT receives the full granularity noise-suppressed ($|E_T| > 2\sigma$)

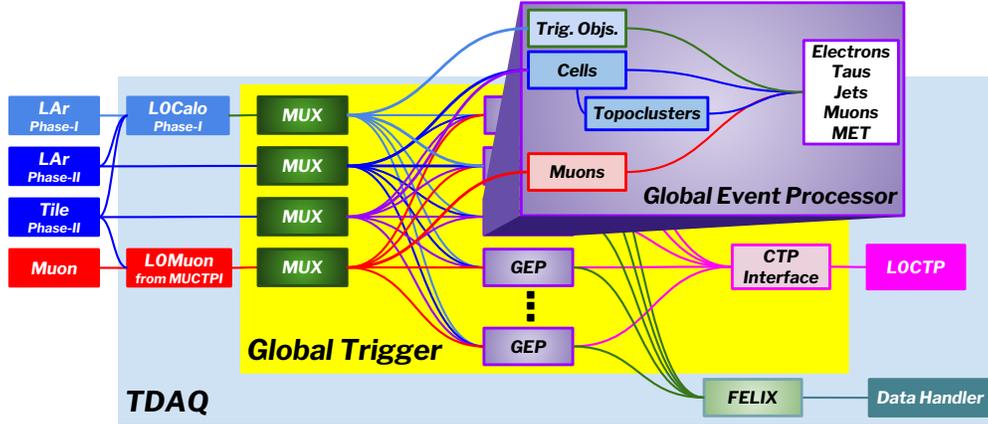


Figure 8.2: Schematic view of the Global Trigger processing [10].

calorimeter data from the LAr and Tile calorimeter front-end electronics. This will be the first time that the Level-0 trigger has access to calorimeter cells. This new information will enable the implementation of topological clustering, in turn allowing for improved TOBs definitions. The GT will also receive TOBs from the L0Calo FEX processors and from the MUCTPI, which can be used as seeds for GEP algorithms.

The LAr and Tile inputs are expected to arrive first in $1.4 - 1.7 \mu\text{s}$. The data from L0Calo is expected to arrive in $2 - 2.6 \mu\text{s}$, followed last by the L0Muon inputs at $4.5 - 5.3 \mu\text{s}$. The overall latency is always dictated by the muon latency. Because of the longer arrival time of the muon information, algorithms requiring inputs from the MUCTPI will be run later in the pipeline.

Hardware

The hardware is based on a common design, called the Global Common Module (GCM), to minimize the complexity of the system. A GCM with dedicated firmware is used for each of the three main components of the GT: a multiplexing module, a Global event processing module, and a demultiplexing CTPi module. Each GCM can support two independent FPGAs, where each FPGA can represent a GEP, MUX, or CTPi node, and a central processing chip for monitoring, control, and readout. The FPGA board will be the Versal Premium VP1802 developed by the company Advanced Micro Devices (AMD), providing substantial I/O and Random Access Memory (RAM) capabilities to handle the large bandwidths and buffering.

Firmware

The various algorithms that compute the trigger objects will be executed on the GEP hardware modules and will represent the majority of the resource consumption of the GT system.

Topoclustering, e/γ , and τ algorithms are expected to consume less than 1% of the resources. The resource usage will be dominated by the jet finding and trigger hypothesis algorithms. Resource usage for the latter can be estimated from the current L1Topo usage to be between 36% and 68%. This leaves about 20% of the FPGA resources for jet finding. The physics performance of the algorithms is limited by latency and resource allocation, while algorithm scheduling on the board is constrained by their role in the overall dataflow, e.g. the inputs they require. A preliminary plan for algorithm scheduling on a single GEP module is shown in Fig. 8.3. The FPGA is divided into four Super Logic Regions (SLR) pipelined in latency intervals of $1.2\ \mu\text{s}$. The first SLR0 receives LAr data, so that regional algorithms, such as topological clustering, can start processing. After $1.2\ \mu\text{s}$, the processing of the current BC is moved to SLR1, while SLR0 receives LAr data for a new event. Once in SLR1, inputs from the Tile calorimeter arrive, so that the regional algorithms can ultimate their trigger objects. The topoclusters are now ready and can be used as inputs by downstream algorithms, including jet finding, which will take up the majority of the resources in SLR2. Lastly, SLR3 will receive the muon information and will run topological algorithms on the final trigger objects. This picture might evolve with time.

8.2 Trigger performance studies

The trigger algorithms used to reconstruct the event and produce the trigger objects are implemented in reprogrammable firmware, which makes the GT effort primarily a firmware project. From the hardware side, the algorithms are constrained by the number of I/O ports, latency, and bandwidth requirements set by the Global Common Module. From the physics side, each algorithm is developed with the goal of providing a high signal efficiency. The development of a new candidate algorithm is thus performed with two equally important objectives: providing a high physics performance, while keeping the footprint in terms of FPGA resources within the hardware limitations. The first step in the development is generally to demonstrate that the physics performance of the algorithm is sufficient to retain ATLAS physics goals. This is usually done in software, as it provides a faster turn-around and allows one to concentrate on the physics questions. Once the algorithm is mature, the next step is to perform a preliminary firmware simulation in order to provide an estimate of the resource consumption. This is often done using packages such as High-Level Synthesis (HLS), which automate hardware design by taking as input a high-level algorithmic description in a standard language, such as C/C++, and converting it to a lower-level hardware description language (HDL) code. Often this process requires several iterations in order to simultaneously optimize the two objectives. In particular, if the resource usage significantly

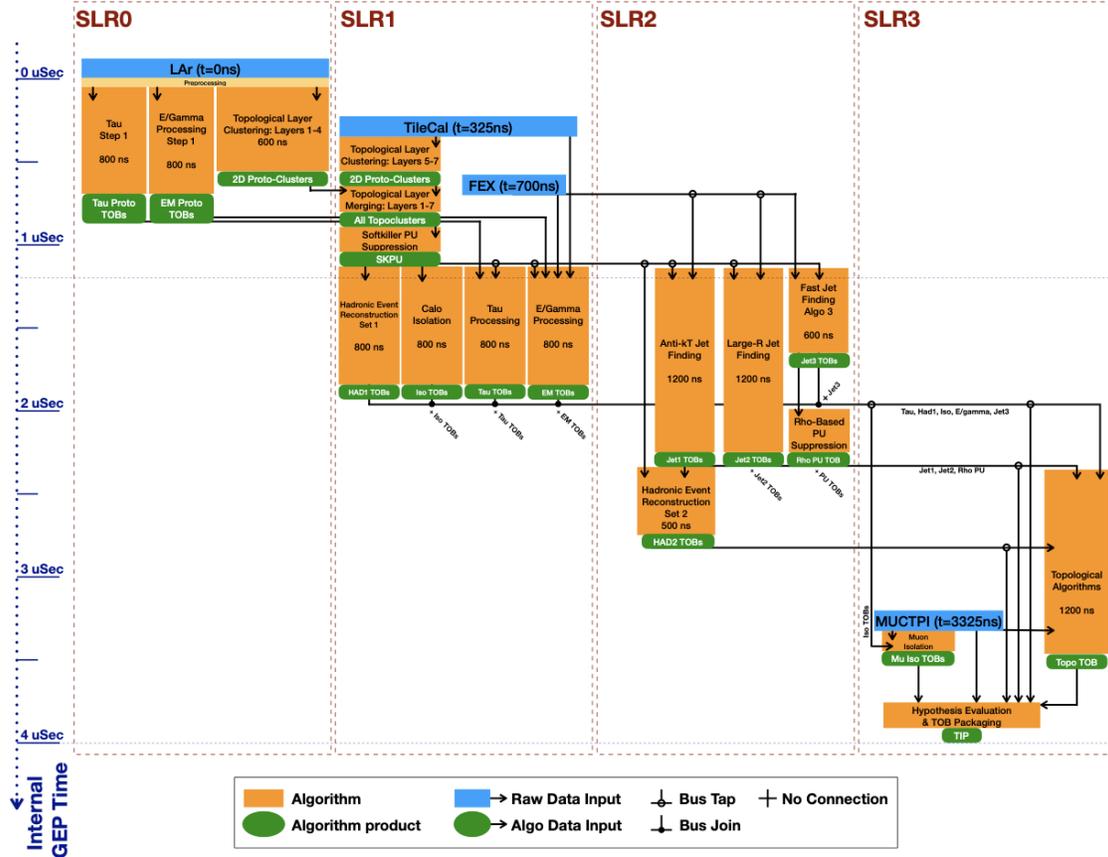


Figure 8.3: Preliminary plan for GT algorithm dataflow on one FPGA on a GEP module. The FPGA is divided into four Super Logic Regions (SLRs) for pipelined data processing. This picture might evolve with time. Picture courtesy of Wade.

exceeds the allocated lot, it is possible that substantial changes to the algorithm itself are needed, so that the software and firmware development proceed in parallel. Once the algorithm has been proven to be a viable option both from a physics and a firmware perspective, the firmware design is typically finalized by a hardware engineer.

In the context of trigger performance studies there are two levels of reconstruction. *Online* reconstruction is performed at the trigger level on the live stream of data arriving from the detector. The development of the online algorithms is constrained by the latency and bandwidth limitations of the trigger environment and often results in coarser objects. After the data has been selected by the trigger and stored to disk, the *offline* reconstruction is performed using the standard ATLAS software. The offline algorithms, run on CPU farms, have very few limitations in terms of resources, and are therefore maximally optimized. The objects output by these two stages will be referred to as *online* and *offline* objects. It should be noted that, in the following, no calibration is applied on the offline objects, as none were

available at the time of these studies.

8.2.1 The GT software simulation framework

Part of this thesis work included the development of the software simulation framework for the GT. The package was written in C++ and Python as part of the ATLAS offline software Athena [105]. A sketch of the functionalities of the framework is shown in Fig. 8.4. The

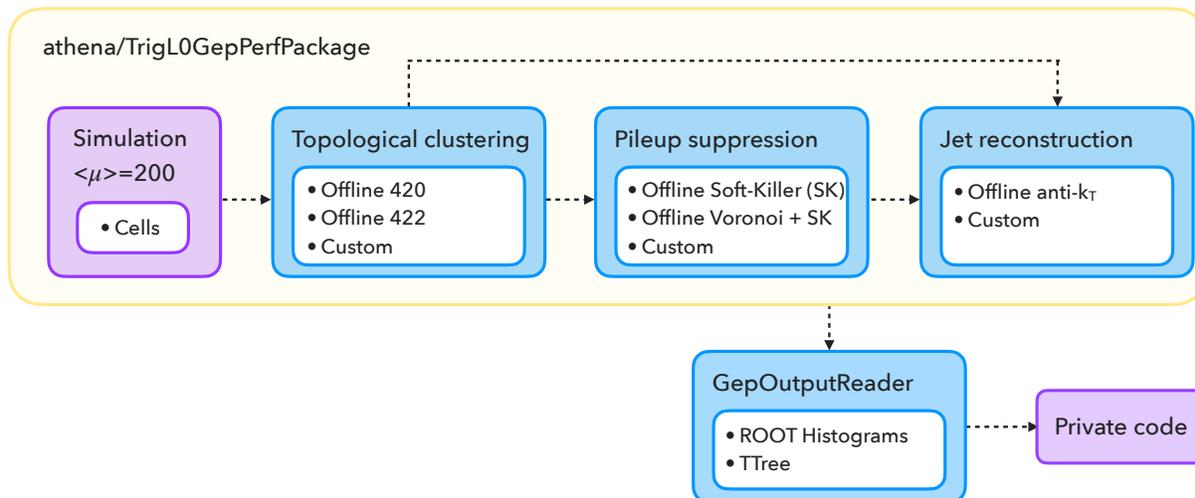


Figure 8.4: Sketch of the software simulation framework for the Global Trigger.

framework was designed to provide an efficient way to develop and study new candidate algorithms for the GT, as well as to provide Level-0 objects as input to physics performance studies for the HL-LHC. The first point was particularly important at the time of this work, as all GT algorithms were under development. The framework was therefore developed with a modular structure, where algorithms could be easily integrated for testing, while shielding the developer from direct interaction with the Athena software. In the diagram, the label “custom” represents a possible plug-in for developers to interface their C++ code with the Athena workflow. In addition, the ATLAS offline versions of the algorithms are always provided for a solid baseline comparison. As most algorithms require topocluster or jet information, the focus was put on the reconstruction chain from calorimeter inputs to jets, which includes topological clustering, constituent-level pileup suppression, and jet reconstruction. This covers the main algorithms in the first three SLRs. Other algorithms, such as e/γ , hadronic event reconstruction, muons and topological algorithms are expected to be added in the future.

The MC samples read by the framework and used in the following studies were produced

with the standard ATLAS MC production path, but with HL-LHC settings. The samples were simulated at a center-of-mass energy of $\sqrt{s} = 14$ TeV, including the simulation of the new detector components, and with the number of simultaneous pp interactions per bunch crossing set to $\langle\mu\rangle = 200$. This represents the extreme pileup scenario that the trigger is expected to handle. Di-jet events were generated using PYTHIA8 [192], with the NNPDF23LO set and A14 ATLAS parameter tuning. These samples were produced in slices of p_T in order to provide sufficient statistics across a wide kinematic range. The first slice, simulated for truth jet p_T in the range $[0, 20]$ GeV, was used as representative of a minimum bias sample, while the remaining slices were taken as representative of QCD multi-jet events. For single-jet trigger studies, a simulation of fully-hadronic $Z' \rightarrow t\bar{t}$ events was used. The sample was generated using POWHEG [198] interfaced with Pythia with the NNPDF23LO set and the A14 tune for the parton shower. For multi-jet trigger studies, di-Higgs events, with each Higgs boson decaying to a pair of b -quarks, were used. The samples were produced for gluon-gluon fusion (ggF) Higgs boson production and assuming the SM trilinear coupling. The matrix element was calculated using POWHEG at NLO, including finite top mass loop calculations. The showering was performed using PYTHIA8.

The primary input objects to the framework are calorimeter cells output by the digitization step of the ATLAS MC simulation path. The cells are fed as input to the topological clustering algorithm. The standard ATLAS offline topoclustering algorithm (see Sec. 4.4.4) provides topoclusters at the EM scale built with the 4-2-0 setting (`Calo420`). The same algorithm is also run with the 4-2-2 setting (`Calo422`), which does not include the outer layer of cells with $|\mathcal{S}| > 0$, resulting in smaller topoclusters. A collection with 4-2-0 topoclusters LCW calibrated (`CaloCal`) is also provided for offline large-R jet reconstruction. As a form of noise suppression, as well as to reduce the bandwidth and the processing time of downstream algorithms, the GT will only receive cells with $E > 2\sigma$. Therefore, the `Calo422` offline collection represents the “best case scenario” for topoclustering at Level-0. As the online topoclustering algorithm for the GT is still under development, the `Calo422` collection was used as representative of the GT topoclusters in the studies presented here.

Next, the topoclusters can be pileup suppressed using the recently developed constituent-level methods described in Sec. 5.5. Both the Soft-Killer (`SK`) and the Voronoi (`Vor`) offline algorithms are provided, including the option of combining them by running `SK` on the Voronoi-subtracted topoclusters (`VorSK`). The modified topocluster collection can then be fed to downstream algorithms, such as jet reconstruction. In ATLAS internal studies, it was shown that running the Voronoi algorithm alone was not as effective. For this reason, in the following studies only the `SK` and `VorSK` options are discussed. No online version of these algorithms was available at the time of these studies.

Lastly, the resulting topocluster collection is fed to the jet reconstruction algorithm. The offline anti- k_t algorithm (see Sec. 5.2) can be run with the choice of $R = 0.4$ or $R = 1.0$. The jet collection produced by running the offline anti- k_t algorithm on the offline Calo422 collection represents the “best case scenario” for small- R jet reconstruction in the GT. The development of a new jet reconstruction strategy for the GT was a major part of this work and will be discussed in detail in the next sections.

8.2.2 Developing a jet trigger

As mentioned in the introduction, a one-jet trigger targets signatures with one energetic small- R jet by applying a p_T threshold on the leading small- R jet in the event. Similarly, multi-jet triggers require the presence of three or four jets above a p_T threshold, where the higher jet multiplicity requirement typically allows to lower the p_T threshold. In the following, the cut on the online jet p_T is referred to as p_T^{cut} .

The performance of a jet trigger is studied with MC simulations in terms of the offline signal efficiency. As shown in Fig. 8.5, the efficiency is analyzed as a function of the offline-

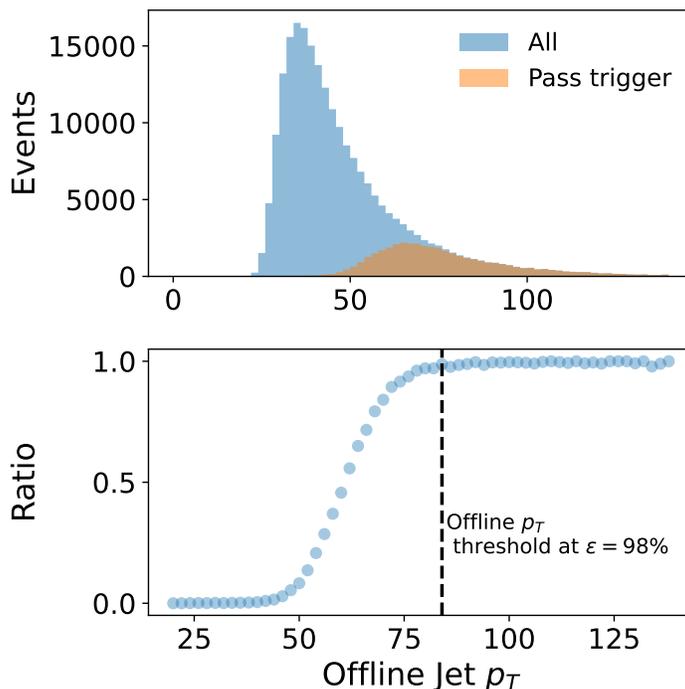


Figure 8.5: Example of trigger efficiency curve (see text for explanation).

version of the variable used to apply the online selection cut. In the case of jets, if the trigger selects on the n^{th} leading online jet p_T , the efficiency is displayed as a function of the n^{th} leading offline jet p_T . This visualization is important, as the offline analysis will only

accept events with the n^{th} leading offline jet p_T above the 98% efficiency point (the dashed line in Fig. 8.5). This ensures that the trigger selection does not bias the MC simulation in unpredictable ways. In the following, we will refer to the offline p_T value at the point where the turn-on curve reaches 98% efficiency as p_T^{thresh} . In practice, the lower the p_T^{thresh} , the broader the p_T range covered by the analysis. Hence, improving the trigger performance means reducing the p_T^{thresh} . Intuitively, this can be obtained by lowering the online p_T^{cut} by means of reducing the rates of high energy background jets. However, it also depends on the online reconstruction performance with respect to offline reconstruction, as explained more in detail in the following.

Recall that offline reconstruction represents the best one can do when virtually no latency or resources limitations are present. Then, one could say that the ideal online algorithm is the one that reconstructs the same objects as the offline one. Consider first the ideal scenario of perfect online reconstruction, with a one-to-one correspondence between online and offline objects. In this case, the efficiency curve would look like a step function at $p_T^{\text{cut}} = p_T^{\text{thresh}}$, with zero efficiency for offline p_T values below the online p_T cut and 100% efficiency for values above. Mistakes in online reconstruction result in a deviation from this scenario. One can identify two ways in which online jet reconstruction can go wrong: one can reconstruct the correct jet, but with an incorrect energy; or one can reconstruct the incorrect jet.

The first case occurs when the online and offline jets represent the same energy deposition in the detector (they have the same (η, ϕ) coordinates), but have different reconstructed transverse momenta. This can happen because offline and online reconstruction use different inputs and possibly different algorithms. For instance, while offline small-R jets are produced using the Calo420 topocluster collection, online jets in the GT are produced using Calo422 topoclusters, which are by construction lower in energy, as they are built from fewer cells. Therefore, one can expect online jets to have lower transverse momenta than offline jets. Similarly, a cone algorithm without any overlap removal strategy will, on average, produce more energetic jets than an algorithm that takes care of removing any energy double-counting, such as anti- k_t .

If the transverse momentum of every online jet differs from the corresponding offline jet by the same factor, this is just a matter of normalization that does not affect the value of p_T^{thresh} (i.e. the same set of events would pass the trigger). However, issues arise when the factor is not constant, but dependent on the phase space. As an example, consider the case of two nearby jets. A cone algorithm might find the two true jets and draw cones around them, but the cones might overlap. In this case, energy is being double-counted and the two jets might see their energy increase. Because nearby jets are characteristic of more boosted scenarios, the cone algorithm would produce an artificial increase in the number of high

energy jets, which would be more pronounced for more boosted signatures. The preferential increase in the rate of high energy jets can determine a higher online p_T^{cut} and cause different events to pass the trigger. An event with well isolated jets, whose online p_T was just above the threshold and did not get augmented, will now fail to pass the higher trigger threshold, resulting in a loss of offline efficiency and hence a higher p_T^{thresh} value. Similarly, a non-zero efficiency below the p_T^{thresh} value could also occur, as online jets that should have not passed the trigger can now have their energy increased sufficiently to pass the selection.

The second case occurs when the reconstructed online and offline n^{th} leading jets correspond to different objects. This can often happen for a seeded algorithm, where a choice has to be made on how close the seeds can be. Assuming no constraint on how close two seeds can be, it is possible that the same offline jet is reconstructed twice. Consider an event with two high energy offline jets. Online reconstruction should produce two high energy online jets and soft 3rd and 4th leading jets. However, if two seeds are found around the leading jet, two online jets will be reconstructed with the energy of the leading jet. The result will be three high energy online jets in the event. This means that the event will most likely pass the three-jet trigger, as the online p_T cut meant for the 3rd leading jet is really being applied to the leading jet. This will result in a non-zero efficiency at a p_T value below p_T^{thresh} .

Consider instead the case where seeds are required to be at a minimum distance of $dR = 0.5$ from each other. Consider then a signal topology where two pairs of nearby jets are created (e.g. $hh \rightarrow b\bar{b}b\bar{b}$) and where, for certain percentage of events, at least one pair is closer than $dR = 0.5$. While offline jet reconstruction will always find four jets, for this subset of events the online cone algorithm will not be able to reconstruct the 4th jet, resulting in an inefficient four-jet trigger. For signal samples characterized by this topology, this can produce an offline plateau inefficiency.

Summed over all events, these contributions change the step-function into a turn-on curve, as the one shown in Fig. 8.5. In practice, mismatches between online and offline objects do not affect the performance of the trigger, as long as the event passes the selection anyway. This is why no jet truth-matching is performed when looking at the signal efficiency curves. However, as the examples described above demonstrate, these errors can have undesirable consequences, as they can result in a higher p_T^{thresh} to be adopted by the analysis. Any efficiency below p_T^{thresh} is also undesirable, as it represents an inefficiency from the point of view of the TDAQ system, which utilizes time and resources to process events that are not usable by most analyses.

The development and optimization of a trigger algorithm has therefore two goals:

1. Reduce the rate of high energy background jets, while keeping the signal efficiency

high, i.e. reduce p_T^{thresh} as much as possible.

2. Improve the jet energy resolution of the online vs. offline jet reconstruction, i.e. make the turn on as steep as possible.

Both can be studied by comparing trigger efficiency curves produced by different algorithms. The first point can be studied by comparing trigger efficiency curves at a fixed trigger rate: a maximum trigger rate is assumed, the online p_T cut is found that allows to remain below the given rate threshold, and the offline trigger efficiency is built. The second point can be studied by adjusting the online p_T threshold of each algorithm so that the offline p_T thresholds are aligned. This decouples the problem from the choice of online p_T cut, and allows to study only the jet energy resolution, where the algorithm with the best resolution is the one with the fastest slope of increase. Since the resolution can be p_T dependent, different choices of offline p_T threshold allow one to test the resolution in different phase spaces.

Clearly, the performance of a trigger algorithm closely relies on the jet reconstruction strategy, with the development and optimization of the two aspects being closely intertwined. In the following, a new strategy for jet reconstruction and triggering for the GT is presented, where the two components are treated as one individual task with the shared goal of improving the trigger efficiency curve.

8.3 A cone jet reconstruction algorithm

The offline anti- k_t algorithm (see Sec. 5.2.3) is the optimal choice for jet reconstruction. Nevertheless, it is also a computationally intensive algorithm, necessitating the calculation of dR distances and execution of $1/p_T^2$ divisions for each iteration. Additionally, anti- k_t is a highly iterative algorithm, with a non-deterministic number of operations. This makes it non-scalable on parallelizable firmware, thereby losing one of the main advantages of fast FPGA hardware. Consequently, its deployment on the Global Event Processor would require a substantial allocation of resources, making it an impractical, if not prohibitively expensive, choice. The work presented here aimed to find an alternative to anti- k_t that would allow to retain the necessary physics performance, while requiring fewer resources.

The main advantages of anti- k_t are IRC safety and its ability to correctly identify the boundary between nearby jets. At the Level-0 trigger, IRC safety is not a requirement, as the goal of the trigger is to accept the right events, not to reconstruct physics-analysis-ready objects. If the event passes the Level-0 trigger selection, the objects are reconstructed again with the offline algorithms first at the Event Filter level, and then offline for use in analyses.

The ability to identify nearby jets is instead highly desirable, if not necessary in order to retain high signal efficiencies for multi-jet triggers.

The jet reconstruction strategy proposed in this work is based on the other class of algorithms: cone algorithms (see Sec. 5.2.2). While the choice of seeding strategy is not straightforward, seeding provides a fixed handle on the number of computations. It also makes jet building highly parallelizable, as each seeded jet can be built independently. The studies presented here focus on small-R jet reconstruction, where the radius parameter is fixed to the standard $R = 0.4$. Different possible extensions of this work to large-R jet reconstruction are envisioned, including using the leading small-R cone jets as seeds for another iteration of the cone algorithm with a larger radius parameter, or as input to a reclustering algorithm. However, this goes beyond the scope of this work.

In the performance studies shown in this section the reference offline jet collection is produced by running offline anti- k_t on `Calo420`³ topoclusters at the EM scale. This will be referred to as `AntiKt420`. The inputs to jet reconstruction in the GT will be the `Calo422` topoclusters. It follows that the upper bound on jet reconstruction performance at the GT level is set by the anti- k_t algorithm run on `Calo422` topoclusters. The development of the cone algorithm is therefore benchmarked against this jet collection, which will be referred to as `AntiKt422`. All selected jets in these studies are required to be in the central region of the detector $|\eta| < 2.5$ and to have $p_T > 10$ GeV.

The performance of the algorithm was benchmarked against target signal simulations. The $Z' \rightarrow t\bar{t}$ sample was used as the representative signal for one-jet trigger studies. It also provided an event topology useful to study the effect of nearby jets for multi-jet triggers. The $hh \rightarrow b\bar{b}b\bar{b}$ sample was used as representative for signals relying on three- and four-jet triggers. In particular, the event topology is characterized by well separated low energy jets, making this type of signature particularly sensitive to the online p_T threshold.

The studies were performed for the most part assuming a fixed rate threshold. In these studies this will be assumed to be 60 kHz for a one-jet trigger and 50 kHz for three- and four-jet triggers, as it was assumed in the Phase II Upgrade Technical Design Report for the TDAQ system [10]. Some studies are performed at fixed offline threshold, in order to compare the energy resolution of different jet collections.

In Sec. 8.3.1 the development process of the new cone jet algorithm is discussed. In Sec. 8.3.2 the performance of the optimized version of the algorithm is benchmarked against the online `AntiKt422` jet collection. The efficacy of offline pileup suppression schemes will also be evaluated. Lastly, Sec. 8.3.4 discusses the results of a preliminary firmware simulation,

³See Sec. 8.2.1 for an overview of the object collections available in the framework.

providing insights into the practical implementation of the algorithm.

8.3.1 Development

Recall that a jet definition is defined by the jet algorithm, the jet inputs, and the recombination strategy (see Sec. 5.2). These choices are discussed in the following.

Inputs

As mentioned earlier, the inputs to jet reconstruction in the GT will be 4-2-2 topoclusters, which, for the purpose of these studies, will be the `Calo422` collection. Pileup suppression is expected to be applied on the topoclusters before these are provided to the jet reconstruction process. However, the jet algorithm development was performed without pileup suppression applied, as no online pileup suppression algorithm was available. This approach also simplified the optimization process by reducing the number of factors involved.

In the absence of any E_T cut on the topoclusters, the number of topoclusters reconstructed per event is several hundreds, as shown in Fig. 8.6. Processing such a large number of inputs for each event is unfeasible, making thresholding the input topoclusters necessary. Additionally, the maximum number of inputs that can be processed must be predetermined in firmware, and the lower this limit, the smaller the algorithm's footprint on the FPGA. However, as this choice depends on the resources available in the firmware, the study of this trade-off was left for future work and only the E_T thresholding was studied here. The E_T cuts at 1, 2, and 3 GeV were considered, motivated by the range of the constituent-level p_T cut produced by the Soft-Killer algorithm reported in Ref. [154]. The effect of the different cuts on the jet energy resolution was studied in different energy regimes using the `AntiKt422` online jet collection. Fig. 8.7 and Fig. 8.8 show the trigger efficiencies for the di-Higgs signal with the offline p_T threshold fixed at 50 GeV and 100 GeV, respectively. The energy resolution worsens with increasing topocluster E_T cut, and the effect is greater for regions of phase space with higher p_T jets. It follows that the choice of the E_T threshold is a trade-off between reducing the number of inputs and keeping the energy resolution high. Note that the performance of the `AntiKt422` jets with no E_T cut on the topoclusters, where the only difference between offline and online reconstruction is the topocluster collection, shows the effect on the jet energy resolution from using the `Calo422` instead of the `Calo420` topocluster collection.

The conclusion from these studies was to exclude the 3 GeV cut as a viable option. The final choice between 1 GeV and a 2 GeV cuts will likely be determined by the available resources on the firmware.

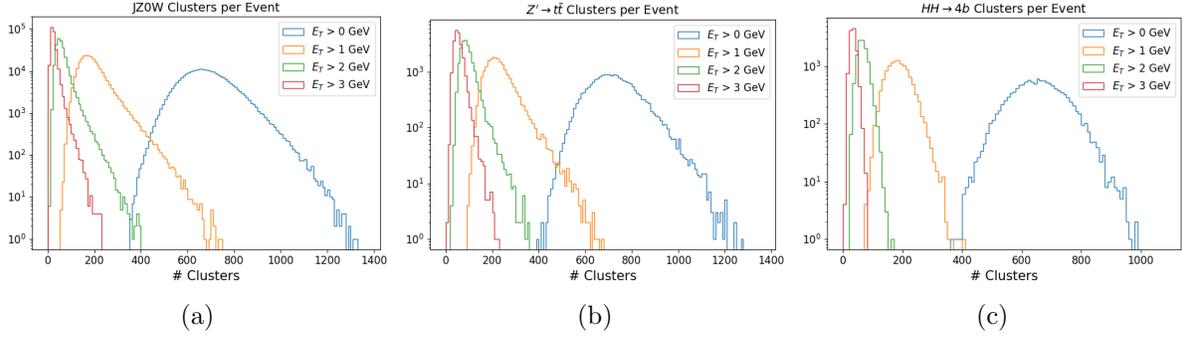


Figure 8.6: Number of topoclusters per event passing different values of E_T thresholding for minimum bias (left), Z' (center) and di-Higgs (right) samples. Plots made by Garrit.

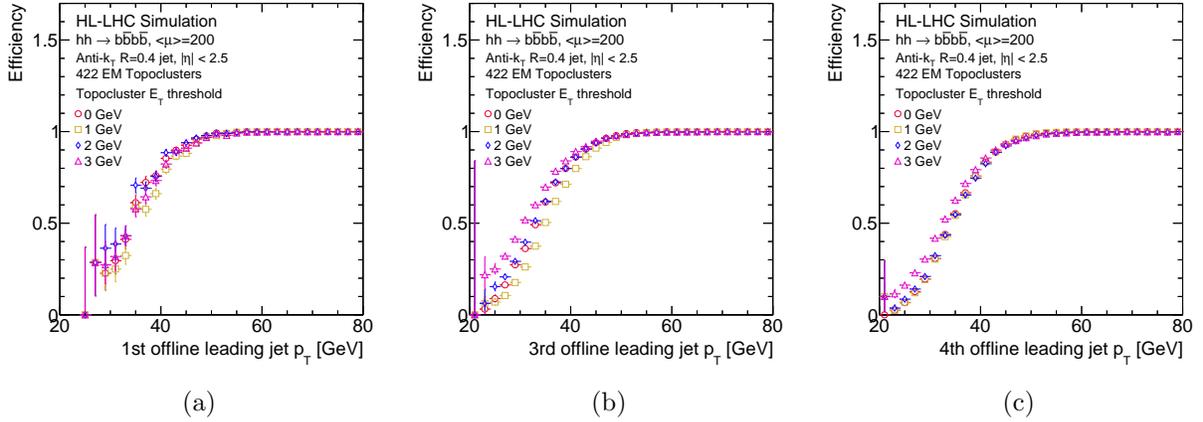


Figure 8.7: One (left), three (center), and four (right) jet trigger efficiency curves built with the `AntiKt422` online jet collection and a fixed offline p_T threshold of 50 GeV using di-Higgs events. The jets are reconstructed using input topoclusters with different E_T thresholding applied: 1, 2, and 3 GeV thresholds and no threshold.

Recombination scheme

The most commonly used recombination strategy is the E -scheme, where the jet four-vector is given by the sum of the four-vector components of its constituents. In the first stages of development, this was the recombination scheme used, again as a way to keep the interacting factors in the optimization process low. Once the algorithm development was mature and a first firmware simulation was performed, the E -scheme computations involving trigonometric functions were observed to increase the FPGA resource utilization beyond acceptable limits. A new scheme was therefore designed, which will be referred to as the “Approximate- E_T scheme”. In this scheme, the transverse energy of the jet is given by the scalar sum of the transverse energy of its constituents, the (η, ϕ) coordinates are given

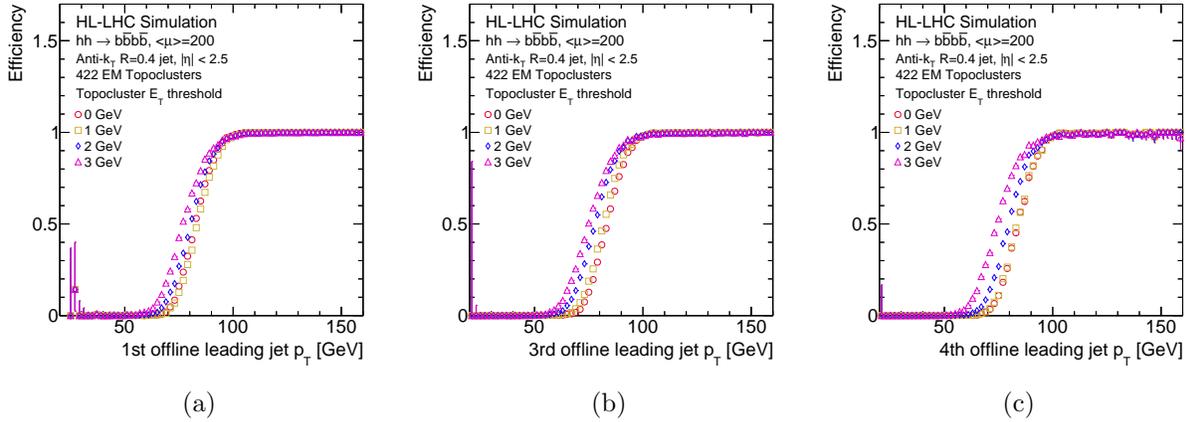


Figure 8.8: One (left), three (center), and four (right) jet trigger efficiency curves built with the `AntiKt422` online jet collection and a fixed offline p_T threshold of 100 GeV using di-Higgs events. The jets are reconstructed using input topoclusters with different E_T thresholding applied: 1, 2, and 3 GeV thresholds and no threshold.

by the coordinates of the seed, and the jet is assumed massless:

$$E_T^{\text{jet}} = \sum_{i \in \text{Clusters}} E_T^i, \quad \eta^{\text{jet}} = \eta^{\text{seed}}, \quad \phi^{\text{jet}} = \phi^{\text{seed}}, \quad m^{\text{jet}} = 0. \quad (8.1)$$

The approximation in (η, ϕ) space assumes the jet axis to be aligned with the seed. The effect is that of a loss of resolution on the location of the jet, which was observed to be more significant the lower the transverse energy of the seed. However, the resulting jet is typically still contained within an area in (η, ϕ) of 0.1×0.1 . The computation of the jet E_T assumes the jet constituents' four-vectors to be aligned. Similarly, this assumption degrades for lower energy jets. Nonetheless, both these approximations were shown to have a negligible effect on the trigger performance.

Energy reconstruction

A cone algorithm needs to be seeded to keep the number of computations under control. However, choosing the appropriate objects to act as seeds and determining the criteria to down-select them were complex challenges. As discussed here, these decisions have multiple interconnected implications. Before tackling this issue, the viability of a cone algorithm was tested by using the best seeds available, the offline anti- k_t jets themselves. As discussed in the previous section, cone jet reconstruction can go wrong in two ways: it can reconstruct the incorrect jet, or it can use the right seed, but reconstruct the incorrect energy. Using offline anti- k_t jets as seeds allowed to decouple the two issues and focus on the latter. Fig. 8.9

shows the comparison of four-jet trigger efficiencies between `AntiKt422` jets and `Cone422` jets seeded with offline anti- k_t jets. These studies are shown for the $Z' \rightarrow t\bar{t}$ signal sample, where the 3rd and 4th leading jets are often close to the 1st and 2nd. Cone jets show a clear over-efficiency below threshold. When separating the signal sample into isolated and non-isolated jets, as shown in Fig. 8.9b and 8.9c respectively, this over-efficiency was shown to originate from the subsets of events with non-isolated online jets. A similar behavior is observed for the three-jet trigger. This study demonstrated the need for an energy overlap removal strategy.

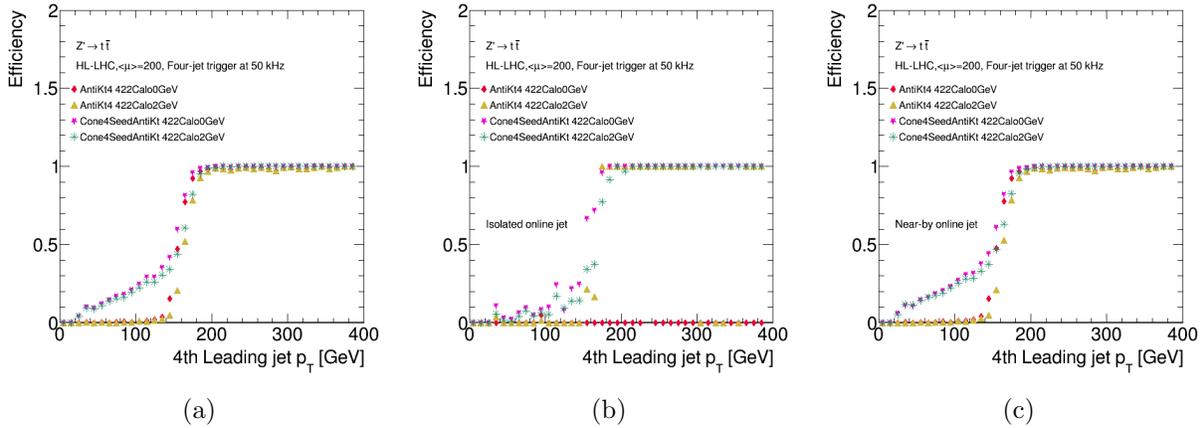


Figure 8.9: Four-jet trigger efficiencies for $Z' \rightarrow t\bar{t}$ signal sample. The online jet collections compared are `AntiKt422` and cone jets seeded with offline `AntiKt420` jets. Fig. (a) shows the inclusive performance, while Figs. (b) and (c) show the performance for events with isolated and non-isolated offline jets, respectively.

Seeding

The seeding strategy is determined by the type of seeds used and by the seed selection process. Obvious candidates for seeding were the trigger objects (TOBs) output by the `jFEX` algorithm (described in Sec. 4.3.4), as these will be provided to the GT by `L0Calo`. However, the minimum dR distance requirement between seeds of the `jFEX` sliding window algorithm was observed to cause a constant inefficiency for multi-jet triggers and signal samples with nearby jets. This comparison is shown in App. 9. Therefore, the next choice was to use the topoclusters themselves. In order to avoid the same issue encountered with `jFEX` TOBs, no minimum distance requirement was imposed between the seeds, with the understanding that this would exacerbate energy overlap and jet double counting. Both issues were solved with an energy overlap removal step, as discussed later.

Similarly to the maximum number of inputs, the maximum number of seeds that can be

processed for any event has to be fixed in firmware. E_T thresholding is therefore required to select interesting topoclusters to be used as seeds and, once the list of thresholded topoclusters is formed, only the N highest E_T seeds should be used to build jets. From now on, when referring to seeds it will be assumed that it is the list of topoclusters with E_T above the given seed E_T threshold. The number of potential seeds per event as a function of possible E_T thresholds is shown in Fig. 8.10. The choice of the maximum number of seeds is another important trade-off, as a smaller number reduces the computations, but can cause some reconstruction inefficiency. The final choice will depend on the available firmware resources, as well as on the number of jets per event required by the topological algorithms downstream. For this reason, the question of the maximum number of seeds was left for future studies, and only the seed E_T threshold was considered here.

Different seed thresholds were studied, from 5 GeV up to 30 GeV. The results showed that seed thresholds of 10 GeV or above would worsen the jet energy resolution and cause inefficiencies, while 5 GeV and 7.5 GeV cuts performed equally well without introducing any inefficiency. For the studies presented here, a seed threshold of 5 GeV was used.

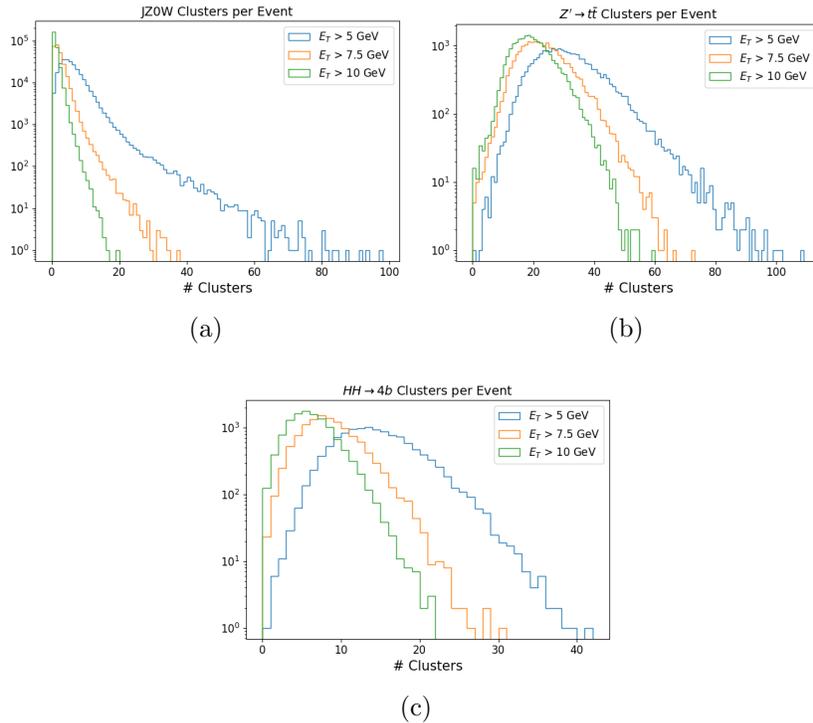


Figure 8.10: Number of topoclusters per event passing different values of E_T thresholding for minimum bias (left), Z' (center) and di-Higgs (right) samples. Plots made by Garrit.

Energy overlap removal

An energy overlap removal strategy was necessary to remove energy and jet double counting. Two strategies were developed. The first strategy, referred to as *seed removal* (SR), aims at removing the possibility of a seed to be a constituent of more than one jet. The list of sorted seeds is parsed in order of decreasing E_T , starting from the leading seed. If any other seed in the list is found within a radius parameter $dR = 0.4$, the lower energy seed is removed. Effectively, this strategy removes the possibility of severe overlap, as jets cannot be closer than $dR = 0.4$. However, it still allows partial overlap for seeds at distances $0.4 < dR < 0.8$. Note that because the seed removal is done in order of decreasing E_T , it is still possible for a jet to contain a topocluster with higher E_T than its seed, if this topocluster was removed from the seed list because of a nearby seed with even higher energy.

The second approach, referred to as *energy overlap removal* (EOR), removes any possible energy sharing with a winner-take-all strategy. After the topoclusters have been assigned to the seeds, if any topocluster belongs to more than one seed, it gets assigned only to the highest energy seed. The jets transverse energies are calculated only afterwards, to avoid repeating the computations. This procedure is similar to the CMS Cone algorithm described in Sec. 5.2.2, but without the iterative step for finding a stable cone. The effect is that of removing any possible overlap between jets. Note that in the case of seeds closer than $dR = 0.4$, one of the two jets remains seedless. This was observed to have no effect in terms of trigger performance, but a combination of seed merging with $dR = 0.4$ and EOR avoids this possibility.

The effect of these strategies on the trigger rates and trigger efficiencies is shown in Fig. 8.11 and Figs. 8.12 and 8.13, respectively. First, note that the one-jet trigger is not significantly affected by the energy double counting. On the other hand, in the case of three- and four-jet triggers, in the absence of an overlap removal strategy (red circles) the rate of high energy background jets is artificially increased, causing the online p_T threshold to be substantially higher. For a signal sample like $Z' \rightarrow t\bar{t}$, where the 1st and 2nd leading jets are significantly higher in energy than the 3rd and 4th, jet double counting allows all the events to pass the 3rd and 4th jet triggers despite the increase in rates. In the di-Higgs signal sample, on the other hand, one expects four well separated leading jets in the same low energy regime. As the signal jets are isolated enough to not see their energy augmented, while having transverse momenta close to the turn-on region, this signature is sensitive to the online p_T^{cut} . The increase in rates and the resulting higher online p_T^{cut} cause a severe loss of trigger efficiency. Some energy double-counting also occurs, producing an over-efficiency below the offline 100% efficiency threshold. Clearly, an energy overlap removal is necessary to keep the rates low and retain sensitivity for this type of signal. Seed removal with $dR = 0.4$

removes the possibility of double counting jets. For the di-Higgs signal, the seed removal strategy is sufficient to remove any over-efficiency. In the Z' sample, some over-efficiency remains due to nearby jets still sharing some energy. The EOR strategy removes any possible overlap and therefore any over-efficiency. The performance is identical for EOR and EOR with seed removal to avoid seedless jets. No strategy is observed to cause inefficiencies. In the following, the EOR strategy will be used.

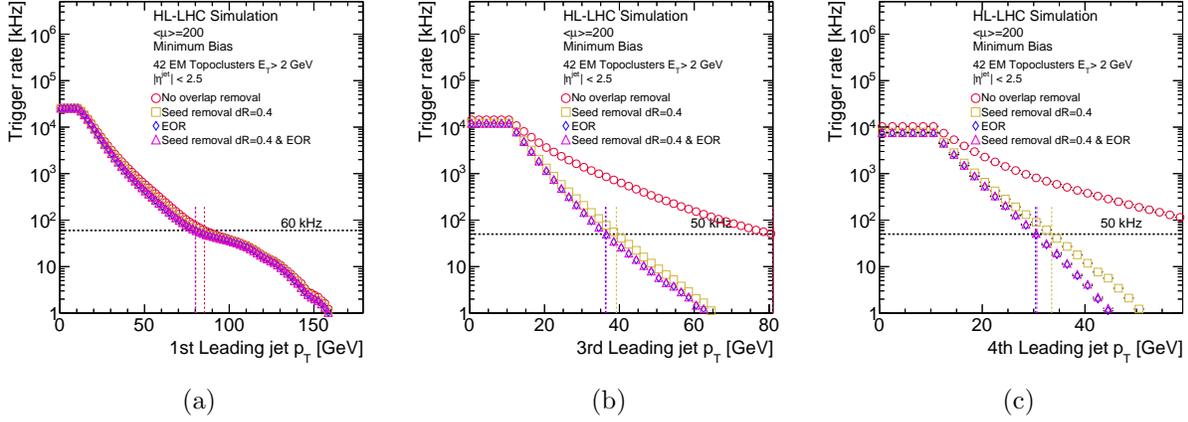


Figure 8.11: One-jet and multi-jet trigger rates using cone jets with different overlap removal strategies.

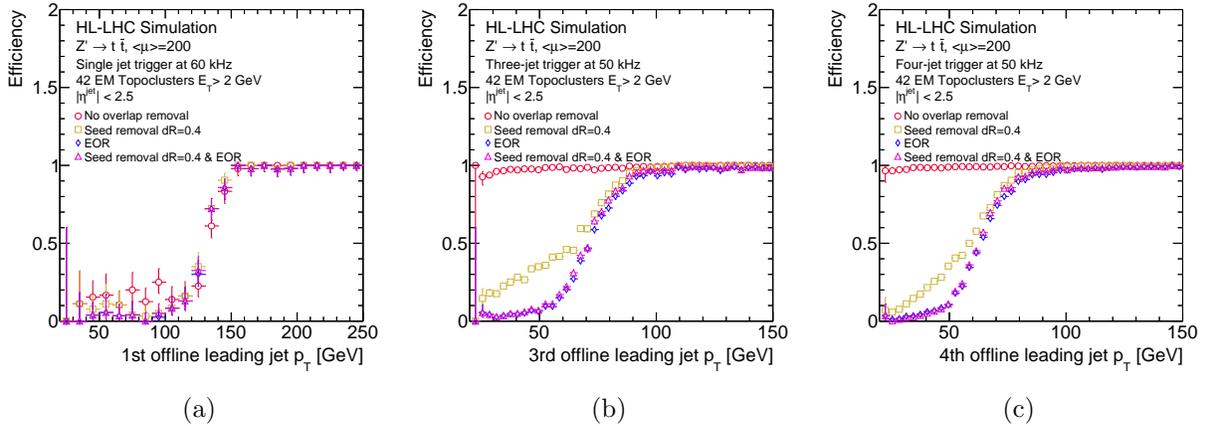


Figure 8.12: One-jet and multi-jet trigger efficiencies for $Z' \rightarrow t\bar{t}$ signal using cone jets with different overlap removal strategies.

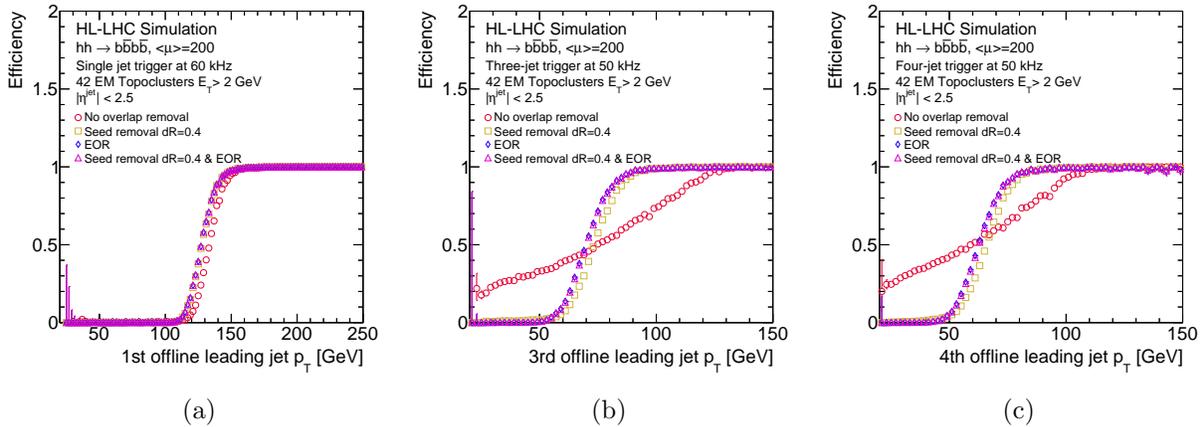


Figure 8.13: One-jet and multi-jet trigger efficiencies for di-Higgs signal using cone jets with different overlap removal strategies.

8.3.2 Physics performance

This section presents the performance of the optimized cone algorithm. The final cone jet collection, referred to as `ConeTopo`, is reconstructed with the following settings:

- **Inputs:** Topoclusters with $E_T > 2$ GeV.
- **Seeds:** Topoclusters with $E_T > 5$ GeV.
- **Recombination scheme:** Approximate- E_T .
- **Overlap removal:** EOR.

The performance is benchmarked against the online jet collection `Antikt422`, which represents the upper limit on performance when using 4-2-2 topoclusters as inputs. The same E_T threshold is applied to select the input list of topoclusters for both algorithms. Similar results are observed for $E_T > 1$ GeV and $E_T > 3$ GeV topocluster thresholds.

Trigger efficiencies

Fig. 8.14 shows the trigger rates for the minimum bias sample. The rates are identical for `ConeTopo` and `Antikt422` jets. Fig. 8.15 and 8.16 show the trigger efficiencies for $Z' \rightarrow t\bar{t}$ and $hh \rightarrow b\bar{b}b\bar{b}$ signals, respectively. Again, the performance is almost identical. The only observable difference is a slight over-efficiency for multi-jet triggers with the $Z' \rightarrow t\bar{t}$ sample. This over-efficiency is not due to nearby jets, as it does not go away after requiring offline jets to be at a distance $dR > 0.8$. It could be due to cone jets over-estimating the jet area of lower energy jets.

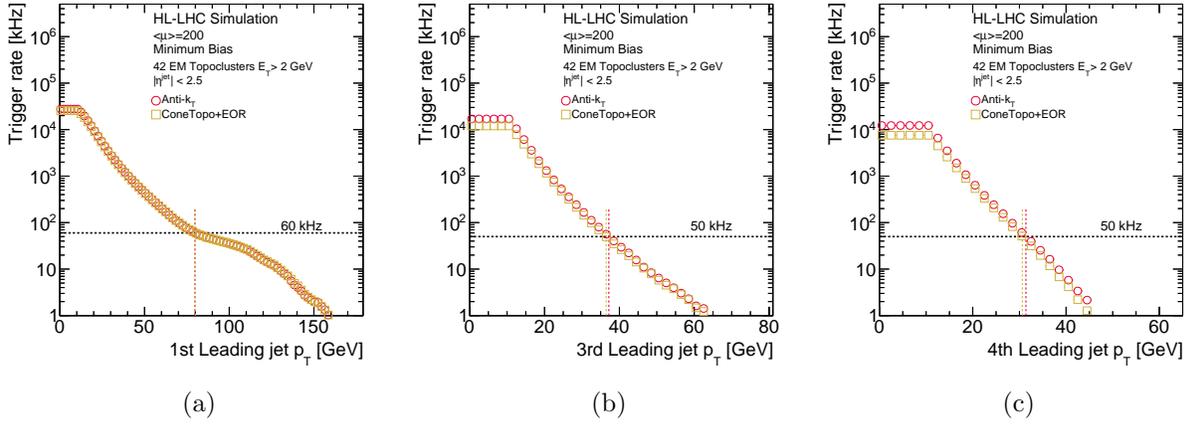


Figure 8.14: One-jet and multi-jet trigger rates comparing ConeTopo and AntiKt422 online jet collections.

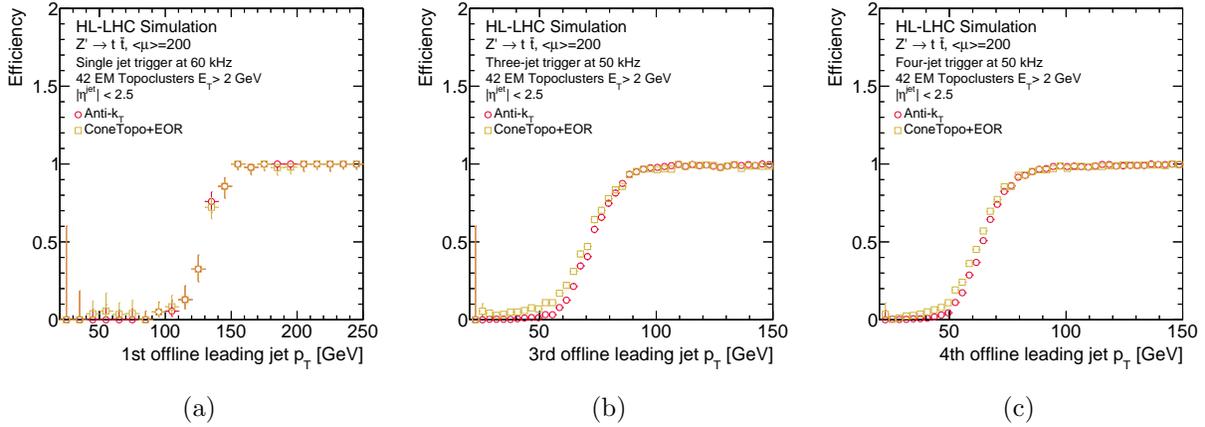


Figure 8.15: One-jet and multi-jet trigger efficiencies for $Z' \rightarrow t\bar{t}$ signal comparing ConeTopo and AntiKt422 online jet collection.

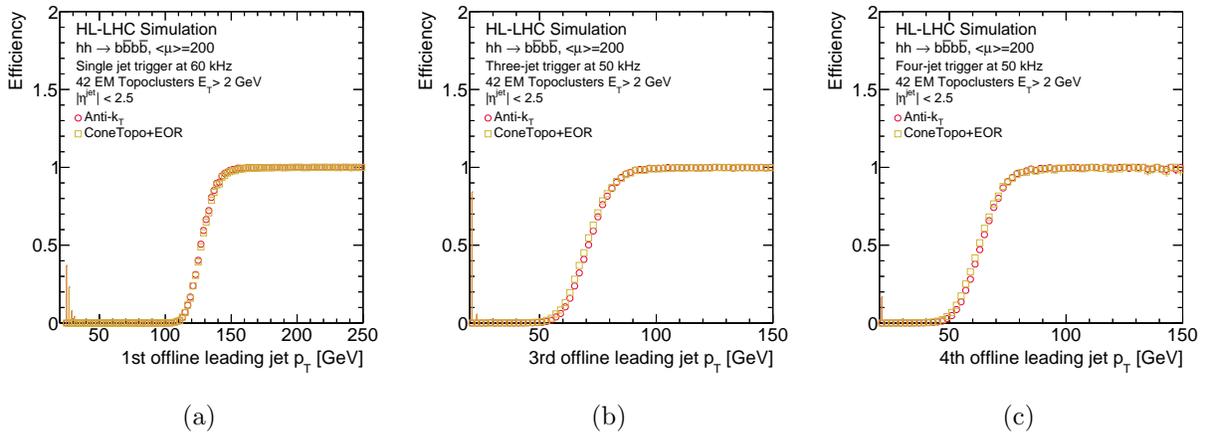


Figure 8.16: One-jet and multi-jet trigger efficiencies for di-Higgs signal comparing ConeTopo and AntiKt422 online jet collection.

Di-Higgs m_{HH} signal efficiency

After examining the trigger performance, the impact of the trigger selection on a metric more closely related to the offline analysis was assessed. In particular, it was important to check that the cone algorithm was not introducing unexpected bias, for instance, through the energy overlap removal strategy. This was studied in the context of the di-Higgs analysis. An important metric for the di-Higgs analysis is the reconstructed invariant mass of the di-Higgs (m_{hh}) system, as only the low m_{hh} region is sensitive to the value of k_λ (see Sec. 8.4). A full set of calibrations and b-tagging was not available for the HL-LHC MC samples used in this study. For this reason, the study was performed using truth information and the final hh state was reconstructed by finding the four reconstructed small-R jets truth matched to the four truth b -quarks using the procedure described in Sec. 5.3.2 with $dR = 0.35$. Only events with all four b -quarks within $|\eta| < 2.5$ and four truth matched reconstructed jets are retained. The signal efficiency as a function of m_{hh} is studied for events that pass a three- and a four-jet trigger, comparing the performance when using online `ConeTopo` and `AntiKt422` jets. As shown in Fig. 8.17, no difference in performance is observed.

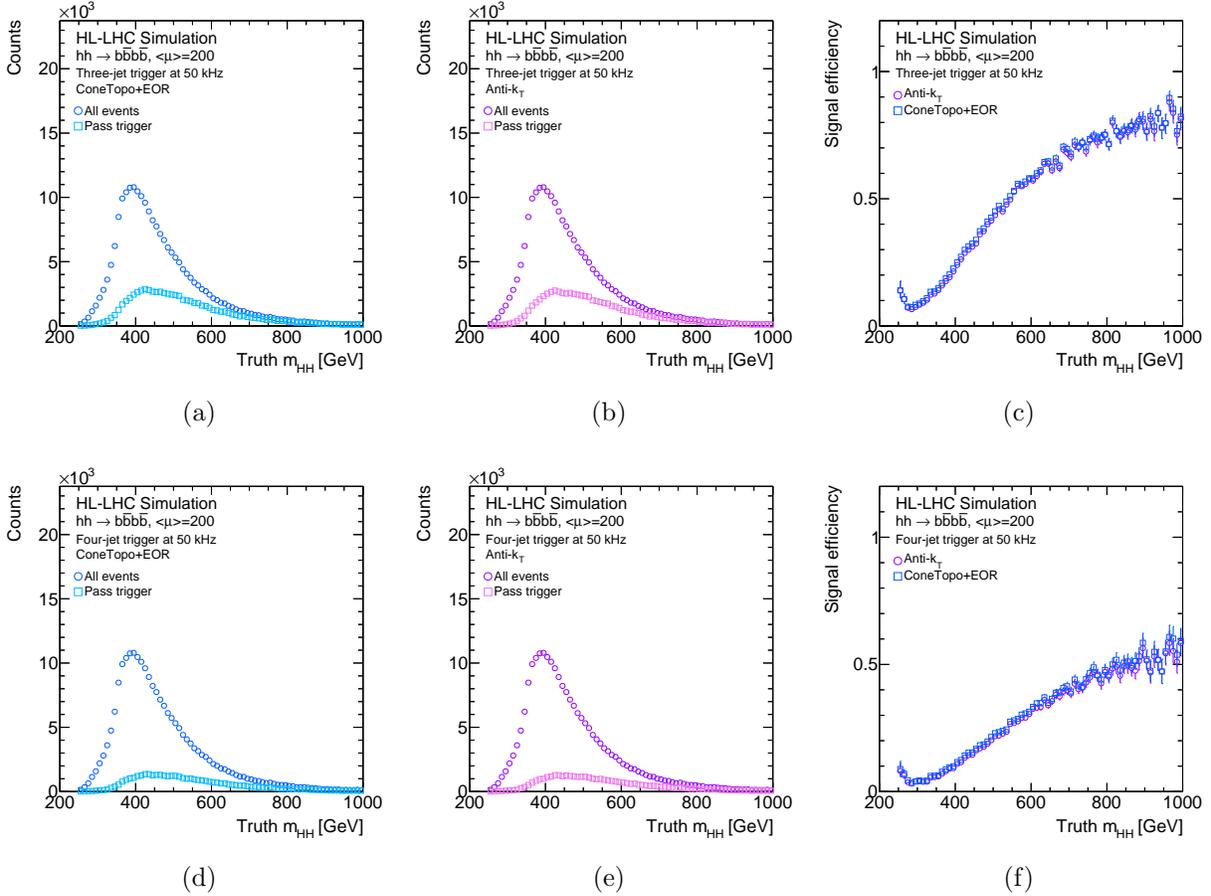


Figure 8.17: Efficiency as a function of reconstructed m_{hh} after passing the three-jet (top row) and four-jet (bottom row) trigger comparing the performance of ConeTopo and AntiKt422 online jet collections.

8.3.3 Constituent-level pileup suppression

As mentioned previously, ongoing work within the GT effort is exploring the feasibility of an online implementation of the Soft-Killer (SK) and Voronoi+Soft-Killer (VorSK) algorithms (see Sec. 5.5 for an overview of these methods). This section examines the robustness of cone jet reconstruction when the offline version of these pileup suppression algorithms is applied to the input topoclusters.

The SK algorithm removes from the event topoclusters that have a transverse energy below an event-dependent threshold indicative of the hadronic activity. The Voronoi algorithm applies a correction factor to the topoclusters transverse energy that is also event-dependent. The effect of Voronoi is that of reducing the topocluster energy, while SK actually removes the topoclusters from consideration. By reducing the energy of the constituents and/or removing candidate constituents, one expects the energy of all jets to decrease on average. This automatically decreases the online p_T^{cut} . However, as mentioned earlier, this is less relevant when the energy decrease affects all jets uniformly, as in this case the effect accounts as a simple normalization difference that will not change the offline p_T threshold. Looking at the trigger efficiencies is therefore important to understand the real effect of reducing the rates.

The n^{th} jet trigger rates and efficiencies are compared using the different pileup suppression methods. The results are shown for the di-Higgs signal sample, as this is a signature significantly affected by the presence of pileup, but similar results were observed for the $Z' \rightarrow t\bar{t}$ signal. The ConeTopo jet reconstruction is used, but identical conclusions were obtained with online AntiKt422 jets. The list of topoclusters input to the cone jet algorithm is thresholded using 1, 2, and 3 GeV cuts. The corresponding results are shown in Figs. 8.18, 8.19, and 8.20, respectively.

As expected, applying pileup suppression on the topoclusters reduces the rate of high energy background jets, as all jets will have on average fewer constituents. The effect is more significant for VorSK pileup suppression, because the jets have not just fewer constituents, but also less energetic ones. The total number of jets (obtained from the first bin) can also decrease, as seeds can be removed. This occurs more often when the lowest $E_T = 1$ GeV threshold is used, as SK rarely removes topoclusters with $E_T \geq 2$ GeV. The signal efficiencies show that SK can bring a significant improvement to the offline analysis sensitivity by reducing the offline p_T threshold. However, this positive effect diminishes with increasing topocluster E_T threshold. The effect of applying Voronoi before running SK is instead negligible. This might indicate that Voronoi is not sufficiently discriminating between signal- and background-induced topoclusters, and the same reduction factor is applied to all.

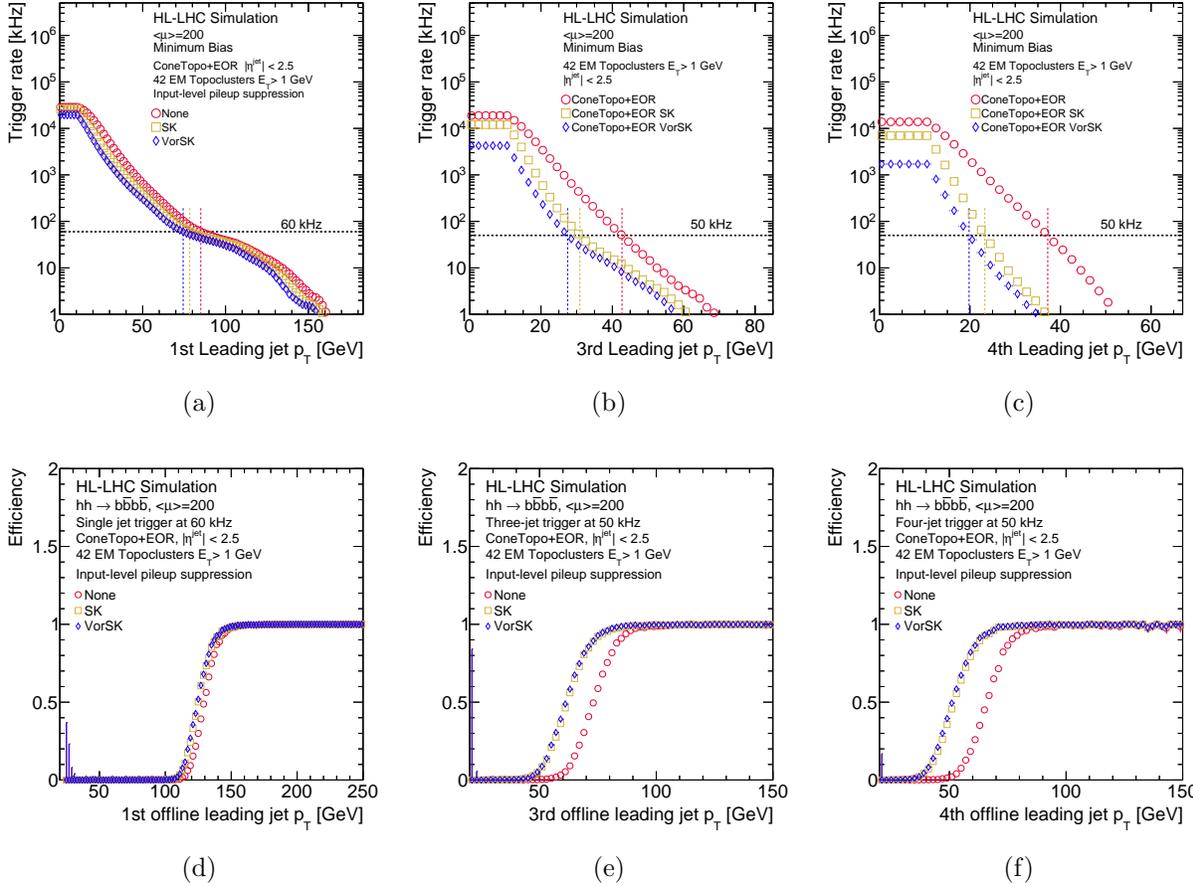


Figure 8.18: Compare trigger rates (top) and trigger efficiencies (bottom) using different constituent-level pileup suppression methods, as well as no pileup suppression. The turn-on curves are built using online p_T cuts derived from the corresponding fixed-rate threshold. The efficiencies are shown for the di-Higgs signal sample and the ConeTopo online jet collection built from input topoclusters with $E_T > 1$ GeV.

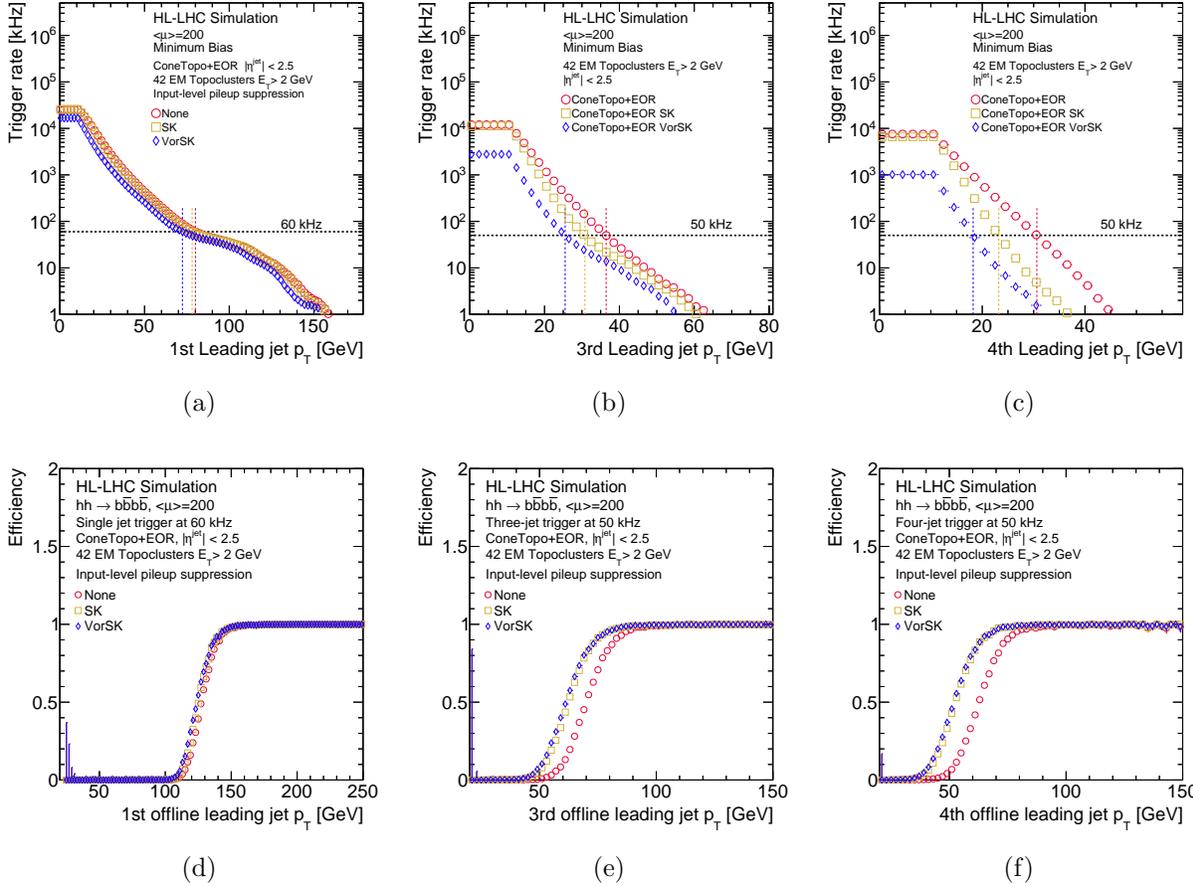


Figure 8.19: Trigger rates (top) and efficiencies (bottom) using different constituent-level pileup suppression methods, as well as no pileup suppression. The turn-on curves are built with online p_T cuts at fixed-rate threshold. The efficiencies are shown for the di-Higgs signal sample and the ConeTopo online jet collection built from input topoclusters with $E_T > 2$ GeV.

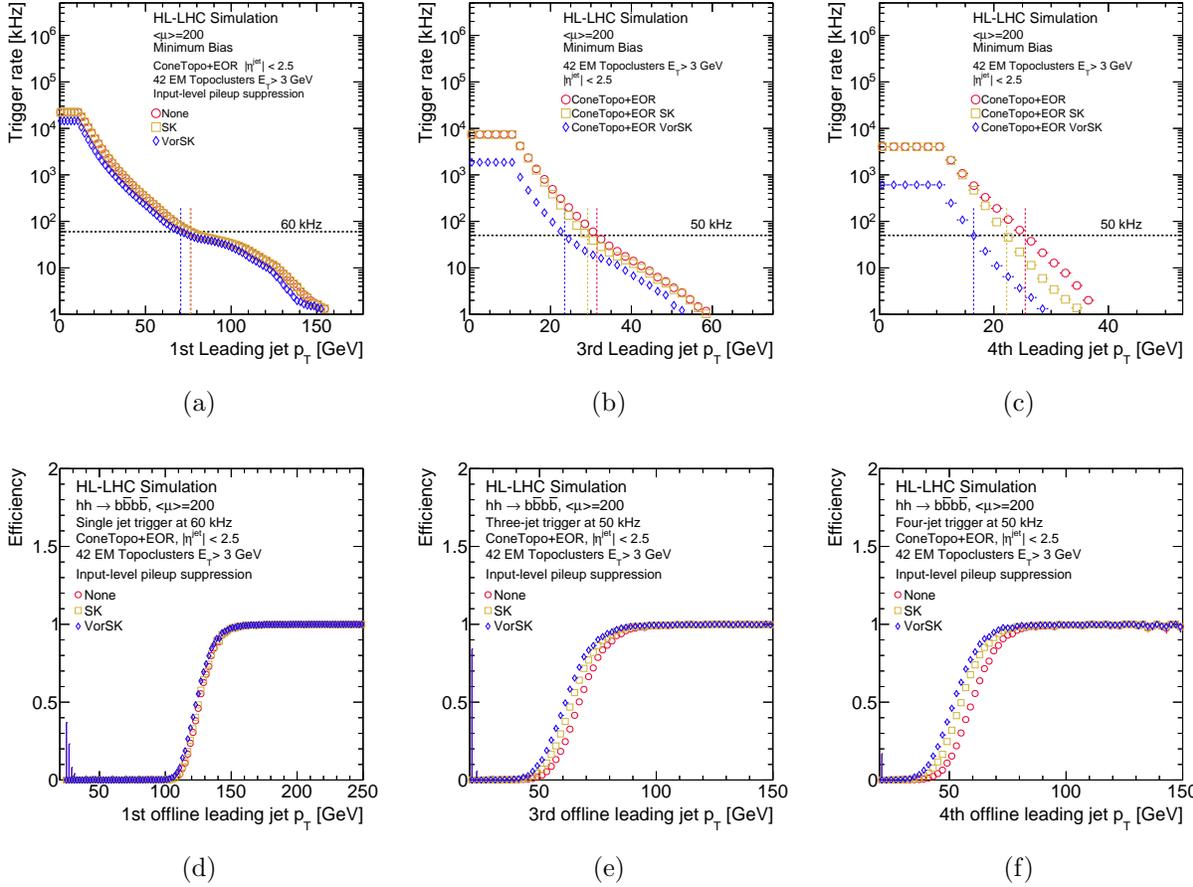


Figure 8.20: Trigger rates (top) and efficiencies (bottom) using different constituent-level pileup suppression methods, as well as no pileup suppression. The turn-on curves are built with online p_T cuts at fixed-rate threshold. The efficiencies are shown for the di-Higgs signal sample and the ConeTopo online jet collection built from input topoclusters with $E_T > 3$ GeV.

8.3.4 Preliminary firmware simulation

A preliminary firmware simulation of the cone jet algorithm was performed in order to get an estimate of FPGA resource utilization. As seen in the previous sections, this simulation informed the development process.

The simulation was performed using the Vitis High-Level Synthesis (HLS) software suite provided by the company Advanced Micro Devices (AMD), the same company manufacturing the FPGA boards that will be used in the GT. The package provides a software developer friendly interface to synthesize hardware code starting from a software algorithm written in C/C++. For the purpose of this study, HLS was used to have a fast turn-around for an estimate of the utilization of resources. However, this workflow would similarly speed up future optimization work in terms of throughput, power, and latency. The workflow to go from the software version of the algorithm, to the register-transfer level (RTL) abstraction used in hardware description languages, is the following:

1. **Software algorithm.** Develop the algorithm in software. If not already the case, provide a version written in C. In practice, the main requirement is to avoid dynamic memory allocation. As the set of resources is fixed on the FPGA, dynamic creation and freeing of memory cannot be implemented.
2. **Testbench.** Provide a set of input samples for testing. For each sample, provide the expected output for validating the results. Write a *testbench* that reads the input test files, runs the software algorithm, and compares the output with the expected result. This will be used to check that the C function is functionally correct prior to synthesis. It is also used to verify the RTL output.
3. **C-simulation.** The testbench is used to compile and execute the C simulation and validate that the C design of the algorithm produces the expected output.
4. **Synthesis:** Synthesize the C algorithm in an RTL implementation. Vitis HLS will effectively compile the C code into hardware description language. Both VHDL and Verilog are provided.
5. **C-RTL Cosimulation.:** Use the C testbench to validate the RTL design and to confirm that the hardware implementation produces the same output as the C-level code.
6. **Analysis.** Fine-tune the hardware design with code directives. Produce different RTL versions and analyze the designs by looking at the reports of the resource utilization, latency, and throughput.

7. **IP-Block.** Export RTL design as an IP block that can be integrated into the hardware.

In the studies presented here, steps 1–5 were performed. The software algorithm to be analyzed was the cone jet algorithm described in the previous section, with a few differences. The energy overlap removal step was not included in the simulation. It was also assumed that the input list of topoclusters would be provided already sorted by E_T . This is a reasonable assumption, as several downstream algorithms using topoclusters as inputs might require some type of sorting, motivating centralizing this step in the first SLRs. The C version of the algorithm was provided assuming the non-optimized values of `N_MAX_SEED` = 10 and `N_MAX_CONSTITUENT` = 40.

To perform the simulation, the user has to input the clock period, the clock uncertainty, and the FPGA target. The target clock period was set to the GT rate of 240 MHz, or 4.17 ns. The uncertainty was left as the default one, taken to be 12.5% of the clock period. The target device was left as the default target device, a *Virtex UltraScale+* FPGA, as Vitis did not offer the target device for the GT.

The clock latency is the latency of a clock cycle. Therefore, the latency of the algorithm is defined as the number of cycles it takes to produce the output multiplied by the clock latency. When synthesizing the hardware implementation, HLS determines which operations occur during each clock cycle according to the target clock frequency and the time it takes for the operation to complete on the target device. This is referred to as *scheduling*. The next step is *binding*, where the software organizes the scheduled operations onto the chip, determining which hardware resources will implement them. Once the sequence of operations is finalized, it is extracted as an RTL design, which is analyzed to obtain the performance and resource estimates. The resource utilization is examined in terms of the following resources: the number of digital signal processors (DSPs), specialized units for multiplication and arithmetic; the number of look-up tables (LUTs), units for logic and storage functionalities; and the number Flip-Flops (FFs), binary shift registers used to register data in time with the clock.

The result of the synthesis of the cone algorithm was the following. The estimate of the fastest achievable clock frequency was 3.027 ns, while the latency of the algorithm was 378 cycles. This gives a latency of 1.144 μ s, which is within the GT requirements. The resources used to implement the design are shown in Tab. 8.1, both in terms of absolute numbers and of fraction of resources of a single SLR on a VP1802, the target device for the GT. The estimates are close to the expected FPGA budget for jet reconstruction, currently set at 20% of one SLR. Note that the simulation was performed using floating point data-types, but a more accurate and conservative estimate should be obtained with fixed-point precision.

The simulation served to understand roadblocks in the algorithm design. For instance, earlier results prompted the investigation of a different recombination scheme to reduce the

	DSP	FF	LUT
Absolute numbers	681	131719	81755
% VP1802 single SLR resources	19	7.8	9.7

Table 8.1: FPGA resources utilization of the cone jet algorithm from a preliminary firmware simulation.

FPGA footprint. Ultimately, the purpose of this study was to confirm that the algorithm could be a viable option for the GT. The optimization of the firmware was deferred, as this required a final version of the cone jet algorithm, which will only be possible once a more mature picture of the other algorithms on the Global Event Processor will be available. Nonetheless, these preliminary results were indeed promising.

8.4 Pileup-jet rejection with neural networks

The cone jet reconstruction algorithm described in the previous section was shown to have equivalent performance to the offline anti- k_t algorithm run on online topological clusters (`AntiKt422`). This was already a significant result, as the `AntiKt422` collection represents the upper bound on jet reconstruction performance in the GT. Nevertheless, the question remained of whether this performance sufficed to meet the physics objectives of the experiment. If not, it would become crucial to find new avenues to further improve the algorithm’s effectiveness.

One of the main difficulties foreseen at the HL-LHC is the extreme pileup environment, with up to 200 secondary interactions per bunch crossing. As described in Sec. 5.5, pileup can significantly impact object reconstruction. In the case of jets, pileup introduces a positive bias in the reconstructed transverse momentum and causes an overall resolution degradation of the reconstructed kinematic quantities. While these effects were discussed in the context of offline reconstruction, they similarly affect the online reconstructed jets, with additional consequences. On one side, the positive bias has the effect of artificially increasing the rate of high energy background jets, which, as discussed in the previous section, pushes the acceptance thresholds to higher p_T values. On the other, the smearing due to pileup fluctuations worsens the online jet p_T resolution, making the turn-on curve less steep. In addition, pileup represents a source of noise in the event reconstruction and identification process, further complicating the trigger selection task and requiring more sophisticated trigger algorithms to retain the same signal-to-background discrimination power. Clearly, pileup mitigation is an important factor in hadronic triggers performance and it is particularly relevant for signals sensitive to the trigger p_T thresholds, such as $hh \rightarrow b\bar{b}b\bar{b}$.

Measurement of di-Higgs production is one of the most pressing experimental goals of the collaboration, being a direct probe of the Higgs boson self-coupling λ , which is still unmeasured (see Sec. 3.1). This is a challenging measurement, as the production cross section of a Higgs boson pair is very low, with two Higgs bosons being produced every one in a trillion events. Sensitivity to λ^4 is complicated further by a destructive interference between the two contributing diagrams, shown in Fig. 8.21. Only the low m_{hh} mass region remains sensitive to possible deviations from the SM Higgs self-coupling, an experimentally difficult phase space dominated by pileup. Due to these difficulties, the latest results from ATLAS and CMS using the full Run 2 dataset have only been able to set limits [7, 8]. However, the HL-LHC is expected to reach the ultimate sensitivity, with the current projected signal significance with (without) systematic uncertainties at 4.0σ (4.5σ) [9]. Advances in trigger, reconstruction, and analysis strategies in the incoming years could push these predictions to the level of a 5σ discovery. This unprecedented opportunity makes di-Higgs production one of the flagship signatures of the HL-LHC upgrade.

The decay channel $hh \rightarrow b\bar{b}b\bar{b}$ ($hh4b$) has one of the largest sensitivities thanks to the largest decay branching ratio of the Higgs boson into b -quarks. In the low mass non-resonant region, the $hh4b$ final state is characterized by four low energy jets, a region of phase space dominated by multi-jet background and pileup. Measurement of non-resonant $hh4b$ critically relies on multi-jet trigger thresholds and is one of the key challenges and drivers of the HL-LHC trigger upgrade, starting from the GT. In this section, a new method for mitigating the impact of pileup on the Level-0 multi-jet trigger performance targeting $hh \rightarrow b\bar{b}b\bar{b}$ signal is investigated.

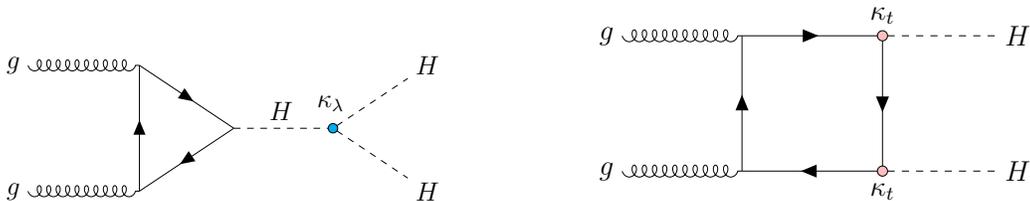


Figure 8.21: Leading order diagrams contributing to gluon-gluon fusion di-Higgs production cross section. Only the left diagram depends on the Higgs self-coupling λ .

8.4.1 Pileup jet identification

Pileup collisions are uncorrelated from the hard scatter and produce an approximately uniform distribution of low transverse momentum particles in the detector volume. When run-

⁴Results for measurements of the Higgs boson couplings are typically presented in terms of coupling modifiers. For instance, the coupling modifier of the Higgs trilinear self-coupling is given by $\kappa_\lambda = \lambda_3^{\text{Obs}}/\lambda_3^{\text{SM}}$.

ning a jet reconstruction algorithm, these low energy deposits can end up being recombined into a jet. As the pileup levels increase, the number of these soft *pileup particles* increases as well and the overlap of these low energy depositions can lead to the reconstruction of high energy topoclusters and jets. In the following, a jet whose transverse momentum is mostly due to pileup particles will be referred to as *pileup jet*, while a jet originating from a hard quark or gluon produced in the hard scatter is referred to as *signal jet*. It is precisely these high energy pileup jets that are problematic for the trigger, as they fictitiously increase the rates of high energy jets, pushing the p_T thresholds up. This is particularly relevant for multi-jet triggers. To improve the trigger performance, one would like to identify and reject high energy pileup jets before the trigger selection is applied. This is expected to reduce the background rates, which in turn allows to reduce the p_T thresholds. In this study, the use of deep learning to identify and reject pileup jets in the GT is investigated⁵.

The likelihood of a jet to be a pileup jet is determined by the amount of pileup contamination that contributes to the jet energy. Different pieces of information are typically used to identify pileup particles offline, but not all of them will be available in the GT. Tracking information is an effective tool, as it allows to determine the number of associated tracks originating from pileup vertices. However, this information will not be accessible in the ATLAS Level-0 trigger. Another good metric to identify pileup particles is that they are soft. Because pileup contributes uniformly to the event kinematics and the level of pileup fluctuates between events, how soft a particle has to be for it to be identified as pileup is event-dependent. A metric of the pileup activity in the event is necessary to make the most well-informed decision on this cutoff. This is, in fact, the strategy adopted by offline pileup suppression algorithms, such as Soft-Killer. Whether a metric of the pileup event density or Soft-Killer itself will be available in the GT is still under study. One last piece of information remains, which is the local energy and multiplicity distribution of the jet's constituents. In fact, this information will be accessible for the first time in the Level-0 trigger thanks to the ability to reconstruct topoclusters from the full granularity calorimeter information. How this information can be useful is discussed next.

Pileup particles are uniformly distributed in the detector and, when clustered into a jet, they cause a uniform smearing of the jet image. This results in distinctly different energy profiles between signal and pileup jets, as shown in Fig. 8.22. These plots were produced in the following way. The jets⁶ were built with all the topoclusters above a given E_T threshold,

⁵Pileup contamination of signal jets can also occur, causing a loss of energy resolution. Therefore, another avenue to improve the trigger performance would be to improve the jet energy resolution and make the turn-on curve steeper, which, for instance, could be implemented as a regression task of the true online jet energy. While this is an option worth investigating, this work focused only on the jet-rejection strategy.

⁶These studies were performed using jets produced with an earlier version of the cone algorithm. The jets

without pileup suppression. After the jets were formed, each jet’s constituents were compared with the Soft-Killer pileup suppressed topocluster collection to determine whether a given topocluster would have been pileup suppressed or not. The jet energy profile was then plotted by separating the contribution from the constituents that would have been pileup suppressed and the ones that would have not. This is shown for both signal (QCD)⁷ and pileup jets. The content of each bin is given by the sum of the E_T of the jet constituents at the given dR distance from the jet center. The histograms are shown in bins of reconstructed jet p_T : [15, 35], [35, 50], and [50, 70] GeV. Radiation that is deemed “pileup-like” by Soft-Killer is uniform and low in energy, producing a linearly increasing energy profile. This feature is identical for signal and pileup jets, subject to the same uniform pileup contamination. After pileup-suppressed topoclusters are removed, the energy profile of signal jets peaks close to the center of the jet and falls off rapidly at large radii, while the profile of pileup jets in the lowest p_T bin remains uniform. As the jet p_T increases, the jet energy profile of pileup jets becomes increasingly more signal-like even after pileup suppression, losing most of the discrimination power for jets with p_T above 50 GeV. Nevertheless, as the region of interest for multi-jet trigger rates falls below this threshold, local information on the jet’s constituents is a promising discriminant.

The goal of this study was therefore to determine whether the local distribution of a jet’s constituents could be sufficient to identify pileup jets. Clearly, this approach does not address the fact that the difference in energy profile is still dependent on the pileup event-density. In order to address this, the use of additional information from the output of offline particle-level pileup suppression algorithms was also investigated, representing the upper bound on the performance of this pileup jet rejection technique in the GT.

were built with input topoclusters with $E_T > 1$ GeV and with a seed removal strategy that used $dR = 0.3$. This is the cause of the upward shift in some histograms at $dR = 0.3$, where jets overlap starts being allowed.

⁷The signal jets in this study are jets reconstructed in di-jet events and truth matched to a truth quark.

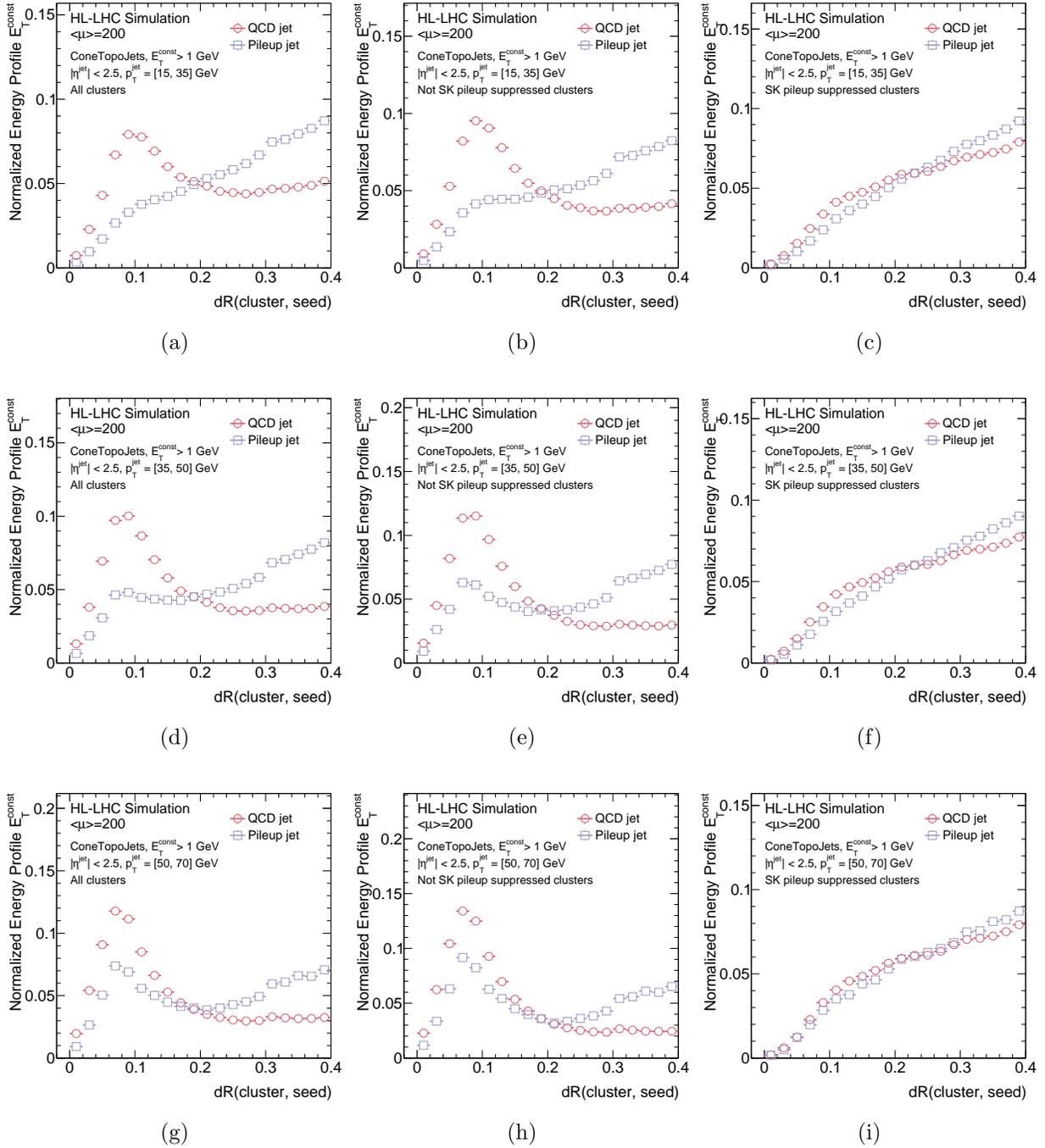


Figure 8.22: Jet energy profile of cone jets built with seed removal using $dR = 0.3$. Comparing all constituents (left), only constituents that would have not been pileup suppressed (center), and constituents that would have been pileup suppressed (right). The jets are separated into p_T bins (from top to bottom): $[15, 35]$, $[35, 50]$, and $[50, 70]$ GeV.

8.4.2 Neural network development

A deep neural network (DNN) was developed as a jet-by-jet classifier to output the probability p of a jet to be signal ($p = 1$) or pileup ($p = 0$). Two different DNNs were trained using different input variables.

Training samples

The training samples were formed starting from the multi-jet samples described in Sec. 8.2.1. The subset of pileup jets was selected from the minimum bias sample, which is generated with truth di-jet transverse momenta in the range $[0, 20]$ GeV, while requiring the jets not to be truth-matched. The subset of signal jets was obtained from the combination of several p_T slices of multi-jet samples, with truth di-jet transverse momenta up to 800 GeV. The jets were further required to be truth-matched to one of the two truth quarks. For this reason, in the following “signal jet” and “QCD jet” is used interchangeably. This choice of signal jets avoided the issue of training the networks on the same set of $hh4b$ events of interest.

The jet collection used for the training jets were **ConeTopo** jets built with topoclusters with $E_T > 1$ GeV and the seed removal strategy with $dR = 0.4$ applied. Energy overlap removal was not applied in order to preserve the circular shape of the jets and avoid confusing the network during training. For the same reason, only jets whose leading E_T constituent corresponded to the seed were accepted, to avoid the rare case of a higher energy seed removed by the seed removal strategy having entered a jet built from a lower energy seed.

The selected signal and background samples are characterized by distinctly different p_T distributions, with signal jets covering a wide p_T range, and background jets peaking at small values. To prevent the network from classifying merely based on jet p_T , different measures were implemented. First, the training was performed only in the region around the region of interest, targeting jets with p_T between 25 and 50 GeV. Jets below this threshold were excluded as they would have negligible impact on the rates determining the online trigger p_T cut. For jets above this threshold, the discrimination power in the local distribution of the jet constituents was observed to degrade, as shown in Fig. 8.22h. In addition, the samples were reweighted to have a uniform p_T distribution and balanced class normalization.

The dataset was split into training, testing, and validation subsets. After requiring the jets to have $p_T \in [25, 50]$ GeV, the number of training samples was reduced to approximately 500,000.

Input variables

A set of input variables was optimized to describe the N constituents with highest E_T . Three types of information were identified to describe each topocluster in the jet: the spatial

location in the jet’s reference frame, the transverse energy, and some metric of the likelihood of the topocluster of being a pileup-particle. Different input variables were considered for each case. In the following, the selection process is described, while the distributions are provided later with the network performance.

The number of leading topoclusters to provide to the network was fixed at $N = 10$. This was motivated by the fact that the jet constituent multiplicity distributions peak below 10 (see Fig. 9.10) and the discrimination power of the input variables between signal and background for the n^{th} leading topocluster decreases for increasing n , as will be shown in Fig. 8.25 and 8.24. The inclusion of an input variable providing the number of jet constituents was tested to compensate for this approximation, but was observed to not bring any improvement. The expectation is that an even smaller value of N might be used, but this choice was not optimized.

The location of a given constituent was provided both in terms of $\Delta\eta$ and $\Delta\phi$ distances between the constituent and the jet, as well as simply in terms of the ΔR . The additional information from the coordinates was found to not bring any improvement, so only the $\Delta R(\text{constituent}, \text{jet})$ values were provided. Note that, by construction, the coordinates of the jet are identical to the coordinates of the seed and of the leading topocluster.

The raw energy of the constituents is highly correlated to the jet p_T and providing this information to the network resulted in the classification being based almost exclusively on the jet p_T . In order to remove this dependence, but still provide a metric representative of the jet’s energy profile, the transverse energy of the constituents was normalized to the transverse energy of the leading constituent. Equivalent results were obtained by normalizing to the jet p_T , with no improvement observed by providing both.

The likelihood of a topocluster being pileup suppressed depends on the topocluster E_T in relation to the event-wide pileup density. However, it is unclear whether such a metric will be available, so two strategies were employed. The first model was trained without any information related to pileup suppression. This will be referred to as DNN-A. The second model was trained including for each jet constituent a boolean flag of whether the topocluster was suppressed by the SK algorithm, in practice providing the same information as what shown in Fig. 8.22. This will be referred to as DNN-B. The training was performed using cone jets built with non-pileup suppressed topoclusters. The performance of DNN-B is expected to be an upper bound on what can be obtained from this technique. If DNN-A is sufficient to improve the trigger performance, the conclusion is that SK is not needed in the GT. If, instead, DNN-A is not sufficient, then DNN-B allows to test whether including information from the SK algorithm can further improve the trigger performance more than running jet reconstruction on pileup suppressed topoclusters.

8.4.3 Training and performance evaluation

The DNNs were trained using Keras with TensorFlow [197] backend. The model hyperparameters were optimized using a grid-search. The same architecture was used for both networks, as no significant variations were observed by varying the hyperparameters. The final model was a deep fully-connected NN with two hidden layers with 50 nodes each and ReLU activation functions. The model had one output node with sigmoid activation function representing the probability of a jet to be signal. The NN was trained optimizing the binary cross-entropy loss with the Adam optimizer, using a learning rate of 1.e-4, 80 epochs, and a batch size of 500.

The input variables were defined to describe each of the 10 leading topoclusters in the jet. The final set of input variables were chosen to be the distance $\Delta R(\text{cl}, \text{lead cl})$ and the energy ratio $E_T^{\text{cl}}/E_T^{\text{lead cl}}$ between each topocluster and the leading topocluster in the jet. Information on the leading topocluster was removed, as by construction the leading topocluster has $\Delta R = 0$ and $E_T^{\text{cl}}/E_T^{\text{lead cl}} = 1$. In addition, a boolean pileup-suppression flag was included only for the training of DNN-B. This gives 18 input variables for DNN-A and 27 input variables for DNN-B. The training dataset was reweighted to have flat jet p_T and class distributions, and only jets with p_T in the range [25, 50] GeV were used for training. Note that no reweighting was applied to the validation and testing datasets.

The final metrics computed on the validation dataset for DNN-A and DNN-B are shown in Tab. 8.2. The *accuracy* is the percentage of correct predictions, or the sum of the true positive and true negative rates. The *recall* represents the rate of true positives: the percentage of signal samples correctly identified as signal. The *precision* is inversely proportional to the rate of false positives: the higher the precision, the larger the percentage of samples identified as signal that are true signals. The area under the ROC curve (*AUC*), or the true positive rate as a function of the true negative rate, represents the trade-off between signal efficiency and background rejection. The higher the area, the smaller the trade-off. The precision-recall curve (*PRC*) represents the trade-off between accurate positive results and relevant positive results. A high area under the PRC curve indicates low false positive and low false negative rates. The precision is equivalent between the two models, while DNN-A has a lower accuracy and recall, resulting in lower AUC and PRC. In particular, DNN-A has a harder time at accurately classifying signal jets, with a higher rate of false negatives (true signal predicted as background).

An unbiased evaluation of the model performance was obtained on the unseen testing dataset. Fig. 8.23 shows the comparison between DNN-A and DNN-B performance in terms of raw output scores, confusion matrices, and ROC curves. In particular, for the latter,

	DNN-A	DNN-B
Accuracy	0.73	0.76
Precision	0.80	0.81
Recall	0.77	0.82
AUC	0.76	0.80
PRC	0.82	0.85

Table 8.2: Training metrics computed on validation dataset for DNN-A and DNN-B. See text for explanation.

specific working points (WPs) at fixed signal efficiency are provided for comparison and for later use. As expected from the training results, DNN-A has a slightly lower performance. From the confusion matrices, which are built using the default classification score at 0.5, one can see that DNN-A has a higher rate of signal jets identified as background (23% instead of 18%). While this difference is not dramatic, one has to look at WPs relevant for the scope of the trigger, where only a minimum signal loss can be tolerated. For signal efficiencies above 85%, DNN-A shows a significantly larger background efficiency than DNN-B. As shown later, this difference will have a significant impact in trigger performance.

In Fig. 8.24 and 8.25, the input variables are shown for selected n^{th} leading topoclusters. In order to visualize what the network is learning, the distributions are shown separately for samples tagged according to the true and predicted label. Note that the predicted label is set by using a cut score of 0.5. The results are shown for DNN-B and similar results were obtained for DNN-A. The agreement between the true and predicted distributions indicates that the network is learning the true PDFs. Deviations appear mostly in the distribution of the dR between the higher energy topoclusters and the leading one, as the network learns that signal jets are more collimated and with the leading constituents carrying most of the jet transverse energy, while background jets are more diffuse and with a more even energy sharing among the constituents. Fig. 8.26 shows similar distributions for out-of-training variables. From Figs. 8.26a and 8.26b, one can see that the network learns that signal jets have, on average, fewer constituents and fewer pileup suppressed constituents. Figs. 8.26c and 8.26d show that the network is not biased by the E_T of the leading topocluster as a result of reweighting the training dataset to have a flat jet p_T distribution, and that it correctly learns the correlation between the E_T of the leading topocluster and the p_T of the jet. Lastly, Fig. 8.27 compares the true and predicted distributions in terms of the jet p_T for DNN-A and DNN-B. While the p_T reweighting mostly succeeds in removing the bias from the jet p_T , some p_T dependence remains. From these plots it is clear that the source of false negatives (true signal classified as background) identified in the previous discussion comes from low transverse momentum signal jets. The misclassification is more pronounced

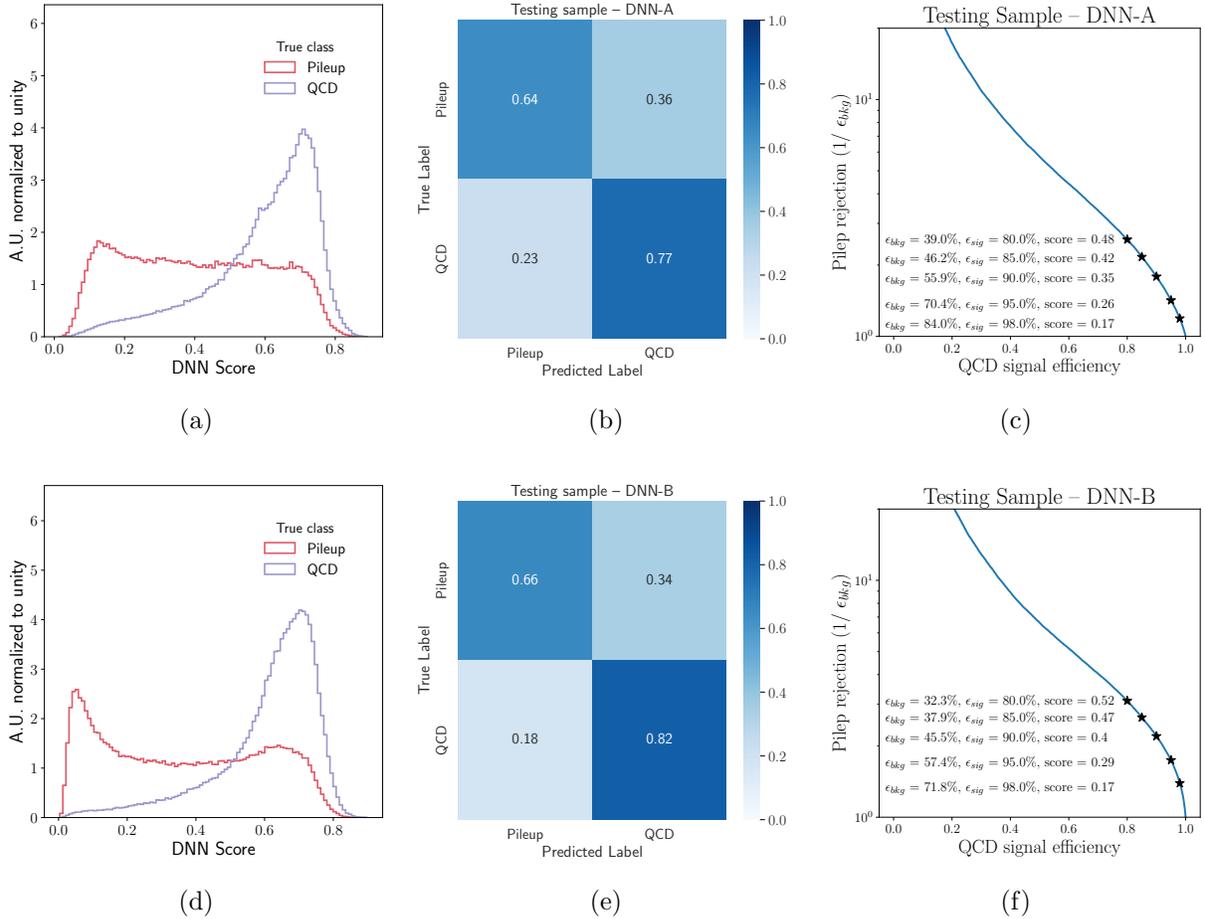


Figure 8.23: Comparison of testing performance in terms of output score (left), confusion matrix (center), and ROC curve (right) between DNN-A (top) and DNN-B (bottom).

for DNN-A, as expected. These low energy signal jets are the ones with a less collimated energy profile and without a handle on the pileup event density through the Soft-Killer flag, a greater number of these jets is wrongly identified as pileup.

8.4.4 Trigger performance

The two DNNs were deployed in the Global Trigger simulation framework to study their effect at on the trigger performance. The same `ConeTopo` jet collection used for training was used here, with the addition of energy overlap removal. However, similar results were observed for the cone jet collection without EOR, as well as for `AntiKt422` jets. In the following, a jet is said to be “in-training” if its transverse momentum is in the training range [25, 50] GeV. For each model, the procedure was as follows.

Every reconstructed jet that would normally enter the trigger workflow is passed through

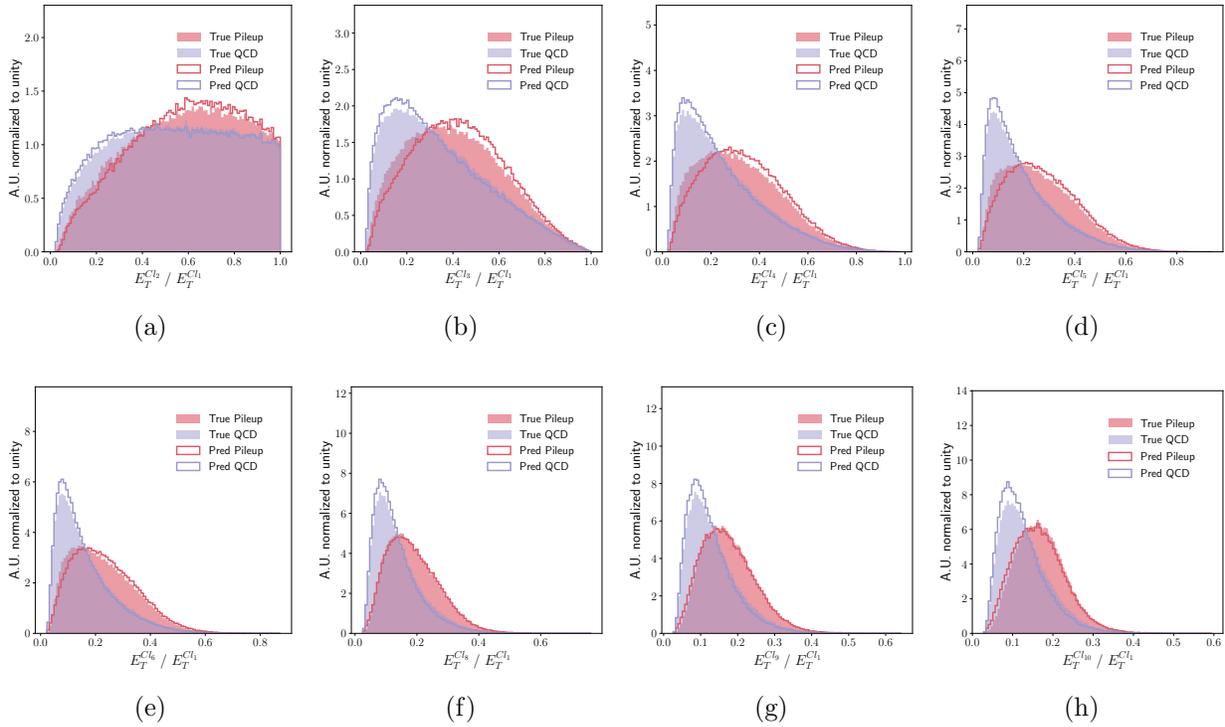


Figure 8.24: Transverse energy of the n^{th} leading constituent normalized to transverse energy of the leading constituent. The distributions are separated according to the true and predicted label.

each of the two DNNs. The scores for the four leading jets with in-training p_T are shown in Fig. 8.28 for the minimum bias, $hh4b$, and di-jet samples. Note that the DNNs correctly identify jets from the di-Higgs sample as signal-like. More confusion is present for the di-jet sample (only the $[400 - 800]$ GeV truth p_T slice is used in these studies), where the third and fourth jets are typically soft. Note that the two leading jets in the di-jet sample have very few statistics, as they populate higher p_T bins.

Next, a cut on the score at a fixed signal efficiency WP (according to Figs. 8.23c and 8.23f) is applied to all jets that enter the study. Each WP produces a new jet collection of “DNN pileup suppressed” jets, where only jets that have a score above the given cut are retained. The “baseline” jet collection without any DNN selection is also shown for comparison. The p_T spectra of the fourth leading jet after the DNN selection at different WPs are shown in Fig. 8.29 and 8.30 for DNN-A and DNN-B, respectively. Notably, these plots show that to retain a high enough signal efficiency for low p_T jets, the background rejection of DNN-A is significantly reduced, while for DNN-B more than 50% of background jets are rejected across the full p_T range. This has direct consequences on the rates.

The trigger rates are built with the minimum bias sample for the different jet collections.

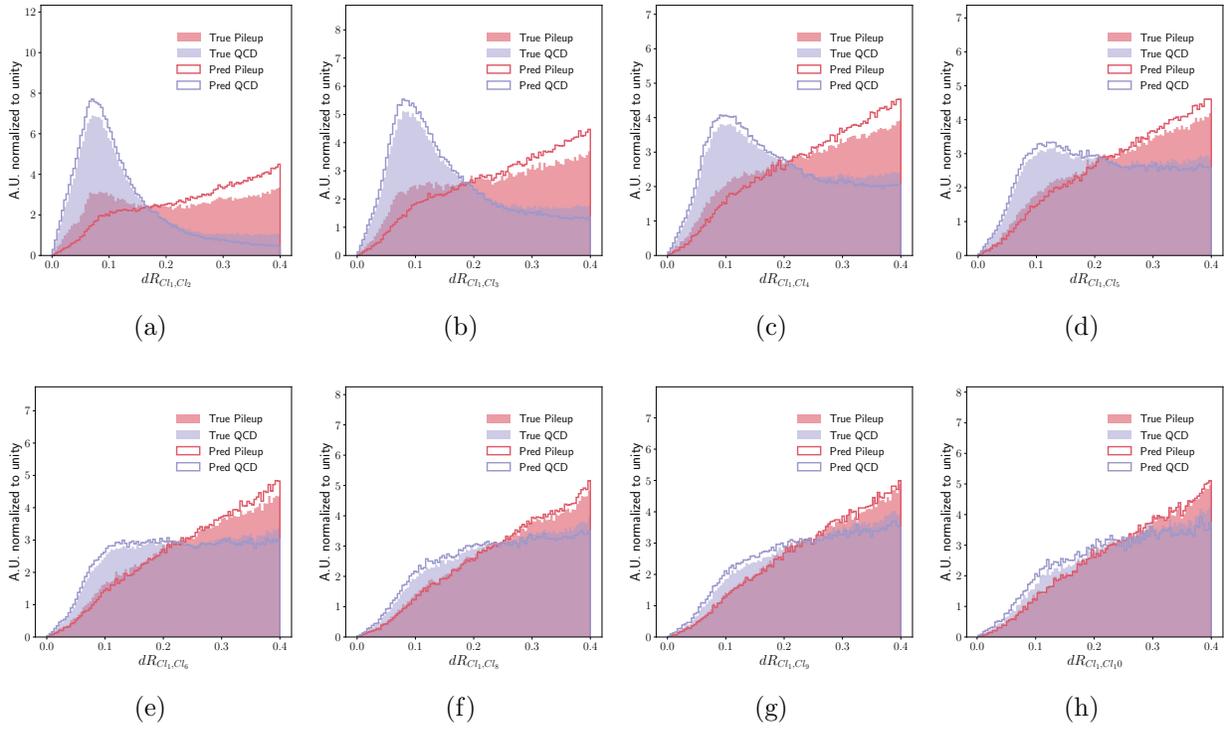


Figure 8.25: Distance ΔR between the n^{th} leading constituent and jet. The distributions are separated according to the true and predicted label.

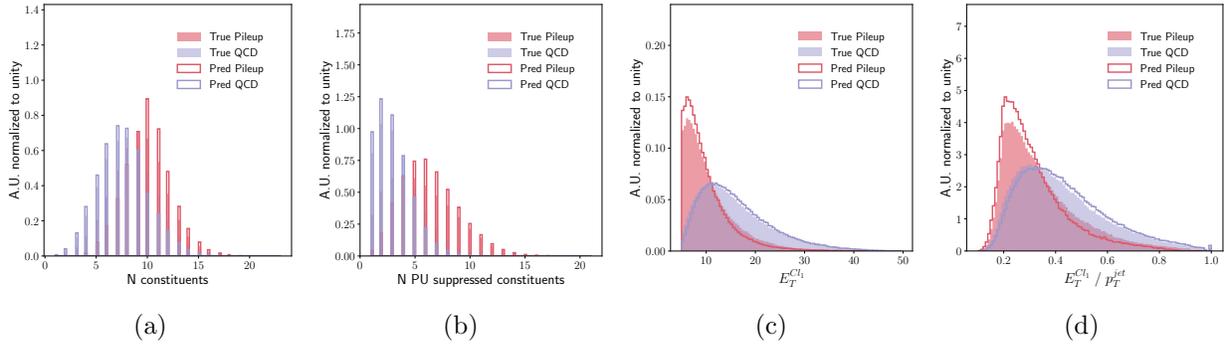


Figure 8.26: Selected out-of-training variables. The distributions are separated according to the true and predicted label.

These are shown in Fig. 8.31 for both models. Note that two strategies are compared in these plots. One strategy applies the DNN cut on all jets, while the other only on in-training jets. As these plots show, applying the DNN cut on all jets has a negligible impact on the rates, mostly because very few minimum bias events have fourth-leading jets with p_T above 50 GeV. It was therefore decided to only apply the DNN cut on jets with $p_T < 50$ GeV in order to retain the maximum signal efficiency. As expected, the lower the WP signal

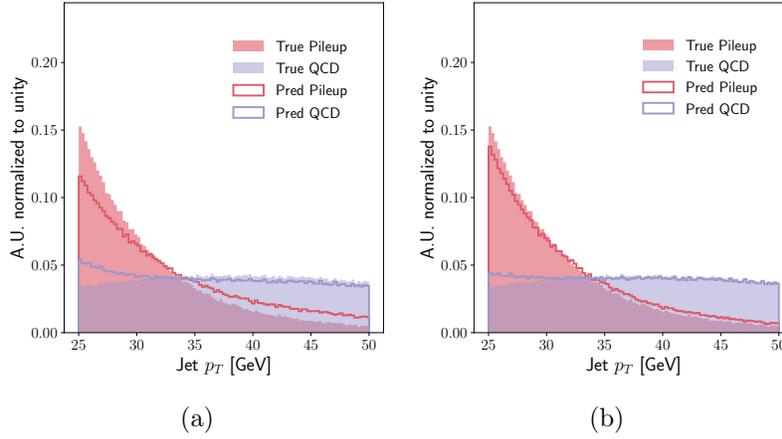


Figure 8.27: Out-of-training jet p_T distribution tagged according to true and predicted label for DNN-A (left) and DNN-B (right).

efficiency, the more background jets are rejected, and hence the larger the decrease in high energy pileup jets. Because DNN-B reaches a higher background rejection at fixed signal efficiency, it can obtain lower online p_T cuts.

Lastly, the effect on the trigger efficiencies was studied. Fig. 8.32 shows the trigger efficiencies comparing different WPs for both models. Increasing the background rejection with tighter WPs is observed to worsen the resolution while not improving the 100% efficiency threshold, so the 95% WP was selected for both models as the best performing option. Fig. 8.33 shows the final comparison of the best WPs for DNN-A and DNN-B. For comparison, the plot includes also the trigger efficiencies obtained by running the same cone jet reconstruction algorithm without DNN cut on pileup suppressed topoclusters using both Soft-Killer alone and Voronoi+Soft-Killer. The lower background rejection of DNN-A, which keeps the rates and the online p_T cut higher, results in only a minor improvement with respect to the baseline scenario. On the other hand, DNN-B has a visible impact on the offline p_T threshold. However, when compared to applying Soft-Killer on the input topocluster collection, it results in a similar performance.

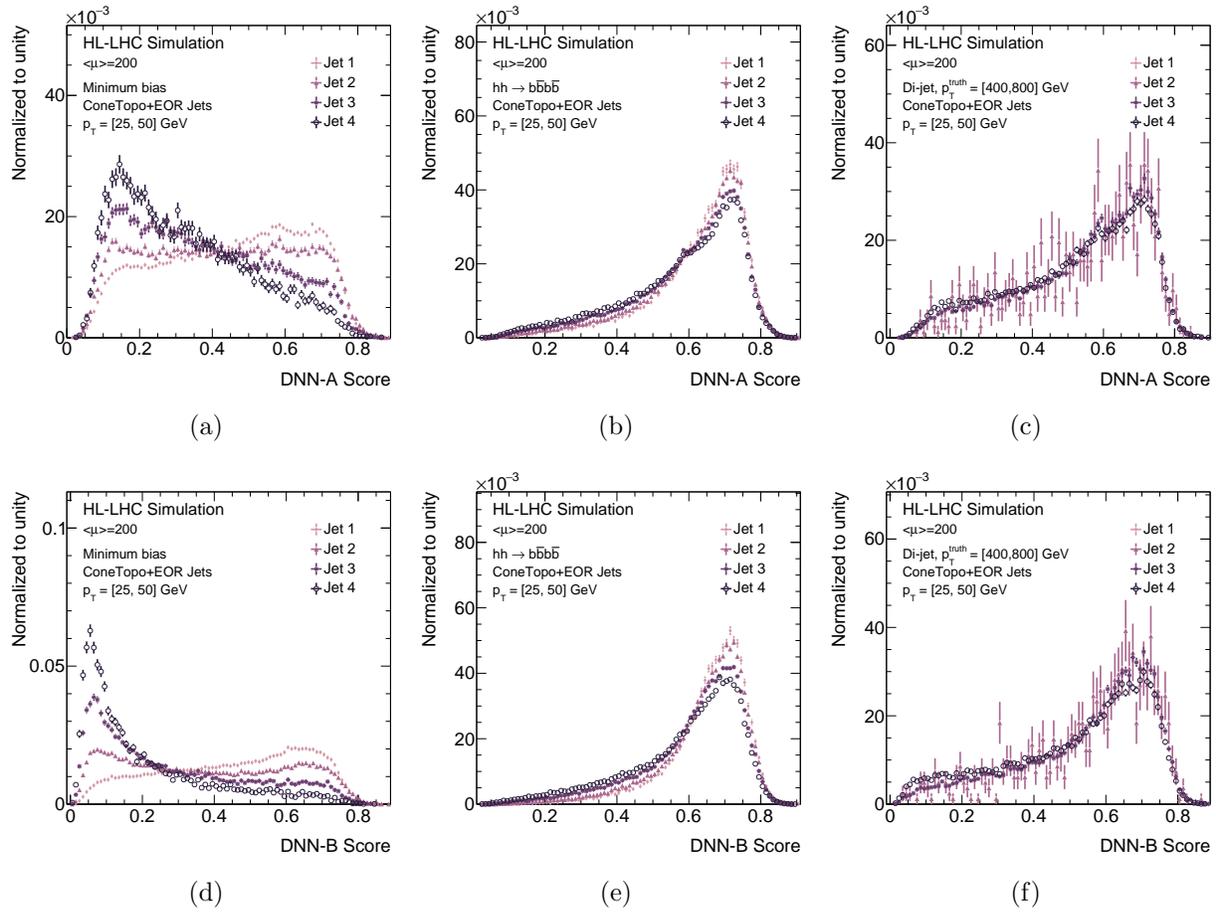


Figure 8.28: Output DNN-A (top) and DNN-B (bottom) scores for the four leading jets in minimum bias (left), di-Higgs (center), and di-jet (right) events, when the jet has an in-training p_T .

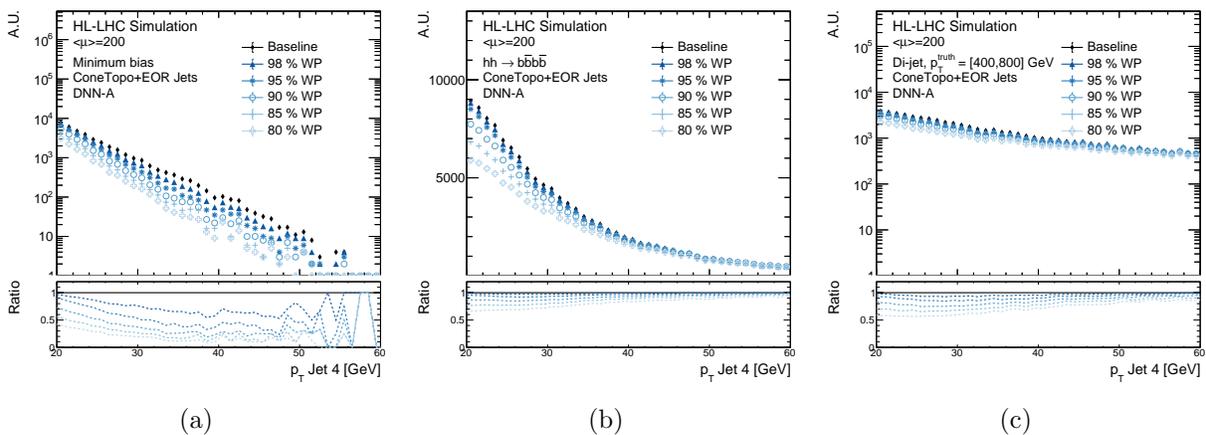


Figure 8.29: Fourth-leading jet p_T distribution after applying DNN-A selection at different WPs in minimum bias (left), di-Higgs (center), and di-jet (right) events.

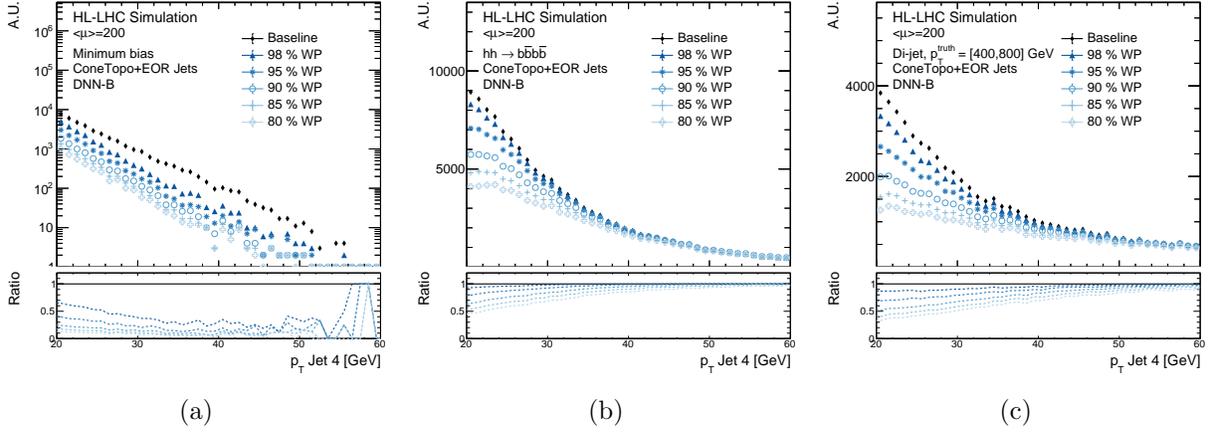


Figure 8.30: Fourth-leading jet p_T distribution after applying DNN-B selection at different WPs in minimum bias (left), di-Higgs (center), and di-jet (right) events.

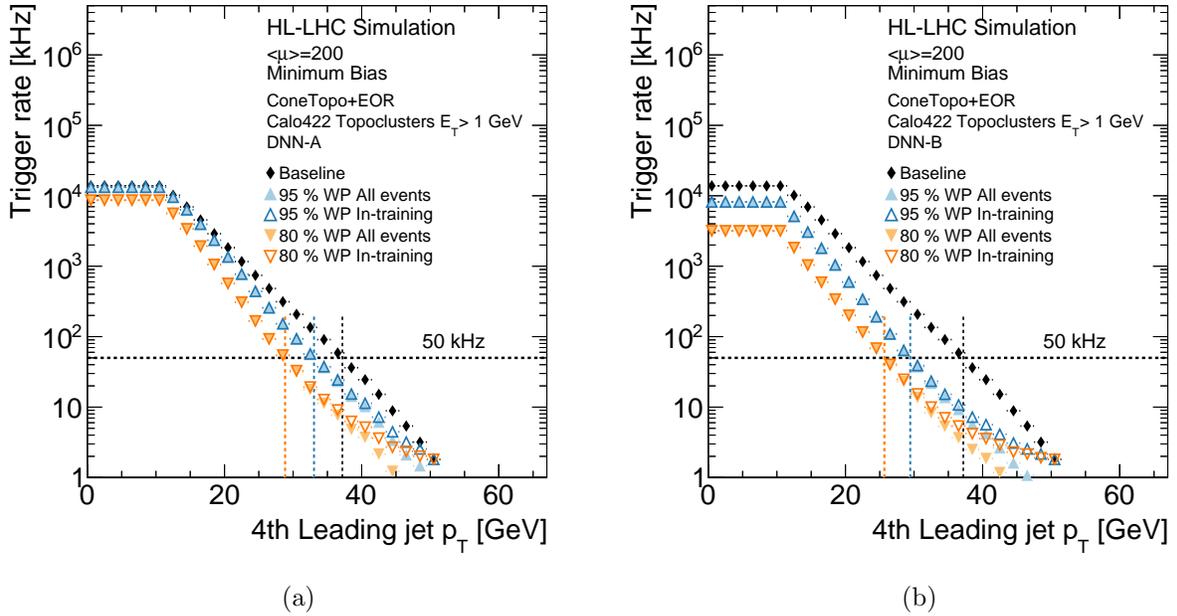


Figure 8.31: Four-jet trigger rates produced with minimum bias sample after applying the DNN selection at 80% and 95% WPs, as well as when using the baseline jet collection. For each WP, the rates are compared when applying the cut to all jets or only to in-training jets with $p_T \in [20, 50]$ GeV. The results are shown for DNN-A (left) and DNN-B (right).

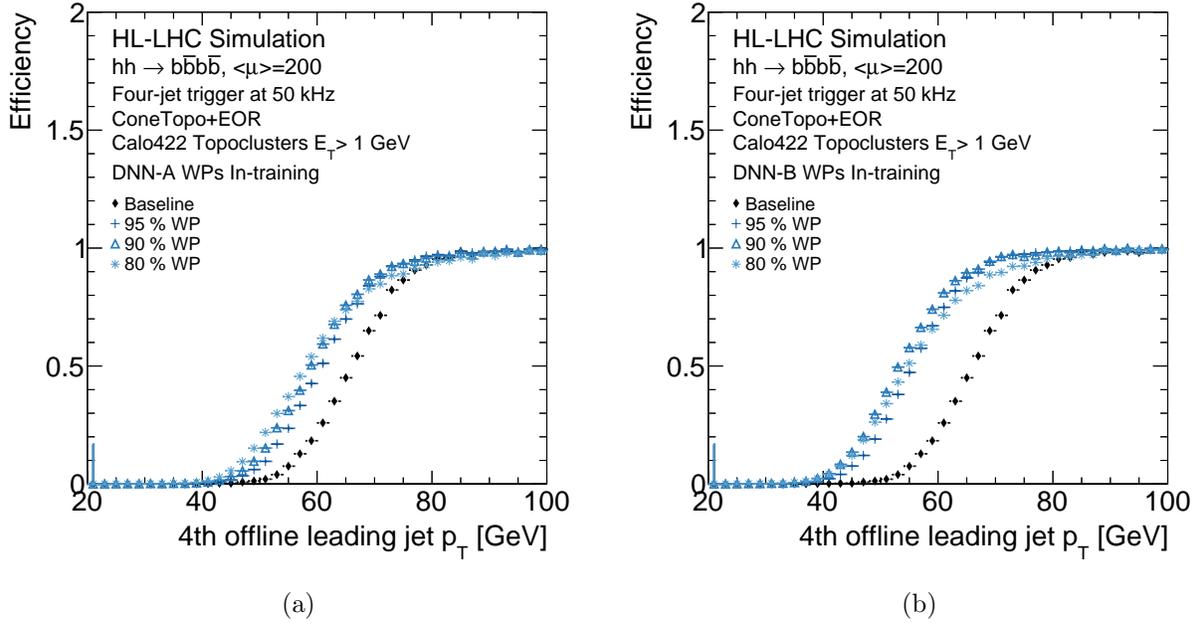


Figure 8.32: Four-jet trigger efficiencies for $hh4b$ signal sample after applying the DNN selection at different WPs to jets with $p_T < 50$ GeV, as well as when using the baseline jet collection. The results are shown for DNN-A (left) and DNN-B (right).

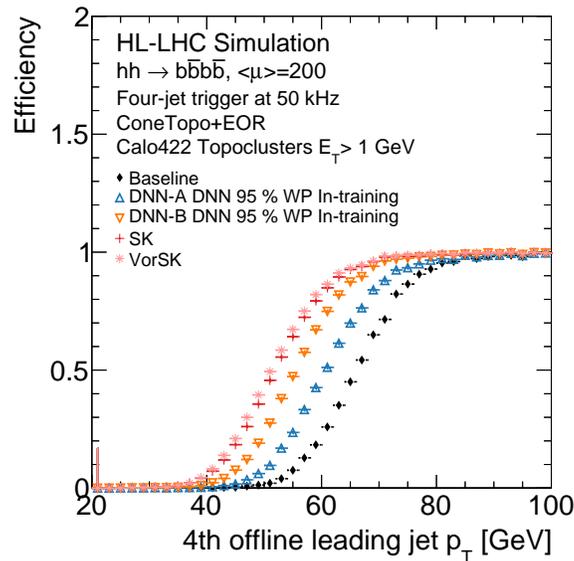


Figure 8.33: Four-jet trigger efficiencies for $hh4b$ signal sample after applying the DNN-A and DNN-B selections at 95% WP, as well as the baseline jet collection and the jet collections produced with the same ConeTopo+EOR algorithm run on SK or VorSK pileup suppressed topocluster collections.

Chapter 9

Conclusion and outlook

Search for new heavy resonances

The search for heavy resonances has been the focus of intense efforts by the ATLAS Collaboration in looking for physics beyond the Standard Model (SM). Several well-motivated models predict that new heavy particles should appear at the TeV scale and decay into highly Lorentz-boosted SM bosons. These models are often interpreted in the context of two general frameworks, the Heavy Vector Triplet model, which predicts an additional SU(2) triplet, and the Two-Higgs-Doublet-Model, which predicts the simplest extension of the SM scalar sector, by including an additional scalar SU(2) doublet. Since the largest branching ratios of W, Z, and Higgs boson decays are into a pair of quarks, boosted jet tagging plays an essential role in this type of searches.

The work performed in this thesis contributed to the search for such new heavy resonances by looking for decays of the new particles into two SM bosons (VV or VH) in semi-leptonic final states. To increase the physics reach of this type of search, a new analysis strategy based on deep-learning algorithms was implemented, with several potential extensions envisioned for the future.

The main contribution of this work was the development of a new deep neural network for the identification of the hadronic decay as coming from a Higgs boson, a W boson, a Z boson, a top quark, or light quarks and gluons. The development of the Multi-Class Tagger (MCT) focused first on large-R jet classification for general boosted decays and was then extended to the resolved jet topology in the context of the analysis. The development included the training of the DNNs, as well as the deployment within the analysis. The latter included the design of a new orthogonalization strategy of the final regions of the analysis, to compare with previous strategies in other VV and VH combination efforts.

The conclusion was that the new MCT strategy does not cause any loss in sensitivity, while it allows to recover up to 20% loss in sensitivity at high resonance mass with respect to the previous efforts. Lastly, the modeling of the MCT scores was studied in the pre-selection and control regions of the analysis. After accounting for normalization differences between background and data by deriving normalization scale factors, the MCT was shown to be

well-modeled and to not need calibrations. Although not discussed in this thesis, the way the MCT was envisioned allows for a straightforward extension to aid in the definition of top- and QCD-enriched control regions. Output scores of the MCT would also be candidate high-level inputs to a possible event-level classifier. These are ideas that can be explored in the future, both in the VV/VH semi-leptonic search, and in other similar searches for heavy resonances.

Upgrade of the HL-LHC Level-0 Trigger

The LHC will soon undergo a major upgrade that will raise the center-of-mass energy to $\sqrt{s} = 14$ TeV and bring the instantaneous luminosity up to $5 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$. The resulting High-Luminosity LHC (HL-LHC) will bring a tenfold increase in the data collected by ATLAS, which will extend the physics reach of the experiment largely beyond the original design. At the same time, the increase in luminosity will inevitably generate higher levels of pileup and radiation, requiring substantial upgrades of the ATLAS detector and TDAQ system to face the harsher conditions. A significant part of this thesis work has involved contributions to the Phase II upgrade of the hardware-based trigger system in preparation for the HL-LHC. This included the development and maintenance of a software simulation framework for the study of new firmware algorithms, as well as the development of a new jet reconstruction and triggering strategy.

A cone algorithm was developed for jet reconstruction in the Global Trigger. While a coarser option than anti- k_t , the standard for offline jet reconstruction, the cone algorithm has been shown to provide equivalent performance. This includes performance metrics such as trigger rates, trigger efficiencies, and signal efficiencies for the specific offline analysis targeting di-Higgs production. Several parameters had to be optimized and their effect and correlations had to be understood to arrive to this result. Different viable options have been identified, with their respective advantages and drawbacks. The finalized set of parameters will be determined by the trade-off between physics performance and hardware resource consumption. Both of these factors are deeply interconnected to the requirements of the other algorithms that will run on the hardware, most of which are still under development. Nevertheless, this research demonstrated what choices and trade-offs will need to be addressed before arriving to the final version, and conclusively established the cone jet algorithm as a viable option for the Global Trigger, with promising avenues for extension to τ -lepton and large-R jet reconstruction.

Due to the nature of hadron-hadron collisions, pileup is an ever-present issue at the LHC, affecting the reconstruction of physical observables and stressing the detector and TDAQ systems. This is particularly true for a multi-jet trigger, which looks for jets in kinematic

regions dominated by pileup, and on which important signatures rely, such as $HH \rightarrow b\bar{b}b\bar{b}$. As the LHC moves towards the high-luminosity era, pileup mitigation will become increasingly challenging, but also necessary to retain the physics reach of the experiment.

This work introduced a novel deep-learning approach for pileup mitigation, with the goal of reducing the rates of high energy background jets in the Level-0 trigger. To accomplish this, two neural networks were trained for the identification and removal of pileup jets using only information about the constituent topological clusters, making full use of the full granularity calorimeter information that will become accessible for the first time at the first stage of the trigger.

The first model (DNN-A) utilizes only energy and spatial information of the topological clusters, while the second model (DNN-B) was additionally provided with boolean flags representing the outcome of the Soft-Killer pileup suppression algorithm for each constituent. Both neural networks were found to reduce the rate of background pileup jets. However, including the results of the offline pileup suppression in the input data proved necessary to achieve substantial improvements in the trigger efficiencies. These results suggested that an event-level characterization of the pileup density is needed in the input data to observe a substantial reduction in the trigger thresholds. Nevertheless, this study demonstrated that significant discrimination power between signal and pileup jets is available in the energy profile of the jet constituents, making this an interesting area for future developments.

Outlook

High energy physics is approaching an exciting phase, as the HL-LHC will open up new search channels, previously inaccessible cross-sections, and more precise measurements of SM observables. New revolutionary discoveries might be around the corner and it is critical that we have all the tools at our disposal ready to get the most out of the data. The unprecedented challenges of data-intensive physics research have made it increasingly clear that standard approaches used to extract meaningful physics have to be rethought. The multi-class jet tagger and the neural network for pileup-jet rejection discussed in this work are illustrative of the broader array of applications where deep learning can significantly enhance our ability to analyze complex data and further our understanding of the universe.

BIBLIOGRAPHY

- [1] S. Weinberg, “Gauge hierarchies,” *Phys. Lett. B* **82**, 387 (1979).
- [2] G. F. Giudice, “Naturally speaking: The naturalness criterion and physics at the LHC,” in *Perspectives on LHC Physics* (WORLD SCIENTIFIC, 2008) pp. 155–178.
- [3] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer, “Heavy Vector Triplets: bridging theory and data,” *J. High Energ. Phys.* **2014**, 60 (2014).
- [4] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, “Theory and phenomenology of two-Higgs-doublet models,” *Phys. Rep.* **516**, 1 (2012).
- [5] D. E. Morrissey and M. J. Ramsey-Musolf, “Electroweak baryogenesis,” *New J. Phys.* **14**, 125003 (2012).
- [6] A. Noble and M. Perelstein, “Higgs self-coupling as a probe of the electroweak phase transition,” *Phys. Rev. D* **78**, 063518 (2008).
- [7] ATLAS Collaboration, “Constraints on the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV,” *Phys. Lett. B* **843**, 137745 (2023).
- [8] CMS Collaboration, “A portrait of the Higgs boson by the CMS experiment ten years after the discovery,” *Nature* **607**, 60 (2022).
- [9] B. Di Micco, M. Gouzevitch, J. Mazzitelli, and C. Vernieri, “Higgs boson potential at colliders: Status and perspectives,” *Rev. Phys.* **5**, 100045 (2020).
- [10] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*, Tech. Rep. (CERN, 2017).
- [11] Wikimedia Commons contributors, “Standard Model of Elementary Particles.” Accessed: 2023-09-30.
- [12] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (CRC Press, 1995).
- [13] F. Mandl and G. Shaw, *Quantum Field Theory*, 2nd ed. (Wiley, 2010).
- [14] A. Zee, *Quantum Field Theory in a Nutshell*, 2nd ed. (Princeton University Press, 2010).
- [15] D. Tong, “Lectures on Quantum Field Theory,” (2006).
- [16] I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics: A Practical Introduction, Volume 1 : From Relativistic Quantum Mechanics to QED*, 4th ed. (Taylor & Francis, 2013).

- [17] A. Djouadi, “The anatomy of electroweak symmetry breaking. Tome I: The Higgs boson in the Standard Model,” *Phys. Rep.* **457**, 1 (2008).
- [18] L. Reina, “TASI 2011 lectures on Higgs-boson physics,” arXiv:1208.5504 (2021).
- [19] The Royal Swedish Academy of Sciences, “Here, at last!” (2013).
- [20] S. L. Glashow, “Partial-symmetries of weak interactions,” *Nucl. Phys.* **22**, 579 (1961).
- [21] P. W. Higgs, “Broken symmetries and the masses of gauge bosons,” *Phys. Rev. Lett.* **13**, 508 (1964).
- [22] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons,” *Phys. Rev. Lett.* **13**, 321 (1964).
- [23] S. Weinberg, “A model of leptons,” *Phys. Rev. Lett.* **19**, 1264 (1967).
- [24] A. Salam, “Weak and Electromagnetic Interactions,” *Conf. Proc. C* **680519**, 367 (1968).
- [25] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett. B* **716**, 1 (2012).
- [26] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett. B* **716**, 30 (2012).
- [27] P. A. M. Dirac, “The quantum theory of the emission and absorption of radiation,” *Proc. R. Soc. Lond. A* **114**, 243 (1927).
- [28] Y. Grossman and P. Tanedo, “Just a taste: Lectures on flavor physics,” in *Anticipating the Next Discoveries in Particle Physics* (WORLD SCIENTIFIC, 2018).
- [29] ATLAS Collaboration, “Combined measurement of the Higgs boson mass from the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4l$ decay channels with the ATLAS detector using $\sqrt{s} = 7, 8,$ and 13 TeV pp collision data,” *Phys. Rev. Lett.* **131**, 251802 (2023).
- [30] N. Cabibbo, “Unitary symmetry and leptonic decays,” *Phys. Rev. Lett.* **10**, 531 (1963).
- [31] M. Kobayashi and T. Maskawa, “CP-Violation in the renormalizable theory of weak interaction,” *Prog. Theor. Phys.* **49**, 652 (1973).
- [32] M. Tanabashi *et al.* (Particle Data Group), “Review of Particle Physics,” *Phys. Rev. D* **98**, 030001 (2018).
- [33] ATLAS Collaboration, “A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery,” *Nature* **607**, 52 (2022).
- [34] R. Sancisi and T. S. van Albada, “Dark matter,” *Observational Cosmology.* **124**, 699 (1987).

- [35] R. Massey, T. Kitching, and J. Richard, “The dark matter of gravitational lensing,” *Rep. Prog. Phys.* **73**, 086901 (2010).
- [36] Y. Fukuda *et al.* (Super-Kamiokande Collaboration), “Evidence for Oscillation of Atmospheric Neutrinos,” *Phys. Rev. Lett.* **81**, 1562 (1998).
- [37] Q. R. Ahmad and SNO Collaboration, “Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory,” *Phys. Rev. Lett.* **89**, 011301 (2002).
- [38] LHCb Collaboration, “Test of lepton universality in beauty-quark decays,” *Nat. Phys.* **18**, 277 (2022).
- [39] Muon $g - 2$ Collaboration, “Measurement of the positive muon anomalous magnetic moment to 0.46 ppm,” *Phys. Rev. Lett.* **126**, 141801 (2021).
- [40] D. Buttazzo, G. Degrassi, P. P. Giardino, G. F. Giudice, F. Sala, A. Salvio, and A. Strumia, “Investigating the near-criticality of the Higgs boson,” *J. High Energ. Phys.* **2013**, 89 (2013).
- [41] J. Ellis, “Higgs Physics,” (2015).
- [42] J. M. Lizana and M. Pérez-Victoria, “Vector triplets at the LHC,” *EPJ Web of Conferences* **60**, 17008 (2013).
- [43] J. de Blas, J. M. Lizana, and M. Pérez-Victoria, “Combining searches of Z' and W' bosons,” *J. High Energ. Phys.* **2013**, 166 (2013).
- [44] M. Perelstein, “Little Higgs models and their phenomenology,” *Prog. in Part. and Nucl. Phys.* **58**, 247 (2007).
- [45] M. J. Dugan, H. Georgi, and D. B. Kaplan, “Anatomy of a composite Higgs model,” *Nucl. Phys. B* **254**, 299 (1985).
- [46] K. Agashe, R. Contino, and A. Pomarol, “The minimal composite Higgs model,” *Nucl. Phys. B* **719**, 165 (2005).
- [47] V. Barger, W. Y. Keung, and E. Ma, “Gauge model with light W and Z bosons,” *Phys. Rev. D* **22**, 727 (1980).
- [48] R. Contino, D. Pappadopulo, D. Marzocca, and R. Rattazzi, “On the effect of resonances in composite Higgs phenomenology,” *J. High Energ. Phys.* **2011**, 81 (2011).
- [49] A. Djouadi, “The anatomy of electro-weak symmetry breaking. Tome II: The Higgs bosons in the Minimal Supersymmetric Model,” *Phys. Rep.* **459**, 1 (2008).
- [50] J. E. Kim, “Light pseudoscalars, particle physics and cosmology,” *Phys. Rep.* **150**, 1 (1987).

- [51] M. Joyce and T. Prokopec and N. Turok, “Nonlocal electroweak baryogenesis. Part 2: The Classical regime,” *Phys. Rev. D* **53**, 2958 (1996).
- [52] S. L. Glashow and S. Weinberg, “Natural conservation laws for neutral currents,” *Phys. Rev. D* **15**, 1958 (1977).
- [53] E. A. Paschos, “Diagonal neutral currents,” *Phys. Rev. D* **15**, 1966 (1977).
- [54] ATLAS Collaboration, “Constraints on new phenomena via Higgs boson couplings and invisible decays with the ATLAS detector,” *J. High Energ. Phys.* **2015**, 206 (2015).
- [55] J. F. Gunion and H. E. Haber, “CP-conserving two-Higgs-doublet model: The approach to the decoupling limit,” *Phys. Rev. D* **67**, 075019 (2003).
- [56] P. J. Bryant, “A brief history and review of accelerators,” in *CERN Accelerator School: Course on General Accelerator Physics*.
- [57] “Report of the Long Range Planning Committee to the CERN Council. 83rd Session of Council,” (1987).
- [58] C. L. Smith, “Genesis of the Large Hadron Collider,” *Phil. Trans. R. Soc. A* **373**, 20140037 (2015).
- [59] R. Assmann, M. Lamont, and S. Myers, “A brief history of the LEP collider,” *Nucl. Phys. B Proc. Suppl.* **109**, 17 (2002).
- [60] L. Evans and P. Bryant, “LHC Machine,” *JINST* **3**, S08001 (2008).
- [61] O. S. Brüning and P. Collier and P. Lebrun and S. Myers and R. Ostojic and J. Poole and P. Proudlock, *LHC Design Report. Vol. 1: The LHC Main Ring*, CERN Yellow Reports: Monographs (CERN, 2004).
- [62] The ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider,” *JINST* **3**, S08003 (2008).
- [63] The CMS Collaboration, “The CMS experiment at the CERN LHC,” *JINST* **3**, S08004 (2008).
- [64] The LHCb Collaboration, “The LHCb Detector at the LHC,” *JINST* **3**, S08005 (2008).
- [65] The ALICE Collaboration, “The ALICE experiment at the CERN LHC,” *JINST* **3**, S08002 (2008).
- [66] J. P. Blewett, “200-GeV Intersecting Storage Accelerators,” *eConf* **C710920**, 501 (1971).
- [67] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl, *LHC Design Report. Vol. 3: The LHC Injector Chain*, CERN Yellow Reports: Monographs (CERN, 2004).
- [68] F. Landua, “The CERN accelerator complex layout in 2022.” (2022).

- [69] D. Boussard and T. P. Linnekar, *The LHC Superconducting RF System*, Tech. Rep. (CERN, 1999).
- [70] S. Baird, *Accelerators for pedestrians; Rev. version*, Tech. Rep. (CERN, 2007).
- [71] V. Parma and L. Rossi, *Performance of the LHC magnet system*, Tech. Rep. (CERN, 2010).
- [72] J. Sterling, “Private communication,” (2013).
- [73] S. van der Meer, *Calibration of the effective beam height in the ISR*, Tech. Rep. (CERN, 1968).
- [74] G. Soyez, “Pileup mitigation at the LHC: A theorist’s view,” *Phys. Rep.* **803**, 1 (2019).
- [75] O. Aberle *et al.*, “High-Luminosity Large Hadron Collider (HL-LHC): Technical design report,” CERN Yellow Reports: Monographs (2020).
- [76] The ATLAS Collaboration, “The ATLAS experiment at the CERN Large Hadron Collider: A description of the detector configuration for Run 3,” arXiv:2305.16623 (2023).
- [77] S. Verdú-Andrés, S. Belomestnykh, I. Ben-Zvi, R. Calaga, Q. Wu, and B. Xiao, “Crab cavities for colliders: past, present and future,” *Nucl. Part. Phys. Proc.* **273**, 193 (2016).
- [78] A. Yamamoto, Y. Doi, Y. Makida, K. Tanaka, T. Haruyama, H. Yamaoka, T. Kondo, S. Mizumaki, S. Mine, K. Wada, S. Meguro, T. Sotoki, K. Kikuchi, and H. ten Kate, “Progress in ATLAS central solenoid magnet,” *IEEE Transactions on Applied Superconductivity* **10**, 353–356 (2000).
- [79] ATLAS Collaboration, *ATLAS Inner Detector: Technical Design Report 1*, Tech. Rep. (1997).
- [80] S. Haywood, L. Rossi, R. Nickerson, and A. Romaniouk (ATLAS), *ATLAS Inner Detector: Technical Design Report 2*, Tech. Rep. (1997).
- [81] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Strip Detector*, Tech. Rep. (CERN, 2017).
- [82] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, Tech. Rep. (CERN, 2017).
- [83] J. Pequeno, “Computer generated image of the ATLAS inner detector,” (2008).
- [84] ATLAS Collaboration, “ATLAS pixel detector electronics and sensors,” *JINST* **3**, P07007 (2008).
- [85] ATLAS Collaboration, “The silicon microstrip sensors of the ATLAS semiconductor tracker,” *Nucl. Instrum. Meth. A* **578**, 98 (2007).

- [86] ATLAS TRT Collaboration *et al.*, “The ATLAS TRT end-cap detectors,” JINST **3**, P10003 (2008).
- [87] ATLAS TRT Collaboration *et al.*, “The ATLAS TRT Barrel Detector,” JINST **3**, P02014 (2008).
- [88] B. Mindur (ATLAS), “ATLAS Transition Radiation Tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider,” Nucl. Instr. and Meth. in Phys. Res. A **845**, 257 (2017).
- [89] M. Capeans, G. Darbo, K. Einsweiler, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet (ATLAS), *ATLAS Insertable B-Layer Technical Design Report*, Tech. Rep. (2010).
- [90] ATLAS Collaboration, “Production and integration of the ATLAS Insertable B-Layer,” JINST **13**, T05008 (2018).
- [91] C. Grupen and B. Shwartz, *Particle detectors*, 2nd ed. (Cambridge University Press, 2008).
- [92] C. W. Fabjan and F. Gianotti, “Calorimetry for particle physics,” Rev. Mod. Phys. **75**, 1243 (2003).
- [93] ATLAS Collaboration (ATLAS), *ATLAS liquid-argon calorimeter: Technical Design Report* (CERN, 1996).
- [94] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report* (CERN, 1996).
- [95] ATLAS Collaboration, “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1,” Eur. Phys. J. C **77**, 490 (2017).
- [96] S. Palestini, “The muon spectrometer of the ATLAS experiment,” Nucl. Phys. B Proc. Suppl. **125**, 337 (2003).
- [97] L. Adamczyk, E. Banaś, A. Brandt, M. Bruschi, S. Grinstein, J. Lange, M. Rijssenbeek, P. Sicho, R. Staszewski, T. Sykora, M. Trzebiński, J. Chwastowski, and K. Korcyl, *Technical Design Report for the ATLAS Forward Proton Detector*, Tech. Rep. (CERN, 2015).
- [98] ATLAS Collaboration, “Performance of the ATLAS trigger system in 2015,” Eur. Phys. J. C **77**, 317 (2017).
- [99] ATLAS Collaboration, “Operation of the ATLAS Trigger System in Run 2,” JINST **15**, P10004 (2020).
- [100] ATLAS Collaboration, “Performance of the ATLAS muon triggers in Run 2,” JINST **15**, P09015 (2020).

- [101] R. Achenbach *et al.*, “The ATLAS Level-1 calorimeter trigger,” JINST **3**, P03001 (2008).
- [102] ATLAS Collaboration, “Performance of the upgraded PreProcessor of the ATLAS Level-1 Calorimeter Trigger,” JINST **15**, P11016 (2020).
- [103] R. Simoniello (ATLAS), *The ATLAS Level-1 Topological Processor: from design to routine usage in Run-2*, Tech. Rep. (CERN, 2019).
- [104] H. Bertelsen, G. Carrillo Montoya, P.-O. Deviveiros, T. Eifert, G. Galster, J. Glatzer, S. Haas, A. Marzin, T. Pauly M.V. Silva Oliveira and, K. Schmieden, R. Spiwojs, and J. Stelzer, “Operation of the upgraded ATLAS Central Trigger Processor during the LHC Run 2,” JINST **11**, C02020 (2020).
- [105] ATLAS Collaboration, “Athena,” (2019).
- [106] M. Cacciari, G. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm,” J. High Energ. Phys. **04**, 063 (2008).
- [107] ATLAS Collaboration, *Trigger monitoring and rate predictions using Enhanced Bias data from the ATLAS Detector at the LHC*, Tech. Rep. (CERN, 2016).
- [108] ATLAS Collaboration, *Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System*, Tech. Rep. (CERN, 2013).
- [109] ATLAS Collaboration, *Technical Design Report: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*, Tech. Rep. (CERN, 2020).
- [110] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter Phase-II Upgrade: Technical Design Report*, Tech. Rep. (CERN, 2017).
- [111] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS Trigger and Data Acquisition System – Event Filter Tracking Amendment*, Tech. Rep. (CERN, 2022).
- [112] J. Pequeno, “Event cross section in a computer generated image of the ATLAS detector,” (2008).
- [113] ATLAS Collaboration, “Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2,” Eur. Phys. J. C **77**, 673 (2017).
- [114] ATLAS Collaboration, “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC.” Eur. Phys. J. C **77**, 332 (2017).
- [115] ATLAS Collaboration, “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV,” Eur. Phys. J. C **79**, 639 (2019).
- [116] ATLAS Collaboration, *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*, Tech. Rep. (CERN, 2017).

- [117] ATLAS Collaboration, “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data,” *JINST* **14**, P12006 (2019).
- [118] ATLAS Collaboration, “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV,” *Eur. Phys. J. C* **81**, 578 (2021).
- [119] ATLAS Collaboration, “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV,” *Eur. Phys. J. C* **78**, 903 (2018).
- [120] ATLAS Collaboration, *Object-based missing transverse momentum significance in the ATLAS detector*, Tech. Rep. (CERN, 2018).
- [121] ATLAS Collaboration, “Performance of b -jet identification in the ATLAS experiment,” *JINST* **11**, P04008 (2016).
- [122] ATLAS Collaboration, “ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset,” *Eur. Phys. J. C* **83**, 681 (2023).
- [123] P.A. Zyla *et al.* (Particle Data Group), “Review of Particle Physics,” *PTEP* **2020**, 083C01 (2020), and 2021 update.
- [124] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run*, Tech. Rep. (CERN, 2017).
- [125] ATLAS Collaboration, *Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, Tech. Rep. (CERN, 2017).
- [126] A. Hoecker, “Physics at the LHC Run-2 and beyond,” in *Proceedings of the 2016 European School of High-Energy Physics* (CERN, 2016).
- [127] G. P. Salam, “Elements of QCD for hadron colliders,” in *2009 European School of High-Energy Physics* (2010).
- [128] P. Skands, “Introduction to QCD,” arXiv:1207.2389 (2012).
- [129] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Höche, H. Hoeth, F. Krauss, L. Lönnblad, E. Nurse, P. Richardson, S. Schumann, M. H. Seymour, T. Sjöstrand, P. Skands, and B. Webber, “General-purpose event generators for LHC physics,” *Phys. Rep.* **504**, 145 (2011).
- [130] T. Carli, K. Rabbertz, and S. Schumann, “Studies of Quantum Chromodynamics at the LHC,” in *The Large Hadron Collider: Harvest of Run 1* (Springer International Publishing, Cham, 2015) pp. 139–194.
- [131] S. Agostinelli *et al.*, “Geant4 – a simulation toolkit,” *Nucl. Instrum. Meth. A* **506**, 250 (2003).

- [132] G. P. Salam G. Soyez, “A practical seedless infrared-safe cone jet algorithm,” *J. High Energ. Phys.* **2007**, 086 (2007).
- [133] G. P. Salam, “Towards jetography,” *Eur. Phys. J. C* **67**, 637 (2010).
- [134] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions,” *Phys. Rev. D* **48**, 3160 (1993).
- [135] S. Catani, Y. Dokshitzer, M. Seymour, and B. Webber, “Longitudinally-invariant k_T -clustering algorithms for hadron-hadron collisions,” *Nucl. Phys. B* **406**, 187 (1993).
- [136] Yu. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, “Better jet clustering algorithms,” *J. High Energ. Phys.* **1997**, 001 (1997).
- [137] M. Wobisch and T. Wengler, “Hadronization corrections to jet cross sections in deep inelastic scattering,” arXiv:9907280 (1999).
- [138] M. Cacciari and G. P. Salam, “Dispelling the N^3 myth for the k_t jet-finder,” *Phys. Lett. B* **641**, 57 (2006).
- [139] ATLAS Collaboration, “Jet reconstruction and performance using particle flow with the ATLAS detector,” *Eur. Phys. J. C* **77**, 466 (2017).
- [140] ATLAS Collaboration, “Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,” *Eur. Phys. J. C* **81**, 689 (2021).
- [141] ATLAS Collaboration, “In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector.” *Eur. Phys. J. C* **79**, 135 (2019).
- [142] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual,” *Eur. Phys. J. C* **72**, 1896 (2012).
- [143] ATLAS Collaboration, *Improving jet substructure performance in ATLAS using Track-CaloClusters*, Tech. Rep. (CERN, 2017).
- [144] ATLAS Collaboration, “Optimization of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions.” *Eur. Phys. J. C* **81**, 334 (2021).
- [145] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, Tech. Rep. (CERN, 2017).
- [146] A. J. Larkoski, I. Moulton, and B. Nachman, “Jet substructure at the Large Hadron Collider: A review of recent advances in theory and machine learning,” *Phys. Rep.* **841**, 1 (2020).
- [147] R. Kogler *et al.*, “Jet substructure at the Large Hadron Collider: Experimental review,” *Rev. Mod. Phys.* **91**, 045003 (2019).

- [148] G. Fox, T. Tse, and S. Wolfram, “Event shapes in deep inelastic lepton-hadron scattering,” Nucl. Phys. B **165**, 80 (1980).
- [149] ATLAS Collaboration, “Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector,” J. High Energ. Phys. **2013**, 76 (2013).
- [150] J. Thaler and K. Van Tilburg, “Identifying boosted objects with N-subjettiness.” J. High Energ. Phys. **2011**, 15 (2011).
- [151] A. J. Larkoski, G. P. Salam, and J. Thaler, “Energy correlation functions for jet substructure,” J. High Energ. Phys. **2013**, 108 (2013).
- [152] A. J. Larkoski, I. Moutl, and D. Neill, “Power counting to better jet observables,” J. High Energ. Phys. **2014**, 9 (2014).
- [153] A.J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft drop,” J. High Energ. Phys. **2014**, 146 (2014).
- [154] M. Cacciari, G. P. Salam, and G. Soyez, “SoftKiller, a particle-level pileup removal method,” Eur. Phys. J. C **75**, 59 (2015).
- [155] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup per particle identification,” J. High Energ. Phys. **2014**, 59 (2014).
- [156] ATLAS Collaboration, *Constituent-level pile-up mitigation techniques in ATLAS*, Tech. Rep. (CERN, 2017).
- [157] P. Berta, M. Spousta, D. W. Miller, and R. Leitner, “Particle-level pileup subtraction for jets and jet shapes,” J. High Energ. Phys. **2014**, 92 (2014).
- [158] G. Cowan, *Statistical data analysis* (Oxford University Press, New York, 1998).
- [159] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).
- [160] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, USA, 2016).
- [161] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” Eur. Phys. J. C **71**, 1554 (2011).
- [162] L. Randall and R. Sundrum, “Large mass hierarchy from a small extra dimension,” Phys. Rev. Lett. **83**, 3370 (1999).
- [163] The ATLAS Collaboration, “Search for heavy resonances decaying into a W or Z boson and a Higgs boson in the $l^+l^-b\bar{b}$, $l\nu b\bar{b}$, and $\nu\bar{\nu}b\bar{b}$ with pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,” Phys. Lett. B **765**, 32 (2017).
- [164] ATLAS Collaboration, “Search for heavy resonances decaying into a W or Z boson and a Higgs boson in final states with leptons and b-jets in 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector,” J. High Energ. Phys. **2018**, 174 (2018).

- [165] ATLAS Collaboration, “Search for heavy resonances decaying into a Z or W boson and a Higgs boson in final states with leptons and b-jets in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,” J. High Energ. Phys. **2023**, 016 (2023).
- [166] ATLAS Collaboration, “Search for WW/WZ resonance production in $lvqq$ final states in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,” J. High Energ. Phys. **2018**, 42 (2018).
- [167] ATLAS Collaboration, “Searches for heavy ZZ and ZW resonances in the $llqq$ and $\nu\nuqq$ final states in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,” J. High Energ. Phys. **2018**, 9 (2018).
- [168] ATLAS Collaboration, “Search for heavy diboson resonances in semileptonic final states in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,” Eur. Phys. J. C **80**, 1165 (2020).
- [169] CMS Collaboration, “Search for heavy resonances decaying to WW , WZ , or WH boson pairs in a final state consisting of a lepton and a large-radius jet in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$,” Phys. Rev. D **105**, 032008 (2022).
- [170] CMS Collaboration, “Search for a heavy vector resonance decaying to a Z boson and a Higgs boson in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$,” Eur. Phys. J. C **81**, 688 (2021).
- [171] ATLAS Collaboration, “Search for resonances decaying into a weak vector boson and a Higgs boson in the fully hadronic final state produced in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,” Phys. Rev. D **102**, 112008 (2020).
- [172] ATLAS Collaboration, “Search for diboson resonances in hadronic final states in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,” J. High Energ. Phys. **2019**, 91 (2019).
- [173] CMS Collaboration, “Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} = 13 \text{ TeV}$,” Eur. Phys. J. C **77**, 636 (2017).
- [174] CMS Collaboration, “Search for a heavy pseudoscalar Higgs boson decaying into a 125 GeV Higgs boson and a Z boson in final states with two tau and two light leptons at $\sqrt{s} = 13 \text{ TeV}$,” High Energ. Phys. **2020**, 65 (2020).
- [175] ATLAS Collaboration, “Combination of searches for heavy resonances decaying into bosonic and leptonic final states using 36 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13 \text{ TeV}$ with the atlas detector,” Phys. Rev. D **98**, 052008 (2018).
- [176] ATLAS Collaboration, “Combination of searches for heavy resonances using 139 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,” ATLAS-CONF-2022-028 (2022).

- [177] CMS Collaboration, “Combination of searches for heavy resonances decaying to WW, WZ, ZZ, WH, and ZH boson pairs in proton–proton collisions at $\sqrt{s} = 8$ and 13 TeV,” Phys. Lett. B **774**, 533 (2017).
- [178] CMS Collaboration, “Combination of CMS searches for heavy resonances decaying to pairs of bosons or leptons,” Phys. Lett. B **798**, 134952 (2019).
- [179] CMS Collaboration, “Search for heavy Higgs bosons decaying to a top quark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV,” J. High Energ. Phys. **2020**, 171 (2020).
- [180] ATLAS Collaboration, “Search for Heavy Higgs bosons decaying into two tau leptons with the ATLAS detector using pp collisions at 13 TeV,” Phys. Rev. Lett. **125**, 051801 (2020).
- [181] ATLAS Collaboration, *Summary plots for beyond Standard Model Higgs boson benchmarks for direct and indirect searches*, Tech. Rep. (CERN, 2022).
- [182] ATLAS Collaboration, *Summary of Diboson Resonance Searches at the ATLAS experiment using full run-2 data*, Tech. Rep. (CERN, 2023).
- [183] A. Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network,” Physica D: nonlinear Phenomena **404**, 132306 (2020).
- [184] ATLAS Collaboration, “Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC,” Eur. Phys. J. C **79**, 375 (2019).
- [185] ATLAS Collaboration, “Performance of W/Z taggers using UFO jets in ATLAS,” (2021).
- [186] A. M. Sirunyan *et al.* (CMS), “Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques,” JINST **15**, P06005 (2020).
- [187] W. D. Goldberger and M. B. Wise, “Modulus Stabilization with Bulk Fields,” Phys. Rev. Lett. **83**, 4922 (1999).
- [188] W. D. Goldberger and M. B. Wise, “Phenomenology of a stabilized modulus,” Phys. Lett. B **475**, 275 (2000).
- [189] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, “Warped gravitons at the LHC and Beyond,” Phys. Rev. D **76**, 036006 (2007).
- [190] J. Alwall, M. Herquet, F. Maltoni, and T. Stelzer, “MadGraph 5: Going Beyond,” J. High Energ. Phys. **2011**, 128 (2011).
- [191] NNPDF Collaboration, “Parton distributions for the LHC run II,” J. High Energ. Phys. **2015**, 40 (2015).
- [192] T. Sjöstrand, S. Mrenna, and P. Skands, “A brief introduction to PYTHIA 8.1,” Comput. Phys. Commun. **178**, 852 (2008).

- [193] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, Tech. Rep. (CERN, 2014).
- [194] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, “FeynRules 2.0 – A complete toolbox for tree-level phenomenology,” *Comput. Phys. Commun.* **185**, 2250 (2014).
- [195] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, Tech. Rep. (CERN, 2014).
- [196] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV*, Tech. Rep. (CERN, 2017).
- [197] Martín Abadi *et al.*, “TensorFlow: Large-scale machine learning on heterogeneous systems,” (2016).
- [198] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method,” *J. High Energ. Phys.* **11**, 070 (2007).

APPENDIX A. Analysis

MCT Modeling in 0- and 1-lepton pre-selection regions

This appendix is the continuation of the studies presented in Sec. 7.11.

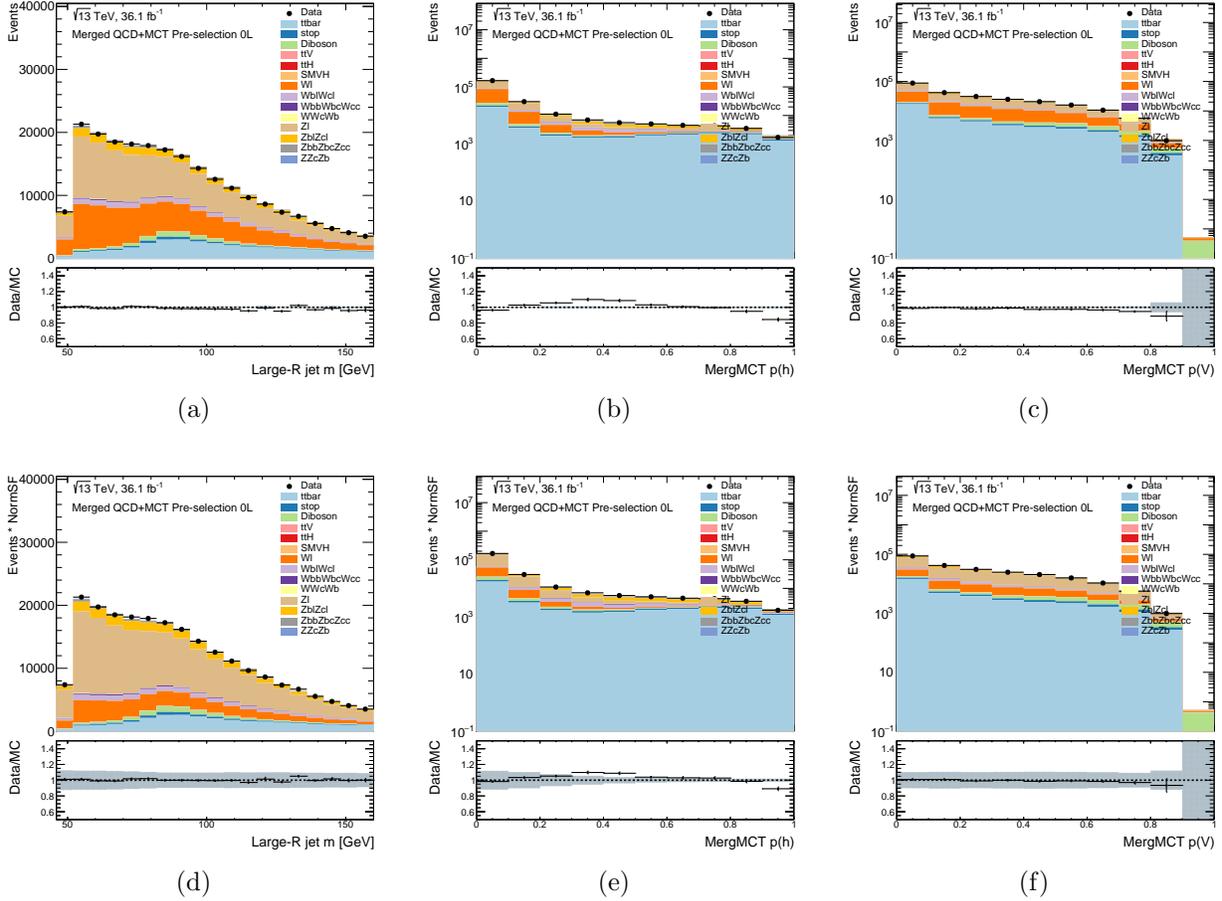


Figure 9.1: Data and MC comparison in the inclusive merged pre-selection region in the 0-lepton channel. Distributions of the large-R jet mass and the raw merged MCT scores $p(h)$ and $p(V)$ are shown before (top) and after (bottom) applying the normalization SFs.

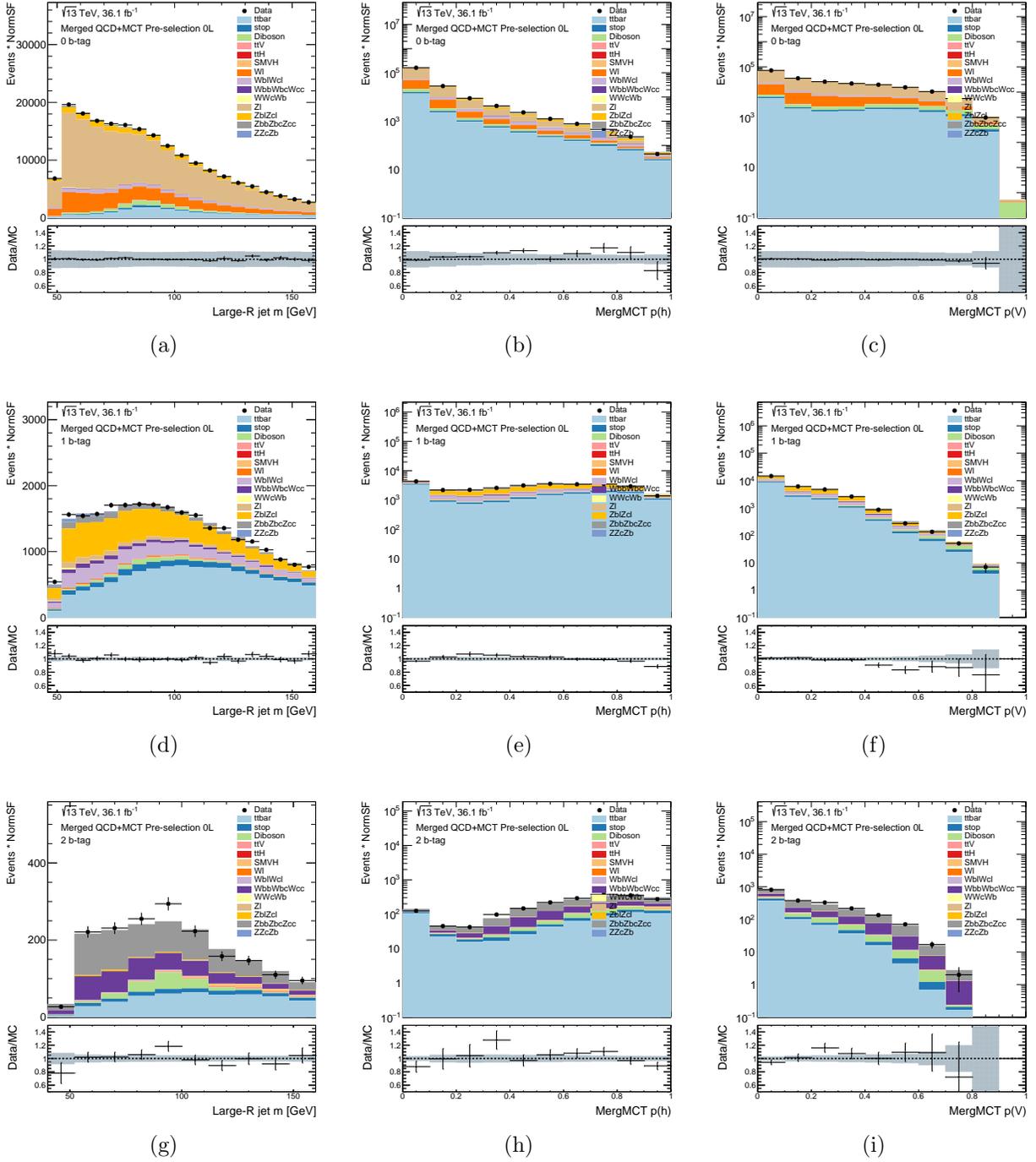


Figure 9.2: Data and MC comparison in the merged pre-selection regions separated by the number of b -tagged jets in the 0-lepton channel. Distributions of the large- R jet mass and the raw merged MCT scores $p(h)$ and $p(V)$ are shown after applying the normalization SFs in the 0 b -tag (top), 1 b -tag (middle), and 2 b -tag (bottom) regions.

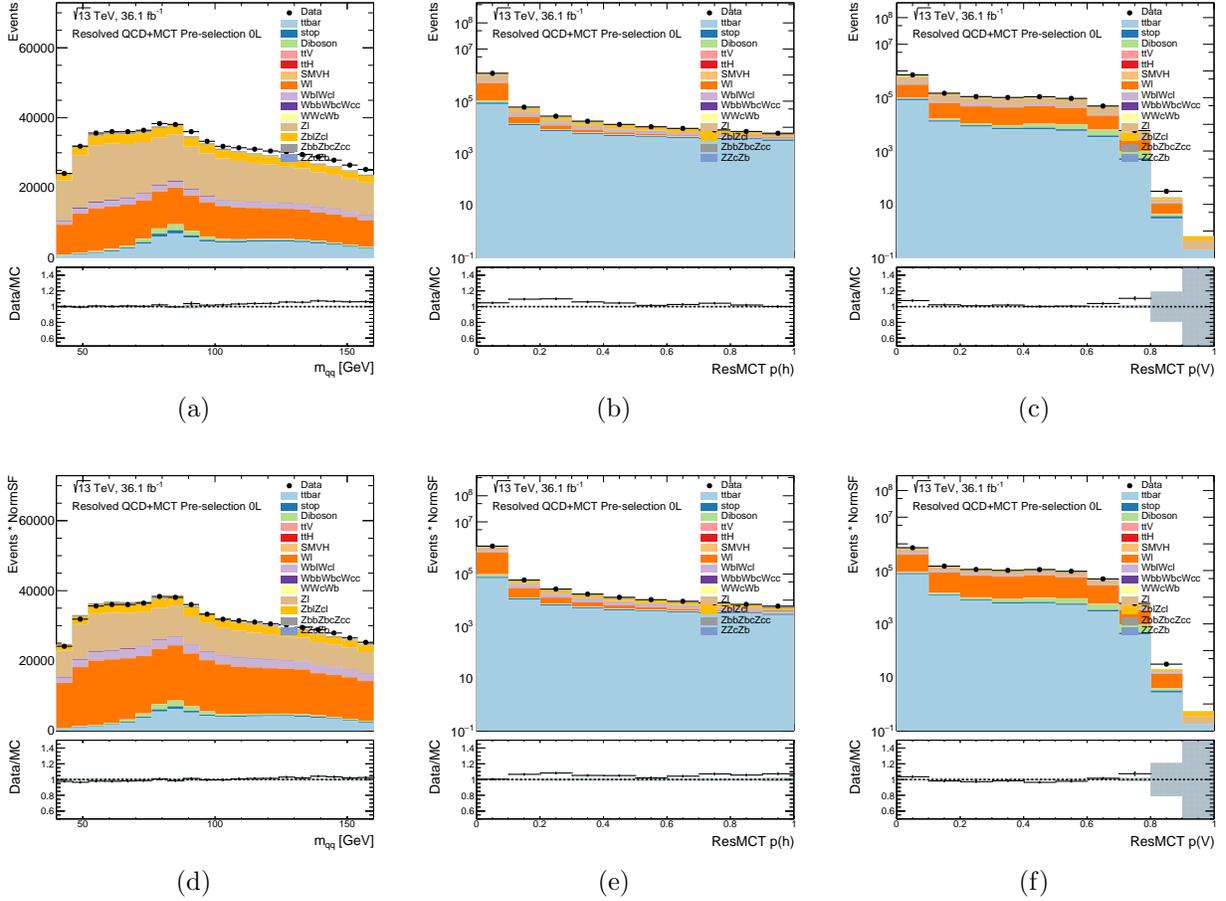


Figure 9.3: Data and MC comparison in the inclusive resolved pre-selection region in the 0-lepton channel. Distributions of the di-jet mass and the raw resolved MCT scores $p(h)$ and $p(V)$ are shown before (top) and after (bottom) applying the normalization SFs.

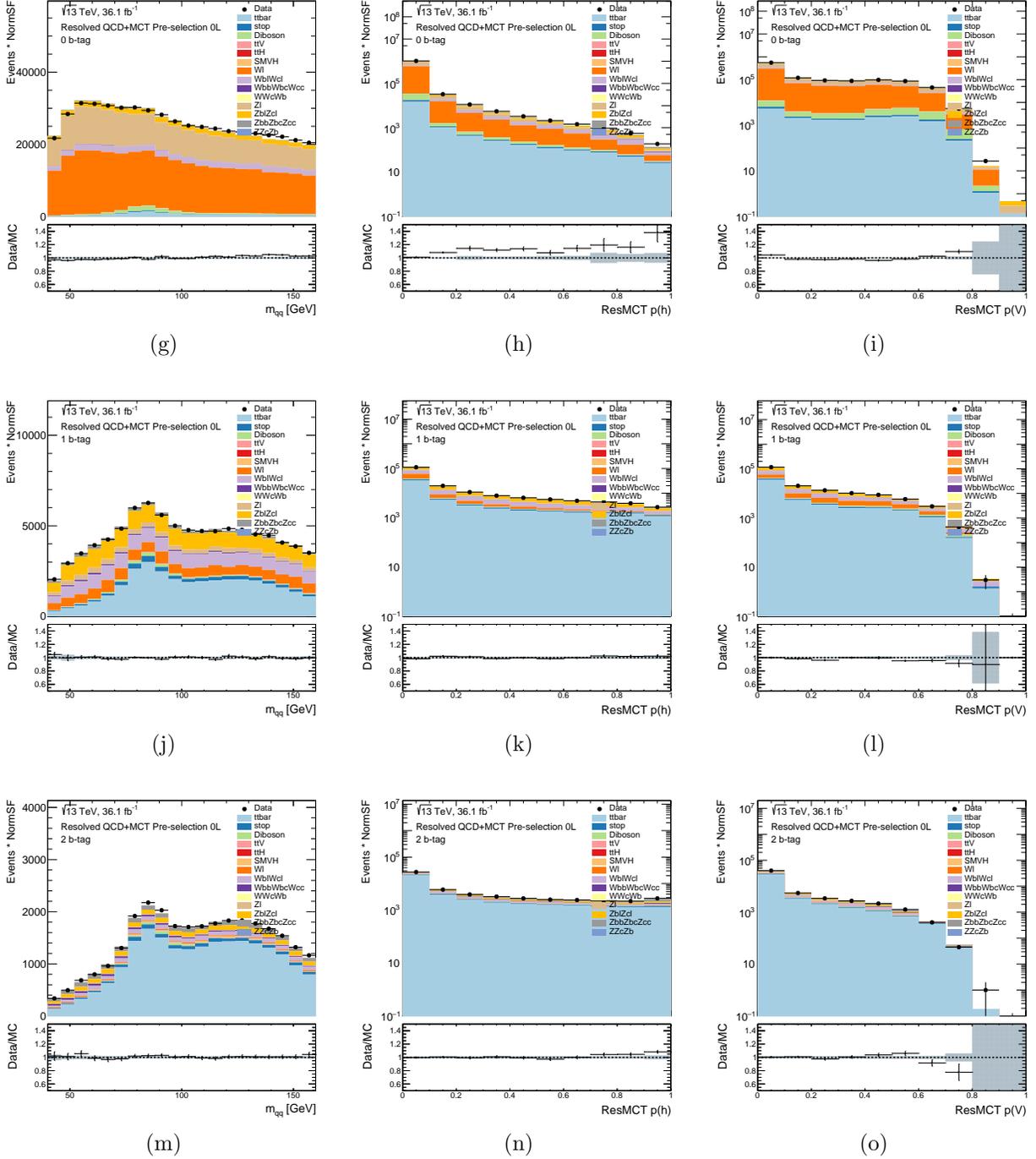


Figure 9.3: Data and MC comparison in the resolved pre-selection regions separated by the number of b -tagged jets in the 0-lepton channel. Distributions of the di-jet mass and the raw resolved MCT scores $p(h)$ and $p(V)$ are shown after applying the normalization SFs in the 0 b -tag (top), 1 b -tag (middle), and 2 b -tag (bottom) regions.

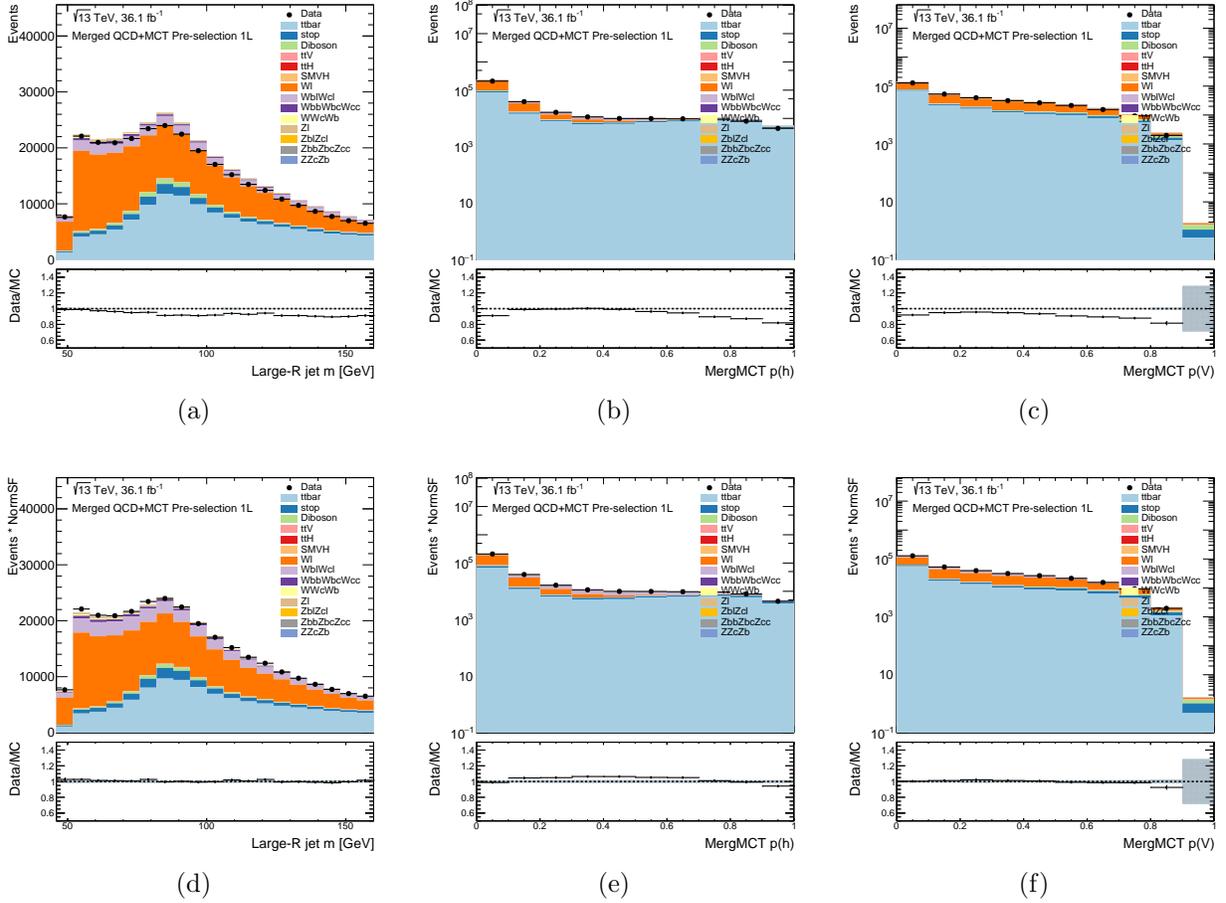


Figure 9.4: Data and MC comparison in the inclusive merged pre-selection region in the 1-lepton channel. Distributions of the large-R jet mass and the raw merged MCT scores $p(h)$ and $p(V)$ are shown before (top) and after (bottom) applying the normalization SFs.

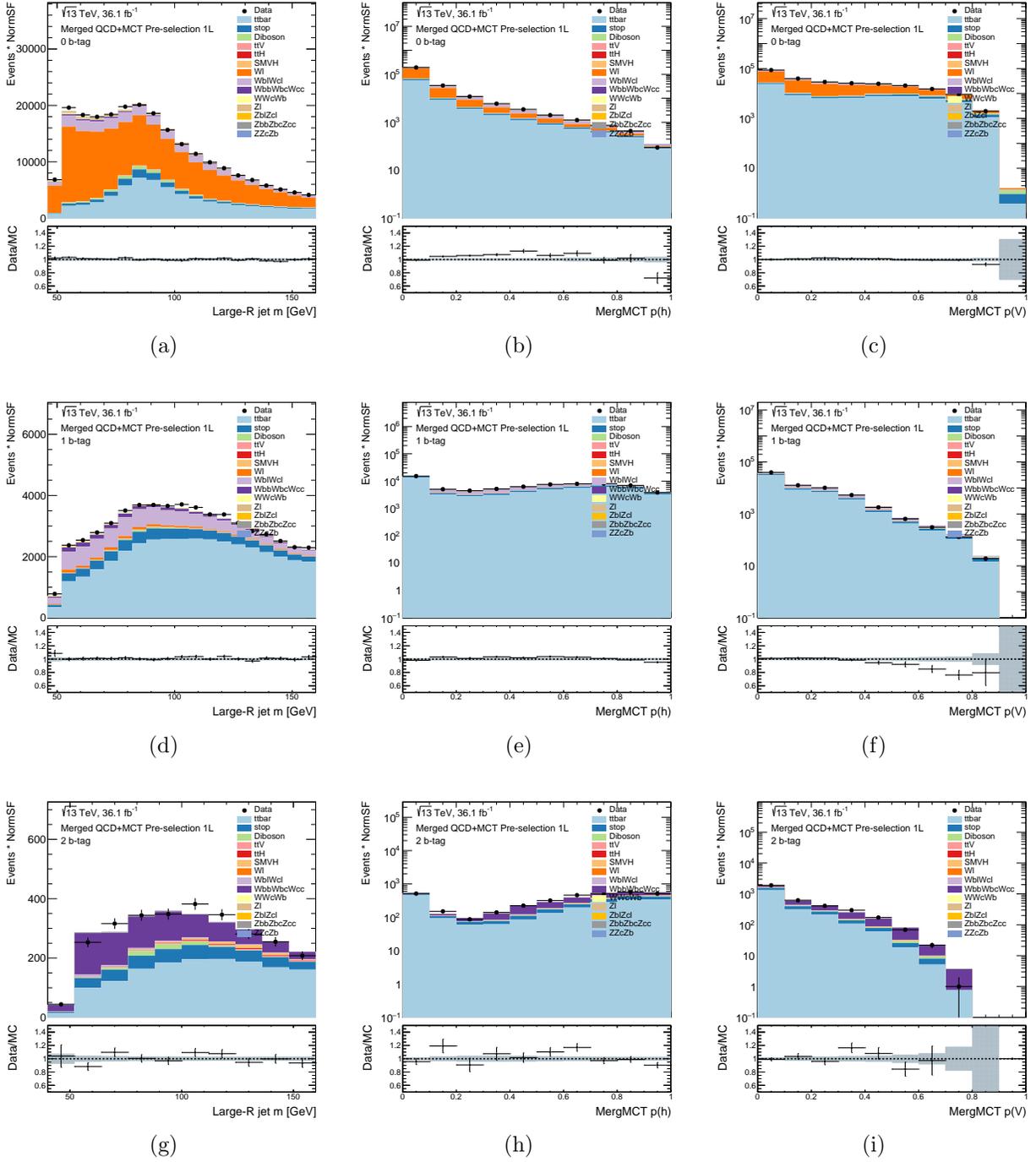


Figure 9.5: Data and MC comparison in the merged pre-selection regions separated by the number of b-tagged jets in the 1-lepton channel. Distributions of the large-R jet mass and the raw merged MCT scores $p(h)$ and $p(V)$ are shown after applying the normalization SFs in the 0 b -tag (top), 1 b -tag (middle), and 2 b -tag (bottom) regions.

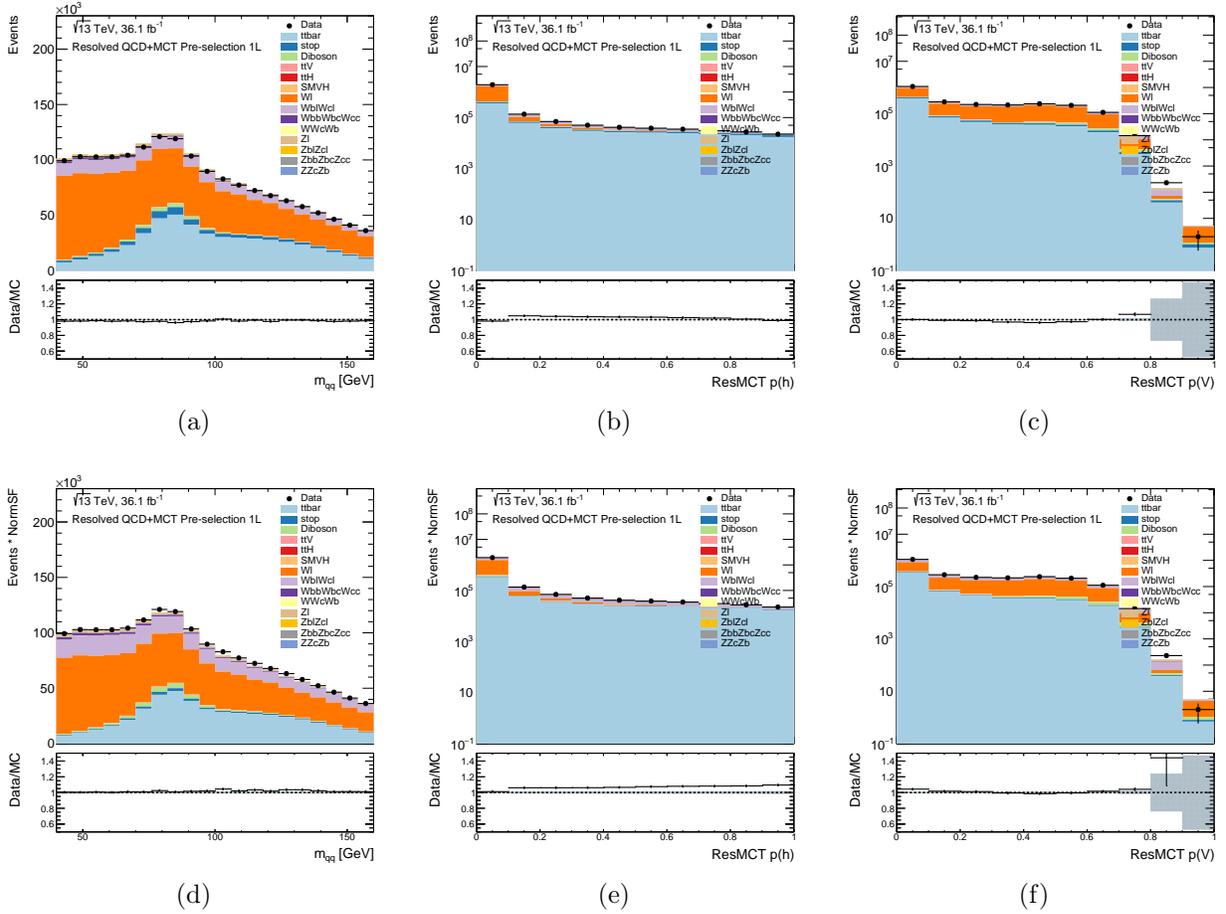


Figure 9.6: Data and MC comparison in the inclusive resolved pre-selection region in the 0-lepton channel. Distributions of the di-jet mass and the raw resolved MCT scores $p(h)$ and $p(V)$ are shown before (top) and after (bottom) applying the normalization SFs.

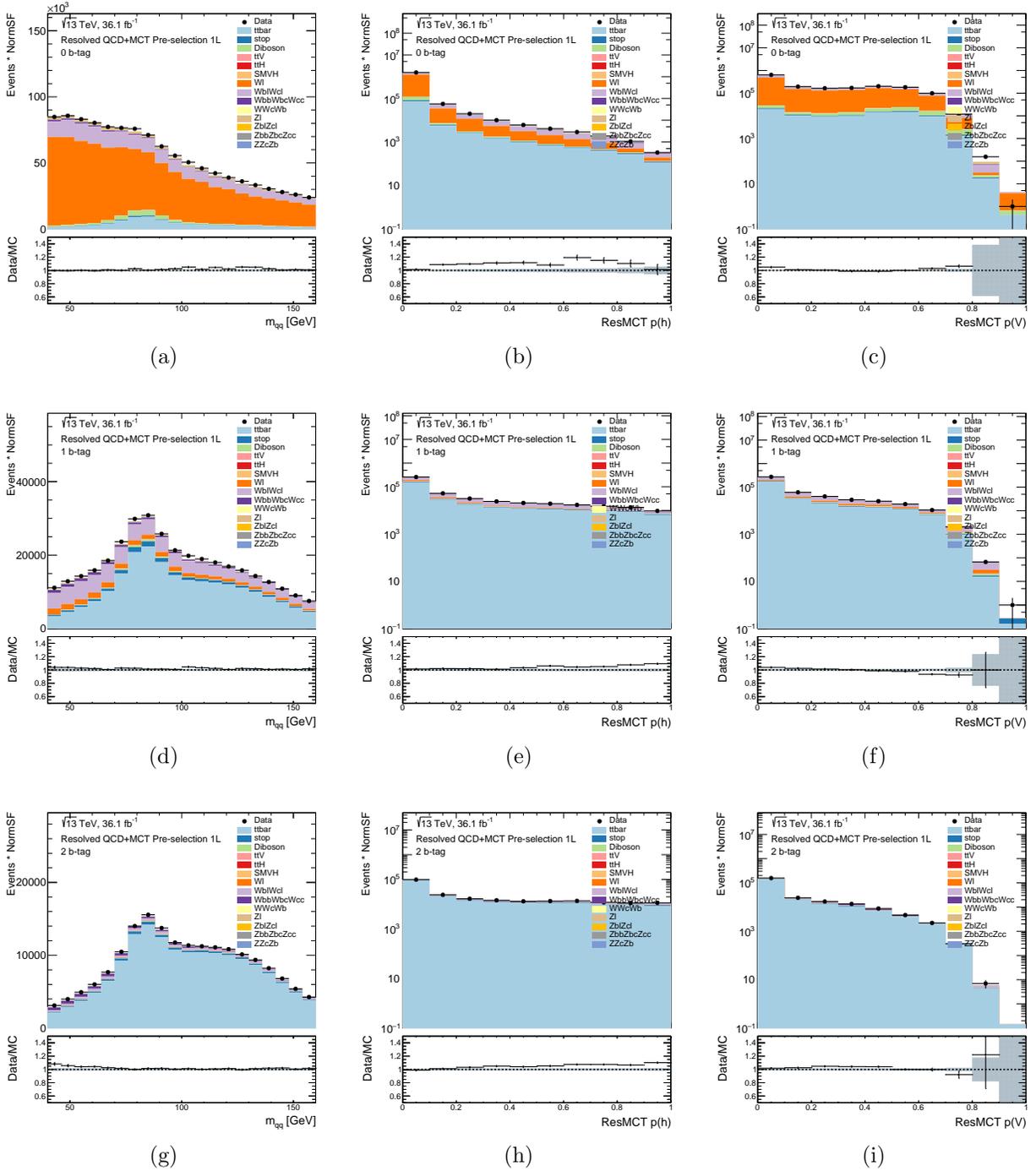


Figure 9.7: Data and MC comparison in the resolved pre-selection regions separated by the number of b -tagged jets in the 1-lepton channel. Distributions of the di-jet mass and the raw resolved MCT scores $p(h)$ and $p(V)$ are shown after applying the normalization SFs in the 0 b -tag (top), 1 b -tag (middle), and 2 b -tag (bottom) regions.

APPENDIX B. Trigger

Cone jets performance with jFEX seeding

This study checked the use of jFEX trigger objects (TOBs) as seeds for the cone jet algorithm discussed in Sec. 8.3 (see Sec. 4.3.4 for an overview of the jFEX algorithm). The resulting jets have the same location as the corresponding jFEX jets, but different energy: while the jFEX jet is built with towers, the cone jet is built with the higher energy resolution of the topoclusters. As with the standard `ConeTopo` jets, the input topoclusters are thresholded according to the specified E_T cut and energy-overlap removal is applied. The resulting jet collection is referred to as `ConeJFEX`. Note that the jFEX algorithm is characterized by a minimum distance requirement between the towers seeding the jFEX objects, which enforces a minimum distance requirement on the seeds of the cone algorithm.

The performance of the `ConeJFEX` collection was compared to the `ConeTopo` and `AntiKt422` jets. Fig. 9.8 shows one- and multi-jet trigger efficiencies for di-Higgs and Z' signals. The trigger efficiencies were built using a common arbitrary online p_T cut of 30 GeV, which allowed to overlay the turn-on curves for better comparison. While equivalent performance was observed for the di-Higgs signal, the Z' sample showed a plateau inefficiency for the `ConeJFEX` online jet collection. The cause of this behavior was identified by separating the events according to how isolated the n^{th} offline leading jet is. Fig. 9.9 shows the `ConeJFEX` trigger efficiencies separated in bins of dR of isolation of the offline jets. For instance, an event enters the $0.4 < dR < 0.6$ bin if the closest dR distance between any pair of the four leading jets is a value between 0.4 and 0.6. Clearly, the plateau inefficiency was mostly originating from the first bin, which had jets closer than $dR = 0.6$, indicating that the `ConeJFEX` algorithm was failing to reconstruct nearby jets. No plateau inefficiency was observed for `ConeTopo` jets thanks to the absence of any restriction on the minimum distance between the seeding topoclusters.

Constituent multiplicity in cone jets

Fig. 9.10 compares the number of constituents in `AntiKt422` jets and in the final version of the `ConeTopo` jets, with the energy overlap removal strategy applied. The results are shown for $E_T > 1$ GeV topocluster thresholding, but similar results were observed for other thresholding options. A requirement on the jet p_T to be larger than 20 GeV was imposed to select typical jets that would pass the trigger. One can note that `AntiKt422` jets tend to have a slightly larger number of constituents. As this is more accentuated for softer jets, it is likely the result of `AntiKt422` finding a balanced boundary between two equally energetic

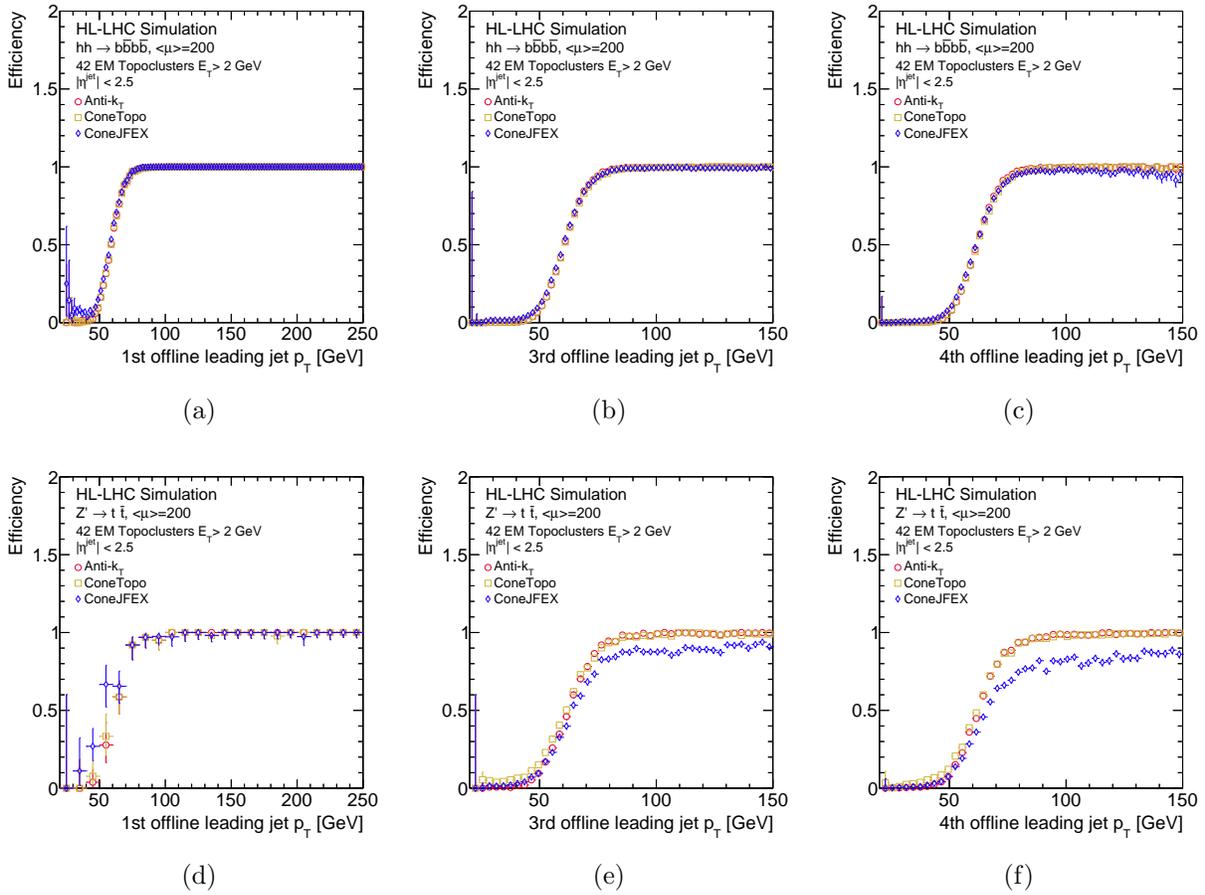


Figure 9.8: Comparison of one-, three-, and four-jet trigger efficiencies when the online jets are reconstructed as AntiKt422, ConeTopo, or ConeJFEX jets. The results are shown for di-Higgs (top) and $Z' \rightarrow t\bar{t}$ (bottom) signals and 2 GeV input topoclusters E_T thresholding.

jets, as opposed to the overlap removal step of the cone algorithm which always adopts a winner-take-all strategy, removing a greater number of constituents from the lower energy jet.

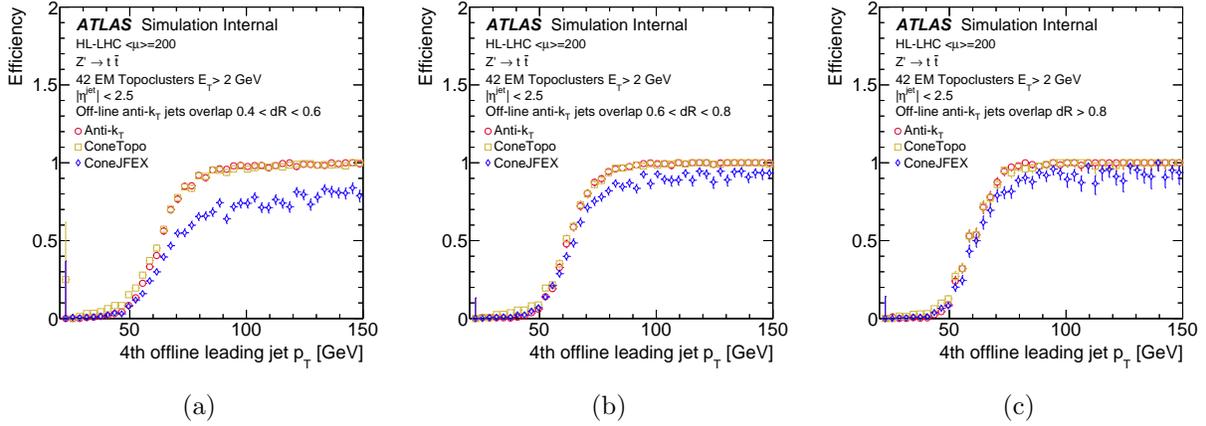


Figure 9.9: Comparison of one-, three-, and four-jet trigger efficiencies in different offline jet isolation bins when the online jets are reconstructed as *AntiKt422*, *ConeTopo*, or *ConeJFEX* jets. The results are shown for $Z' \rightarrow t\bar{t}$ signal and 2 GeV input topoclusters E_T thresholding.

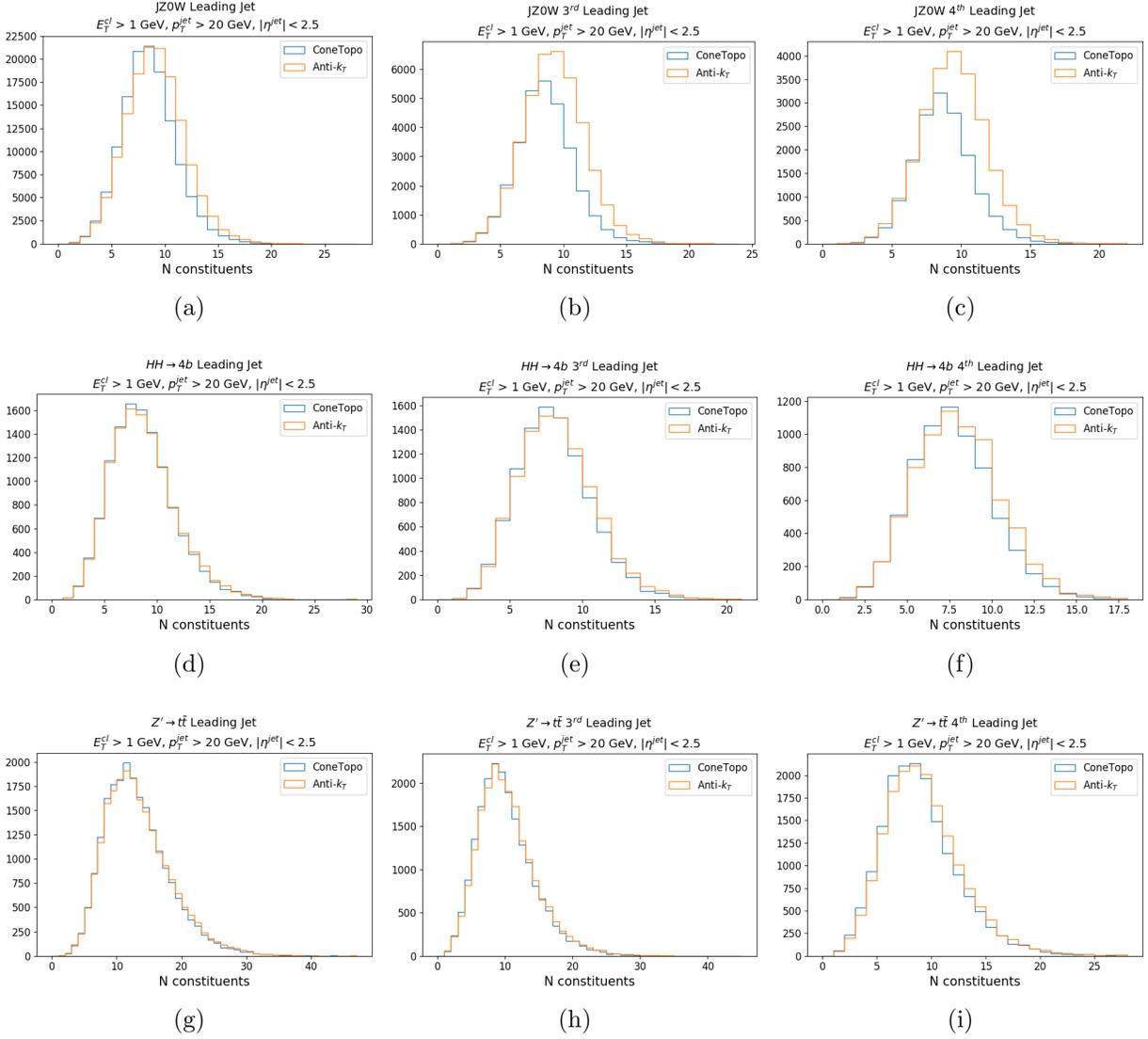


Figure 9.10: Number of constituents in the 1st, 3rd, and 4th leading jet. Comparing AntiKt422 and ConeTopo+EOR jets with $E_T > 1 \text{ GeV}$ topocluster thresholding and a minimum jet p_t of 20 GeV, for minimum bias (left), $Z' \rightarrow t\bar{t}$ (center), and di-Higgs (right) samples. Plots made by Garrit.