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 Science, Mathematics, and Engineering — Dual Major

2023

ABSTRACT

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

By

Maria Mazza

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 a full upgrade of the detector and trigger system, as well as innovative analysis techniques to enhance signal-to-background discrimination. In this work, these research directions are pursued on two fronts.

On one front, the development of a new deep neural network (DNN) jet tagger to compete with traditional analysis techniques is presented and its performance is investigated in the context of a search of new heavy resonances decaying into two standard model bosons (using the Run 2 ATLAS dataset). Two important models that predict such particles are the Heavy Vector Triplet model and the two-Higgs-doublet model, which address important open questions related to the Higgs sector: the naturalness problem and the possibility of an extended scalar sector.

On the other front, the HL-LHC upgrade of the ATLAS hardware-based trigger system will be discussed, with focus on the development of a new multi-jet trigger strategy targeting

$HH \to b\bar{b}b\bar{b}$, one of the flagship signatures of the HL-LHC upgrade. In particular, the physics performance and preliminary firmware simulation of a new jet reconstruction strategy will be presented, as well as the development of a DNN for pileup rejection, both meant to be deployed on fast FPGA hardware.

To Kévin.

E quindi uscimmo a riveder le stelle.
Inferno, XXXIV

.

ACKNOWLEDGMENTS

I would like to start by thanking my Ph.D. advisor Wade Fisher. First and foremost, thank you for the knowledge that you have shared with me. I will always value your teachings, especially on topics of machine learning and statistics, and appreciate your patience during those hour-long discussions, answering every one of my questions until the smallest doubt was clarified. Thank you also for the trust and support; for trusting me to go from student to teacher, by giving me the opportunity to prepare a lecture series for your machine learning class; for giving me the opportunity to work on motivating projects and then the freedom to expand on them and develop my own research direction; and, above all, 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 thank my committee members Wolfgang Kerzendorf, Dean Lee, Saiprasad Ravishankar, and Andreas von Manteuffel for their support and guidance during these past few years, as well as Matt Hirn, who left MSU. I would also like to extend a special thanks to Andreas and C.-P. Yuan for having been excellent professors of high energy theory. Last, but not least, I would like to 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 Raynolds, the undergraduate student I had the chance to mentor, for

working hard and helping me to grow as a mentor.

I would also like to thank the ATLAS collaborators I had the chance to work with, both on the analysis and on the trigger front. In particular, my work on the analysis front 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, Elisabetta, Giulia, 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.

J'aimerais remercier aussi 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 never let me down (even when I ask you to proof-read three 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

Chapter 1

Introduction

The Standard Model (SM) of particle physics 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 soon undergo a major upgrade that will bring a ten-fold increase in total integrated luminosity and that will open up new search channels, previously inaccessible cross-sections, and more precise tests of SM observables. However, high luminosity will also bring unprecedented levels of radiation and pileup. Extracting relevant physics in these extreme conditions will require a full upgrade of the detector and trigger system, as well as innovative techniques to enhance signal-to-background discrimination, both on real-time event selection, as well as in offline analyses. The research interests discussed in this thesis lie at the center of these new directions. The primary interest of this work was improving our understanding of the scalar sector, with a focus on the development of new machine learning strategies for both physics analysis and real-time data acquisition. In particular, I contributed to a search for new heavy resonances that could address important deficiencies of the SM related to the Higgs sector; in the development of new jet tagging deep learning algorithms; and in the high-luminosity LHC (HL-LHC) upgrade of the ATLAS Level-0 hardware-based trigger for the development of new reconstruction strategies and exploration

of new machine learning (ML) algorithms to be deployed on FPGA hardware.

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 (2HDM) [4], which assume the simplest extension of the scalar sector, by predicting the existence of two SU(2) complex doublets. Most of these models predict sizable couplings of the new particles to the SM Higgs and vector bosons, making such final states rich landscapes where to look for new physics.

I was one of the analyzers in the ATLAS search for such new heavy resonances decaying to two SM bosons (VV or VH) in semi-leptonic final states, where one can benefit from the clean signature of the leptonic decay of the vector boson, as well as the larger branching ratio of the hadronic decay. Due to the large multiplicity of 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 main tasks in the event selection of this type of heavy resonance searches is the correct identification of the hadronic decay: in the signal regions, we look for jets originating from a Higgs, W, or Z boson, while $t\bar{t}$ and V+jets are the primary SM background processes. My main contribution within the analysis was the development a new machine learning strategy for the

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 reconstruction strategies for the jets: two R=0.4 jets at low energies, one R=1.0 jet in the boosted regime. I trained two different 5-class deep neural networks (DNN), one for each reconstruction strategy, to output 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. DNN have been shown very effective and are used to separate and orthogonalize the VV and VH signal regions, where they bring significant sensitivity improvements with respect to previous efforts.

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 [6]. Models that predict a first-order phase transition predict large deviations from the SM prediction $\lambda_{HHH} = \frac{m_h^2}{2v}$. Measurement of the Higgs self-coupling would also be a direct test of electroweak symmetry breaking and of the shape of the Higgs potential, 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 HH one of the flagship signature for the HL-LHC.

The ability to measure HH production is also one of the main drivers of the HL-LHC trigger upgrades. Events are selected by the trigger according to a set of p_T thresholds for

specific objects, such as jets, and combinations thereof. In the gluon-fusion dominant HH production mode the destructive interference between the top-box and the triangle diagrams results in the latter contributing mostly to the low m_{HH} region. To retain sensitivity to λ_{HHH} it is therefore pivotal to keep 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 L0 that will 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. I developed the software simulation framework to study candidate algorithms and provide GT objects to do performance studies. I then developed a new jet reconstruction strategy for the firmware. The 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 and proven to be a viable option for the trigger. Despite the success, the p_T threshold might still not be good enough to access $HH \to b\bar{b}b\bar{b}$ events at low m_{HH}, where the p_T threshold is driven by the rate of pileup jets. 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. I developed a new DNN to identify pileup-like jets, simulated its application in the trigger. Preliminary results show that it can bring non-negligible improvements in terms of trigger rates.

Chapter 2

The Standard Model

Particle physics is the study of the fundamental particles of nature and their interactions. The idea that the Universe is composed of fundamental building blocks is as old as 400 BC, when the philosopher Democritus theorized that everything is composed of *àtomos* — "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 surrounded by 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 particles, but 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 very 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 confusion, mistakes, 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).

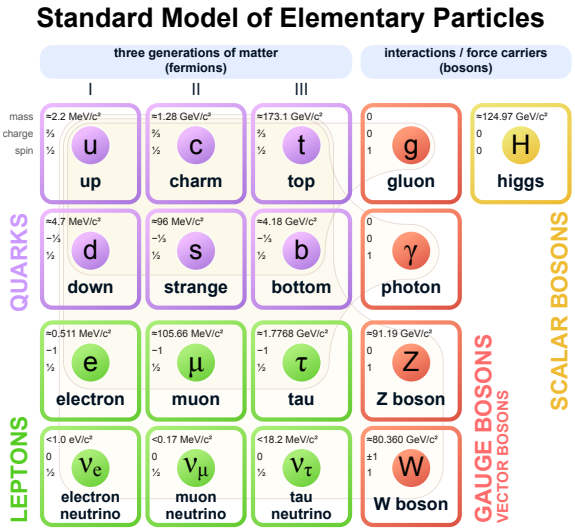


Figure 2.1: The Standard Model of elementary particles [11].

The SM[1] is the mathematical framework of high energy 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 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

[1] A more complete introduction and in-depth explanation of the topics discussed in this chapter can be found in several textbooks and reviews, notably [12–17]

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.

Each fermion particle has a corresponding anti-particle, with equal mass but opposite quantum numbers. 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 flavour 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. 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 colour 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 $\beta-$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. 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. At the time of the Big Bang the Universe was symmetrical, all particles were massless, and all forces were unified into one single fundamental force. Then, 10^{-11} seconds after the Big Bang, the Higgs field changed shape (underwent a phase transition), the symmetrical position became unstable, and the Universe decayed into a lower vacuum energy state that broke the symmetry. Upon the spontaneous symmetry breaking, 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, infinitesimal in size, 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. 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 [18]. However, his model predicted massless gauge bosons and fermions, in disagreement with the experimental observations. In the same years, the concept of a hidden symmetry capable of giving rise to observable phenomena that would break the symmetry, a phenomenon that took the name of "spontaneous symmetry breaking," was proposed. The idea was applied to local gauge theories and 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 was proposed in 1964 independently by Higgs [19], and Englert and Brout [20], opening the possibility of constructing a model with a hidden local gauge symmetry and massive gauge bosons.

The Higgs mechanism was applied to Glashow's gauge theory of the electroweak interaction by Weinberg [21] and Salam [22]. 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 1983 [2]. Lastly, as mentioned earlier, the Higgs boson was finally discovered in 2012 at CERN by the ATLAS [23] and CMS [24] Collaborations, ultimately confirming the validity of the SM, and followed shortly after by a 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 Sec. 2.1, fundamental concepts useful 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. 2.2 and a more detailed discussion of the Higgs sector in presented in 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. In classical physics the concepts of matter

[2] To which also followed a Nobel Prize in Physics in 1984 to Rubbia and Var der Meer for "their decisive contributions to the large project."

and force were clearly distinct, but while the nature of matter was intuitive, the concept of force was problematic. 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 already well established by the beginning of the last century (and later found to be already consistent with special relativity). However, soon after the successes of classical physics, a new realm of physics made its appearance, quantum mechanics.

A systematic quantum theory of fields started with Dirac's 1927 paper [25]. The solutions to Maxwell's equations in free space:

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

are transverse waves whose Fourier components behave like individual harmonic oscillator modes.Upon canonical quantization of the dynamical variables describing the individual modes[3] , Dirac showed the equivalent interpretation of the number of quanta of energy as the

[3]Dirac 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 $n = 0, 1, 2, ...$ can be interpreted as the number of quanta of energy, and a and $a^\dagger$ as annihilation and creation operators of the quanta, interpreted as the excitations of the electromagnetic field.

number of particles moving at the speed of light and satisfying Bose-Einstein statistics, i.e. the number of photons. In quantum mechanics photons are excitations of the electromagnetic field and can be created and annihilated as quanta of the field, but electrons and the other particles still obey the Schrodinger equation. However, 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 quantum mechanics 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. Quantum mechanics provides Heisenberg's uncertainty principle $\Delta E > \hbar c/\Delta x$, which states that the wave-particle duality of nature makes it impossible to measure energy and time simultaneously with arbitrary precision, so that the the more accurately one knows the position or time 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 starts not making sense anymore, as the uncertainty in the energy is now large enough to allow for a swarm of particle-anti-particle pairs to surround the particle. A new formalism other than QM 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 declassed 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. 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. Of particular importance in these developments was Yukawa's paper in 1935, in which he showed how the interaction between particles could be shown to proceed via the exchange of virtual quanta of the field, mediators of the force [16].

The classical electromagnetic field can then be described at the quantum level as a constant creation and annihilation of real and virtual photons, the quanta of the electromagnetic interaction. Similarly, the electron and its anti-particle, the positron, can be viewed as the quanta of the electron-positron field. There is a duality then between the particle and the field. Both pictures can be used to describe the system, but it turns out that fields are the natural way to describe mathematically what is happening at these small scales. Hence, quantum field theory turned out to be the language in which to describe the fundamental laws of nature. 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

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

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 only[4]. The principle of least action requires the variation of the action δS to be zero for small fluctuations of the fields $\phi(x) \to \phi(x) + \delta\phi(x)$. Imposing this requirement brings to the Euler-Lagrange (EL) equations of motion for a field [12],

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

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

[4]Higher order spatial derivatives could enter in principle, but this requirement allows one to treat space and time on equal footing and is sufficient to describe the physics observed by experiment. Ostrogradsky theorem proves that a Lagrangian with higher order time derivatives gives energy densities not bounded from below, which in a QFT are unphysical.

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. 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 locality [5] and renormalizability[6], allows one to identify all the allowed terms and the predictive power of the model can then be tested against experiment. In the following section, it will be shown why symmetries play such a fundamental role 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

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

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

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

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

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

From the EL equations, the first term vanishes. Therefore, the system has a conserved current $\partial^\mu J_\mu = 0$ and a corresponding conserved charged [7] given by,

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

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. As an example, consider the Lagrangian density describing a free Dirac fermion

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

This Lagrangian is invariant under continuous rotations of the phase of $\psi(x)$ as $\psi(x) \rightarrow$

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

$e^{i\alpha}\psi(x)$. 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)

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

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 $A_\mu(x) \to A_\mu + \partial_\mu f(x)$, for any function $f(x)$:

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

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 class of configurations related to each other by a symmetry group. 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

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

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

$$A_\mu(x) \to A'_\mu = A_\mu + \partial_\mu f(x), \tag{2.11}$$

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

$$\psi(x) \to \psi(x)' = e^{iqf(x)}\psi(x), \tag{2.12}$$

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

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

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

Thus, while the global phase of the field is purely convention-dependent, once a choice is made, this has to be taken at 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:

$$A_\mu(x) \xrightarrow{\mathrm{U}(1)} A'_\mu = A_\mu + \partial_\mu f(x) \tag{2.15}$$

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

$$D_\mu \psi(x) = \partial_\mu \psi(x) + ieA_\mu \psi(x) \tag{2.16}$$

which has the required property of transforming like the field itself canceling exactly the contribution from $\bar{\psi}(x) \to \bar{\psi}'(x)$. Note that replacing the ordinary derivative $\partial_\mu \psi(x)$ with the covariant derivative $D_\mu \psi(x)$ is equivalent to the introduction of 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 $1/2m^2 A_\mu A^\mu$ is forbidden as it would break the gauge symmetry, hence the photon is predicted to be massless. To allow the new vector field to propagate in space, one adds a gauge invariant kinetic term, which corresponds to the free field Lagrangian. The final Lagrangian

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

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 in the form of a self-energy term. 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 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 $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ 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. 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 forces that couple the spins of the electrons are invariant under rotation in space. 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. Because of the degeneracy of the ground state, 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 $< 0|\psi(x)|0 >=< 0|V^{\mu}(x)|0 >= 0$, so that in order to break the symmetry a scalar field $\phi(x)$ has to be introduced.

In the following, 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. Specifically, Goldstone's theorem states that for every spontaneously broken continuous symmetry, the theory must contain a massless particle. 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 illustrate 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 Glashow, 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

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

invariant under the global U(1) phase transformation

$$\phi(x) \to \phi'(x) = e^{i\alpha}\phi(x) \qquad \phi(x)^* \to \phi^{*'}(x) = e^{-i\alpha}\phi^*(x) \tag{2.19}$$

The potential of this Lagrangian is

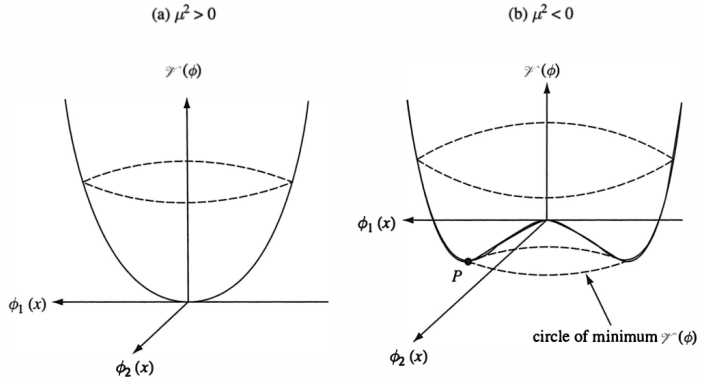


Figure 2.2: Potential for Eq. (2.20) with $\lambda > 0$ [13].

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

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 $\phi(x) = 0$ and the Lagrangian is that of a free complex Klein Gordon field, 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,

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

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

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

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

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

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

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 σ 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, along the 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

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

The Lagrangian has the same potential and is invariant under the coupled U(1) gauge transformations, similarly to Eq. (**??**). Performing the same substitution into $\mathcal{L}$, one obtains

$$\begin{aligned}
\mathcal{L}(x) = &\frac{1}{2}\partial^\mu\sigma(x)\partial_\mu\sigma(x) - \frac{1}{2}(2\lambda v^2)\sigma^2(x) && (2.26)\\
&- \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + \frac{1}{2}(qv)^2A_\mu(x)A^\mu(x) && (2.27)\\
&+ \frac{1}{2}\partial^\mu\eta(x)\partial_\mu\eta(x) && (2.28)\\
&+ qvA^\mu(x)\partial_\mu\eta(x) + \text{interaction terms} && (2.29)
\end{aligned}$$

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 field is spoiled by 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 looks like

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

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

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

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 unifies the description of electromagnetism, weak interactions and strong interactions in the language of a gauge theory. 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 $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ invariant electroweak theory and the $\mathrm{SU}(3)_C$ gauge theory of quantum chromodynamics (QCD), resulting in an $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \times \mathrm{SU}(3)_C$ 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

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

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

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

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,

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

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

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 interaction 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:

$$\begin{array}{llll}
L_1 = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, & l_{R_1} = e_R^- & Q_1 = \begin{pmatrix} u \\ d \end{pmatrix}_L, & u_{R_1} = u_R, \quad d_{R_1} = d_R \\
L_2 = \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, & l_{R_2} = \mu_R^- & Q_2 = \begin{pmatrix} c \\ s \end{pmatrix}_L, & u_{R_2} = c_R, \quad d_{R_2} = s_R \\
\underbrace{L_3 = \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}, \quad l_{R_3} = \tau_R^-}_{\text{Leptons}} & & \underbrace{Q_3 = \begin{pmatrix} t \\ b \end{pmatrix}_L, \quad u_{R_3} = t_R, \quad d_{R_3} = b_R}_{\text{Quarks}} &
\end{array} \tag{2.37}$$

The leptons and quarks are grouped in the three generations and the left-handed components are combined into SU(2) doublets. In the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ electroweak theory gauge invariance determines the conservation of the weak hypercharge Y and of the weak isospin I, while spontaneous symmetry breaking of $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \to \mathrm{U}(1)_{EM}$ brings to the con-

servation 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 $I_i^L = \frac{1}{2}\sigma_i$, where σ_i are the Pauli 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

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

Note that 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 of the form

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

mix left and right-handed components, which violates SU(2) gauge invariance.

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

and transform from one colour state to another under the $SU(3)_C$ group as triplets

$$q_i = \begin{pmatrix} q_i^r \\ q_i^g \\ q_i^b \end{pmatrix} \tag{2.41}$$

with $C = r, g, b$ representing the colour charge. The lepton fields on the other hand are colour singlets and do not interact via the strong force. Only colour singlet combinations are observed in nature, as baryons and mesons

$$B = \frac{1}{\sqrt{6}}\epsilon^{\alpha\beta\gamma}|q_\alpha q_\beta q_\gamma >, \tag{2.42}$$

$$M = \frac{1}{\sqrt{3}}\delta^{\alpha\beta}|q_\alpha q_\beta > . \tag{2.43}$$

This is known as *colour confinement*: quarks are confined within colour-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 colour 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 $\tau^a (a = 1, 2, 3)$ 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

$\mathrm{SU}(3)_C$ group are the set of unitary 3×3 matrices with determinant one, and the generators of the algebra are the matrices $T^a = \frac{1}{2}\lambda_a (a = 1, 2, \ldots, 8)$, where λ_a are the Gell-Mann matrices. The matrices T^a satisfy the commutation relations $[T^1, T^b] = if^{abc}T_c$, where f^{abc} are the SU(3) structure constants, which are real and totally antisymmetric.

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

$$L(x) \to L'(x) = e^{i\alpha_a(x)T^a + i\beta(x)Y} L(x), \tag{2.44}$$

$$R(x) \to R'(x) = e^{i\beta(x)Y} R(x). \tag{2.45}$$

Similarly, in colour space the fermion fields transform as

$$\psi_q(x) \to \psi'_q(x) = e^{i\alpha_a(x)\frac{\lambda^a}{2}}. \tag{2.46}$$

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^i_μ 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^i_μ. The new fields, with their appropriate coupled gauge transformations, are introduced via the covariant derivative

$$D_\mu \psi = \left(\partial_\mu - ig_s T_a G^a_\mu - ig_2 T_a W^a_\mu - ig_1 \frac{Y_q}{2} B_\mu \right) \psi. \tag{2.47}$$

with g_s, g_2, and g_1 the coupling constants of $\mathrm{SU}(3)_C$, $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$, 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

$$\mathcal{L} = \sum_j \bar{\Psi}_L^j i\gamma^\mu D_\mu^L \Psi_L^j + \sum_{j,\sigma} \bar{\psi}_{R\sigma}^j i\gamma^\mu D_\mu^R \psi_{R\sigma}^j \tag{2.48}$$

$$- \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{2.49}$$

where the gauge invariant field strengths tensors are given by

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c \tag{2.50}$$

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g_2 \epsilon^{abc} W_\mu^b W_\nu^c \tag{2.51}$$

$$G_{\mu\nu}^a = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{2.52}$$

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. This theory also turns out to be not renormalizable. To preserve the gauge invariance and renormalizability of the theory 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 Higgs mechanism is introduced in the SM to provide a way for the gauge bosons to acquire mass. The electroweak SM Lagrangian is invariant under the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ 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. A good guess is to attempt to break the SU(2) gauge symmetry. The simplest choice is to add a complex scalar field $\phi(x)$ (providing four degrees of freedom) that is an isospin doublet of SU(2) with hypercharge $Y_\phi = +1$,

$$\Phi = \begin{pmatrix} \phi_+(x) \\ \\ \phi_0(x) \end{pmatrix} \tag{2.53}$$

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 component of the doublet have $Q = 1$, while the lower component has $Q = 0$. 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 $\mathrm{SU}(2) \times \mathrm{U}(1)$ invariant and the new term $\mathcal{L}_\phi$ can be made invariant by introducing the EW covariant derivative from Eq. (2.47),

$$\mathcal{L}_\Phi = (D^\mu \Phi(x))^\dagger [D_\mu \Phi(x)] - \mu^2 \Phi(x)^\dagger \Phi(x) - \lambda (\Phi^\dagger(x)\Phi(x))^2. \tag{2.54}$$

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 $< \Phi >_0$, 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,

$$\Phi_0 = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \quad \text{with } v = \left(-\frac{\mu^2}{\lambda}\right)^{1/2}. \tag{2.55}$$

One can parametrize the scalar field in terms of its deviations from the vacuum state Φ_0 and move to the unitary gauge,

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \eta_1(x) + i\eta_2(x) \\ v + \sigma(x) + i\eta_3(x) \end{pmatrix} \longrightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \sigma(x) \end{pmatrix} \tag{2.56}$$

As observed with the Higgs model, the extra degrees of freedom of the massless $\eta(x)$ bosons are unphysical. These extra degrees of freedom can be "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 neu-

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

$$\phi_0 \to \phi_0 = \exp[i(I_3^W + Y)\beta]\phi_0 = 1 \cdot \phi_0. \tag{2.57}$$

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

Before SSB the SM is an $\mathrm{SU}(3) \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge invariant theory. After SSB the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry is broken, but a new $\mathrm{U}(1)_{EM}$ 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 $\mathrm{U}(1)_{EM}$ 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,

$$|D_\mu \Phi|^2 = \left| \left(\partial_\mu - i g_2 \frac{\tau_a}{2} W_\mu^a - i g_1 \frac{1}{2} B_\mu \right) \Phi \right|^2 \tag{2.58}$$

$$= \frac{1}{2} \left| \begin{pmatrix} \partial_\mu - \frac{i}{2}(g_2 W_\mu^3 + g_1 B_\mu) & -\frac{1 g_2}{2}(W_\mu^1 - i W_\mu^2) \\ -\frac{i g_2}{2}(W_\mu^1 + i W_\mu^2) & \partial_\mu + \frac{i}{2}(g_2 W_\mu^3 - g_1 B_\mu) \end{pmatrix} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \right|^2 \tag{2.59}$$

$$= \frac{1}{2}(\partial_\mu H)^2 + \frac{1}{8} g_2^2 (v + H)^2 |W_\mu^1 + i W_\mu^2|^2 + \frac{1}{8}(v + H)^2 |g_2 W_\mu^3 - g_1 B_\mu|^2. \tag{2.60}$$

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, where the gauge bosons are multiplied by H^2 or vH factors, provide interaction terms between the gauge bosons and the Higgs boson and will be discussed later.

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

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp i W_\mu^2), Z_\mu = \frac{g_2 W_\mu^3 - g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}}, A_\mu = \frac{g_2 W_\mu^3 + g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}}, \tag{2.61}$$

where the first two components of the W_μ field are rewritten as linear combinations of two new vector bosons $W_\mu^\pm$ with electric charge ± 1, while the field B_μ and the third field component W_μ^3 are rewritten as a linear combination of two electrically neutral fields Z_μ and

A_μ orthogonal to each other and related to the original fields via the rotation matrix,

$$B_\mu(x) = -\sin\theta_W Z_\mu(x) + \cos\theta_W A_\mu(x) \tag{2.62}$$

$$W^3_\mu(x) = \cos\theta_W Z_\mu(x) + \sin\theta_W A_\mu(x). \tag{2.63}$$

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

$$\mathcal{L}_\Phi \supset \frac{1}{8}g_2^2 v^2 |W^1_\mu + iW^2_\mu|^2 + \frac{1}{8}v^2|g_2 W^3_\mu - g_1 B_\mu|^2 \tag{2.64}$$

$$\rightarrow (\frac{1}{2}vg_2)^2 W^+_\mu W^{-\mu} + \frac{1}{2}(\frac{1}{2}v\sqrt{g_2^2+g_1^2})^2 Z_\mu Z^\mu. \tag{2.65}$$

and the mass terms can be read off directly

$$M_W = \frac{1}{2}vg_2, \quad M_Z = \frac{1}{2}v\sqrt{g_2^2+g_1^2} \tag{2.66}$$

while the A_μ boson remains massless.

One recognizes then 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 and can be expressed in terms of the W and Z boson masses as

$$\sin^2\theta_W = 1 - \cos^2\theta_W = 1 - \frac{M_W^2}{M_Z^2}. \tag{2.67}$$

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

$$V = \frac{\mu^2}{2}(v+H)^2 + \frac{\lambda}{4}(v+H)^4. \tag{2.68}$$

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

$$\mathcal{L}_H = \frac{1}{2}(\partial_\mu H)^2 - \lambda v^2 H^2 - \lambda v H^3 - \frac{\lambda}{4}H^4. \tag{2.69}$$

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

$$M_H = \sqrt{2\lambda v^2}. \tag{2.70}$$

The remaining terms represent the Higgs self-interactions can be read off directly[8]

$$g_{HHH} = (3!)i\lambda v = 3i\frac{M_H^2}{v}, \tag{2.71}$$

$$g_{HHHH} = (4!)i\frac{\lambda}{4} = 3i\frac{M_H^2}{v^2}. \tag{2.72}$$

The vacuum expectation value is fixed by the values of the vector boson mass terms through the relation

$$M_W = \frac{1}{2}g_2 v = \left(\frac{\sqrt{2}g^2}{8G_\mu}\right)^{1/2} \implies v = \frac{1}{(\sqrt{2}G_\mu)^{1/2}} = 246.22\,\text{GeV}, \tag{2.73}$$

[8] According to the Feynman rules, the couplings for these vertices are obtained by multiplying the coupling term from the Lagrangian by a factor of $-i$ and by a factor $n!$, where n is the number of identical particles interacting at the vertex.

where G_μ is the Fermi constant. On the other hand, 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 $m_H = 125.11 \pm 0.11\,\text{GeV}$ [26]. The Higgs self-coupling is instead still unmeasured, although increasingly stringent limits are being set on its allowed value. The value of the Higgs-self coupling determines the shape of the Higgs potential, in turn connected to several fundamental questions related to cosmology. 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 (HL-LHC) upgrade (see Sec. 3.0.1).

The Yukawa interactions

Fermion mass terms cannot be included *ad hoc* in the Lagrangian because they would violate gauge invariance. At the same time, they do not appear via the Higgs mechanism, like the gauge boson masses do. However, the introduction of the Higgs field still proved useful, as it provided a new way to add mass terms via new couplings between the fermion fields and the Higgs field.

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

$$\mathcal{L}_{Yukawa} = -(Y_l)^{ij}\bar{L}^i_L \Phi l^j_R - (Y_d)^{ij}\bar{Q}^i_L \Phi d^j_R - (Y_u)^{ij}\bar{Q}^i_L \tilde{\Phi} u^j_R + h.c. \tag{2.74}$$

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 for the three generation of quark fields is being used:

$$Q_L^i = \begin{bmatrix} u_L^i \\ d_L^i \end{bmatrix} = \left(\begin{bmatrix} u_L \\ d_L \end{bmatrix}, \begin{bmatrix} c_L \\ s_L \end{bmatrix}, \begin{bmatrix} t_L \\ b_L \end{bmatrix} \right) \tag{2.75}$$

with the corresponding generic 3×3 matrices $Y^q (q = u, d)$. For a fixed choice of quark flavor i and j, after SSB and rotation to the unitary gauge, the following terms appear:

$$\mathcal{L}_{Yukawa,q} = y_d^{ij} \bar{d}_L^i d_R^j (v + H) + y_u^{ij} \bar{u}_L^i u_R^j (v + H). \tag{2.76}$$

These look like candidates for fermion mass terms and fermion coupling terms to the Higgs field. However, the matrices Y^q 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 flavour 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

$$M_{\text{diag}}^q = V_L^{q\dagger} M^q V_R^q, \qquad \text{where} \qquad m_{ij} = y_{ij} \frac{v}{\sqrt{2}}. \tag{2.77}$$

Upon diagonalization only the terms 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,

$$\mathcal{L}_{\text{Yukawa, tb}} = -\frac{1}{\sqrt{2}}(\hat{y}_b \bar{b}_L b_R + \hat{y}_t \bar{t}_L t_R)(v + H) \tag{2.78}$$

$$= \underbrace{-\hat{y}_b \frac{v}{\sqrt{2}} \bar{b}_L b_R - \hat{y}_t \frac{v}{\sqrt{2}} t_L t_R}_{\text{Mass terms}} \underbrace{- \hat{y}_b \frac{1}{\sqrt{2}} \bar{b}_L b_R - \hat{y}_t \frac{1}{\sqrt{2}} t_L t_R}_{\text{Higgs couplings}}, \tag{2.79}$$

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

$$m_f = \hat{y}_f \frac{v}{\sqrt{2}} \quad and \quad g_{Hff} = i\frac{m_f}{v}. \tag{2.80}$$

Note that the couplings are proportional to the corresponding fermions mass and conserve flavor.

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:

$$\bar{\Psi} \not{D} \Psi \supset \frac{g_2}{\sqrt{2}} \bar{u}_L i \gamma_\mu d_L W^\mu \tag{2.81}$$

$$\rightarrow \frac{g_2}{\sqrt{2}} \bar{u}_L i \gamma_\mu (V_L^u V_L^{d\dagger}) d_L W^\mu = \frac{g_2}{\sqrt{2}} (\bar{u}_L, \bar{c}_L, \bar{t}_L) \gamma_\mu V_{CKM} \begin{bmatrix} d_L \\ s_L \\ b_L \end{bmatrix} W^\mu, \tag{2.82}$$

$$V_{CKM} \equiv V_{uL} V_{dL}^\dagger = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \tag{2.83}$$

The three matrices M^u, M^d, and V_{CKM} cannot be diagonalized simultaneously, resulting in quark flavour violating interactions. The matrix V_{CKM} is the CKM matrix, named after Cabbibo [27], and Kobayashi and Maskawa [28]. 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 [29]

$$|V_{CKM}| = \begin{bmatrix} 0.97 & 0.22 & 0.004 \\ 0.22 & 0.97 & 0.04 \\ 0.009 & 0.04 & 1 \end{bmatrix} \tag{2.84}$$

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:

$$g_{Hff} = i\frac{m_f}{v}, \tag{2.85}$$

$$g_{HVV} = -2i\frac{M_V^2}{v}g_{\mu\nu}, \tag{2.86}$$

$$g_{HHVV} = -2i\frac{M_V^2}{v^2}g_{\mu\nu} \tag{2.87}$$

The tree level couplings of the Higgs boson to fermion mass eigenstates are flavor diagonal and CP conserving. They are 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 it couples to the gluon, as it does not carry colour 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 the 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. Nonetheless, while 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 five 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 an extra order of magnitude is needed to explain the electron mass. The structure of fermion masses originates solely from the Yukawa couplings, which are added as free parameters in the SM. In fact, 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

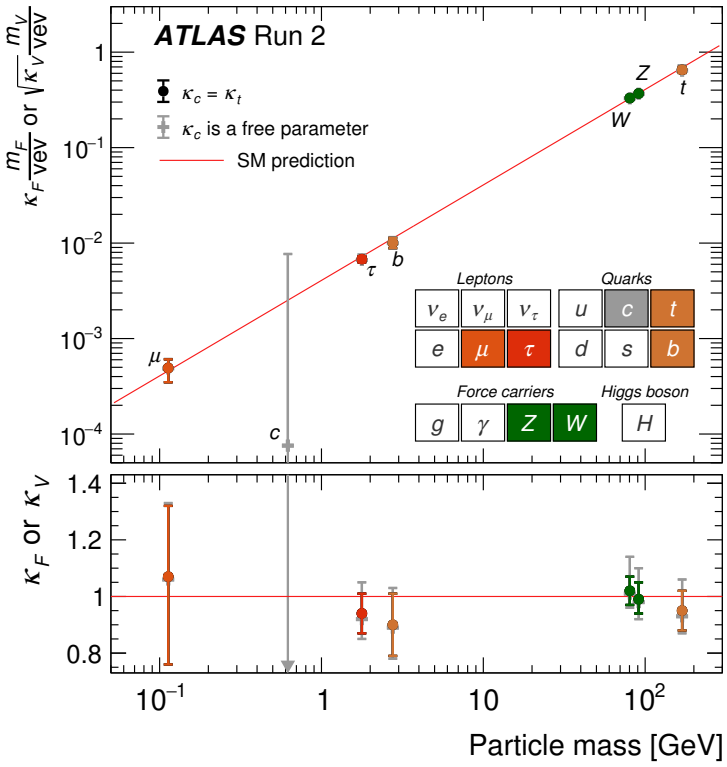


Figure 2.3: Reduced Higgs boson coupling strength modifiers and their uncertainties. [30].

of the four fundamental forces of nature. This makes physicists regard the SM as an effective low energy theory, valid only up to a certain energy scale. This scale has 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 [31] and gravitational lensing (the curvature of space-time near gravitating mass) [32], 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 [33, 34] 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 flavours 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 [35] and the measurement of the anomalous magnetic moment of the muon (Muon $g-2$) [36].

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 in BSM physics scenario 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 importance of the Higgs boson in the SM model makes many consider it the key to explain many of the 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 parameters [1], whose values are set by experimental measurements. 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, showing the centrality of the Higgs boson in the SM and its phenomenology.

The scalar sector of the SM remains greatly unexplored experimentally, as the Higgs self-coupling and the shape of the its 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.

[1] Assuming three generations and excluding neutrino masses

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, such as the Heavy Vector Triplet (HVT) model. The open question remains also of whether there is only one Higgs boson or if there might be an extended scalar sector. Two important theoretical frameworks used to search for answers to these questions are the HVT model, which predicts the existence of a new SU(2) triplet, and the Two-Higgs-Doublet Model, which predicts an extended scalar sector. These models are relevant for the analysis discussed in Chap. 7.

3.0.1 The Higgs self-coupling

The Higgs potential is fundamentally connected to the origin of electroweak symmetry breaking (EWSB), but while the vacuum expectation value and the Higgs boson mass have been measured with high precision, providing the minimum and the slope at the minimum of the potential, the Higgs self-coupling λ remains unmeasured. The self-coupling λ is responsible not only for the mass of the Higgs boson itself, but also for the shape of the potential, and it is therefore connected to questions regarding the stability of the Universe and the observed baryon asymmetry.

Quantum corrections are observed to affect the shape of the Higgs potential. The measured values of the Higgs and top quark mass indicate that, when running the Higgs self-coupling to high renormalization scales, λ turns negative at a scale $\Lambda \sim 10^{10}\,\text{GeV}$ [37], as shown in Fig. 3.1. 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 the 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?

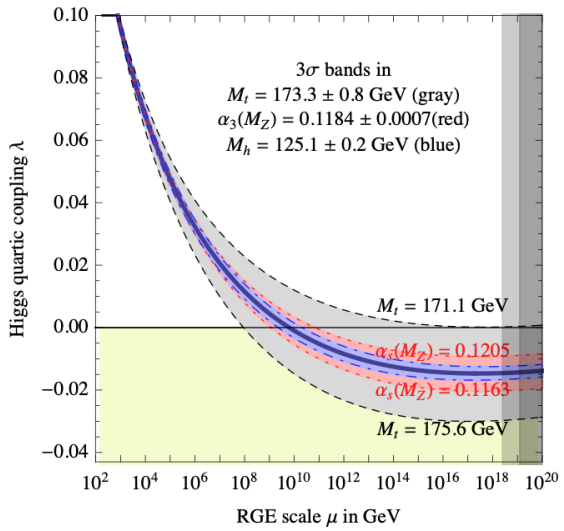


Figure 3.1: Renormalization group evolution of Higgs quartic coupling λ [38].

The quartic 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 occurs during the electroweak phase transition, which is the process by which the Higgs field acquired a vacuum expectation value. The Universe after the Big Bang is thought to have started in the unbroken phase, where the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge invariance was manifest. As the temperature cooled down below $T \lesssim 100\,\mathrm{GeV}$, the Higgs field settled into one of the absolute minima of the potential, spontaneously breaking the original symmetry. In this broken phase, only the $\mathrm{U}(1)_{EM}$ remained. Electroweak baryogenesis is predicted to have taken place during

this phase transition. However, for baryon creation to take place successfully, the phase transition has to be first order, where the departure from thermal equilibrium is violent. Unfortunately, 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). While the sensitivity of the current ATLAS and CMS analyses is limited by statistical uncertainties, the HL-LHC is expected to provide sufficient sensitivity to SM di-Higgs production and a measurement of the Higgs trilinear self-coupling. Di-Higgs production and the experimental challenges related to this measurement will be discussed in more detail in Chap. 8.

3.0.2 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. 3.2. The divergent integrals can be 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

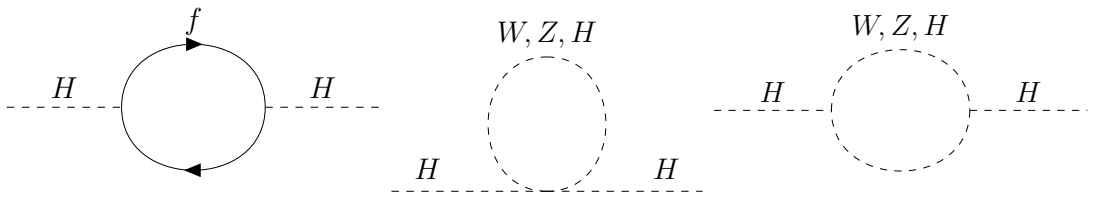


Figure 3.2: Feynman diagrams contributing to the one-loop corrections to the Higgs boson mass in the SM.

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:

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

where m_H^0 is the bare mass from the unrenormalized Lagrangian. The quadratic, rather than logarithmic, divergence 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 $\Lambda \sim 10^{19}\,\text{GeV}$, the renormalized mass square m_H^2 would result from a finely-tuned cancellation of 34 digits between the bare Higgs boson mass square and the counter-term proportional to Λ^2 in order to obtain the observed Higgs mass square of $(\sim 10^2\,\text{GeV})^2$ [37]. This type of cancellation is considered unnatural and is referred to as the *naturalness* problem. This brings about 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.

3.0.3 The Heavy Vector Triplet model

New vector bosons are a common feature 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 $\mathrm{SU}(3)_C \times \mathrm{SU}_L(2) \times \mathrm{U}(1)_Y$ 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 [39, 40] and, from a theoretical point of view, they appear in well-known extensions of the SM, including Grand Unified Theories, Little Higgs [41] models, theories of extra dimensions, and strongly coupled models such as composite Higgs models [42, 43].

While these models are theoretically consistent, there is an inverse problem that makes it 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 not sensitive to all free parameters of the model, but only to the mass and couplings of the predicted new particles, determining the available decay channels, and the strength and location of the peak on the mass spectrum. The HVT [3, 39] model provides a simplified framework with which one can test only the relevant phenomenological parameters c: an experimental search can determine the likelihood between the data and the general model as $L(\vec{c})$; the phenomenological parameters can then be expressed analytically in terms of the parameters of the explicit theory as $\vec{c}(\vec{p})$. 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 $\mathrm{SU}(2)_L$ and with zero hypercharge,

with the charged eigenstates

$$V_\mu^\pm = \frac{V_\mu^1 \mp i V_\mu^2}{\sqrt{2}}, \quad V_\mu^0 = V_\mu^3 \tag{3.2}$$

and the Lagrangian describing the new fields and their interactions with SM particles given by,

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

$$+ i g_V c_H V_\mu^a H^\dagger \tau^a \overleftrightarrow{D}^\mu H + \frac{g^2}{g_V} c_F V_\mu^a J_F^{\mu a} \tag{3.4}$$

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

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

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

where g is the $\mathrm{SU}(2)_L$ gauge coupling. The second line of the equation contains the interactions of V with the Higgs boson

$$i H^\dagger \tau^a \overleftrightarrow{D}^\mu H = i H^\dagger \tau^a D^\mu H - i D^\mu H^\dagger \tau^a H, \tag{3.7}$$

and with the SM left-handed fermions,

$$J_F^{\mu,a} = \sum_f \bar{f}_L \gamma^\mu \tau^a f_L, \tag{3.8}$$

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, 39], 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 new charged and neutral bosons are quasi-degenerate and their masses are assumed to be above $\approx 1\,\text{TeV}$. Thanks to the 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 two 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 $g_F = g^2/g_V c_F$. 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:

$$\mathcal{L}_\pi \ni \frac{g_V c_H}{2} V_\mu^a (\partial^\mu h \pi^a - h \partial^\mu \pi^a + \epsilon^{abc} \pi^b \partial^\mu \pi^c) \tag{3.9}$$

where $\pi_\pm$ and π_0 are the massless Goldstone bosons which reappear in this basis and which, 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 vector bosons and, in particular, the resonance production via vector boson fusion and its 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 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 free parameters M_V and g_V. The model described in Ref. [**?**] 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, so that Model A predicts comparable branching ratios into bosons and fermions. In model A, the total width of the resonance decreases with increasing g_V because of the overall suppression g^2/g_V. The Minimal Composite Higgs Model [**?**] is taken as representative of a strongly

coupled model, referred to as *Model B*, where relatively large $g_V \gtrsim 3$ 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 di-boson 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 $\Gamma/M >> 0.1$. 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. Vector boson fusion can be enhanced by suppressing the coupling of the HVT bosons to fermions. This is done in Model C, where $g_F = 0$ and $g_H = 1$.

3.0.4 The Two-Higgs-Doublet Models

One of the most stringent constraints on the SM are electroweak precision measurements. However, 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 Suppersymmetric Standard Model (MSSM) [44], axions models [45], and baryogenesis models [46]. In particular, one important class of models, called the Two-Higgs-Doublet Models (2HDMs) [4] studies the addition of a new scalar doublet [2].

Flavour Changing Neutral Currents

The introduction of two Higgs doublets in the Yukawa Lagrangian allows for flavour changing

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

neutral currents (FCNCs). Considering the quark terms only, $\mathcal{L}_{\text{Yukawa}}$ is now:

$$\mathcal{L}_{\text{Yukawa}} \supset -\sum_{k=1,2} (Y_d)^{ij,k} \bar{Q}_L^i \Phi_k d_R^j - (Y_u)^{ij,k} \bar{Q}_L^i \tilde{\Phi}_k u_R^j + h.c. \tag{3.10}$$

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 up-type quarks. Upon SSB the mass terms appear as:

$$\mathcal{L}_{\text{Yukawa}} \supset -\sum_{k=1,2} \bar{u}_L^i \underbrace{(Y_u)^{ij,k} \frac{v^k}{\sqrt{2}}}_{M_{ij}^u} u_R^j \tag{3.11}$$

The mass matrix is then:

$$M_{ij}^u = y_u^{ij,1} \frac{v_1}{\sqrt{2}} + y_u^{ij,2} \frac{v_2}{\sqrt{2}}. \tag{3.12}$$

Without further restrictions, Y_u^1 and Y_u^2 are not simultaneously diagonalizable. While in the SM the diagonalization of the mass matrices implies the diagonalization of the couplings of the Higgs boson to the fermions, this is not the case anymore, are flavour changing couplings are present at tree level.

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 Pascos-Glashow-Weinberg theorem [47, 48], 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. For

instance, this can be enforced by requiring the discrete symmetry $\Phi_1 \to -\Phi_1$. In the *Type II* 2HDM, all up-type quarks couple to one Higgs doubletand 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 [49].

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

$$V = m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 \left(\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1 \right) + \frac{\lambda_1}{2} \left(\Phi_1^\dagger \Phi_1 \right)^2 + \frac{\lambda_2}{2} \left(\Phi_2^\dagger \Phi_2 \right) \tag{3.13}$$

$$+ \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \frac{\lambda_5}{2} \left[\left(\Phi_1^\dagger \Phi_2 \right)^2 + \left(\Phi_2^\dagger \Phi_1 \right)^2 \right], \tag{3.14}$$

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

$$\Phi_a = \begin{bmatrix} \phi_a^+ \\ (v_a + \rho_a + i\eta_a)/\sqrt{2} \end{bmatrix}, a = 1, 2. \tag{3.15}$$

Upon spontaneous symmetry breaking, the neutral components of the two doublets ac-

quire vevs v_1 and v_2:

$$< \Phi_1 >_0 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v_1 \end{bmatrix}, \quad < \Phi_2 >_0 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v_2 \end{bmatrix}. \tag{3.16}$$

where the observed value of the vev requires $v^2 = v_1^2 + v_2^2 = (246\,\mathrm{GeV})^2$. Note that the three vevs are related as,

$$v_1 = v\cos\beta, \quad v_2 = v\sin\beta, \quad \text{and} \quad \tan\beta = \frac{v_2}{v_1}. \tag{3.17}$$

where the angle β parametrizes the contributions of each doublet to the observed SM vev. The physical mass eigenstates are obtained by the rotation matrices,

$$\begin{bmatrix} G^0 \\ A \end{bmatrix} = \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\cos\beta \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \tag{3.18}$$

$$\begin{bmatrix} h \\ H \end{bmatrix} = \begin{bmatrix} \sin\alpha & -\cos\alpha \\ -\cos\alpha & -\sin\alpha \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \tag{3.19}$$

$$\begin{bmatrix} G^\pm \\ H^\pm \end{bmatrix} = \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\sin\beta \end{bmatrix} \begin{bmatrix} \phi^\pm \\ \phi^\pm \end{bmatrix} \tag{3.20}$$

where the angle β reappears as the angle that diagonalizes the mass squared matrices of the charged scalars and of the pseudoscalars, while a new angle parameter α diagonalizes the matrices of the neutral scalar fields. 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 obtained with the following field redefinition:

$$H_1 = \cos\beta\Phi_1 + \sin\beta\Phi_2 \tag{3.21}$$

$$H_2 = -\sin\beta\Phi_1 + \cos\beta\Phi_2. \tag{3.22}$$

From the relation $v^2 = v_1^2 + v_2^2$, one gets $< H_1 >_0 = (v_1\cos\beta + v_2\sin\beta)/\sqrt{2} = v$ and $< H_2 >_0 = 0$. Hence, in the Higgs basis one of the two doublets is entirely responsible for the SM vev. In this basis, the physical neutral CP-even states are given by

$$H = (\sqrt{2}ReH_1^0 - v)\cos(\beta-\alpha) - \sqrt{2}ReH_2^0\sin(\beta-\alpha), \tag{3.23}$$

$$h = (\sqrt{2}ReH_1^0 - v)\sin(\beta-\alpha) - \sqrt{2}ReH_2^0\cos(\beta-\alpha), \tag{3.24}$$

The angle $(\alpha-\beta)$ represents the mixing of the neutral scalars. A SM-like Higgs boson exists if $(\sqrt{2}ReH_1^0 - v)$ 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. [50], 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 $|cos(\beta - \alpha)| << 1$. 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. In the alignment limit, on the other hand, even if the tree-level couplings of h to the vector bosons are SM-like, deviations can appear when higher order corrections are included[50].

Chapter 4

The LHC and the ATLAS experiment

4.1 The Large Hadron Collider (LHC)

New accelerator development has historically been driven by particle physics research. In 1932 the first man-induced splitting of the atom was obtained by Cockroft and Walton by accelerating protons to 400 kV using static voltage acceleration. A few weeks later Lawrence's cyclotron, based on the principle of resonance acceleration, produced 1.25 MeV protons also to disintegrate the atom and the era of collider physics began [51]. Since then, the need to look at increasingly smaller distances has meant continuously finding new ways to produce higher energy beams. By the 1970s, it was shared expectation that new physics should appear at substantially higher energies than ever tested before, in the range up to 1 TeV, and several options were considered. 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 energy of 13-15 TeV, equipped with 8 T magnets and with a luminosity of $10^{33}\text{cm}^{-2}s^{-1}$ should be the next choice for the next major facility of CERN [52]. Thus, the Large Electron Positron collider (LEP) [53] tunnel was designed from the start to provide enough space to later install superconducting magnets for a large proton collider [54]. The tunnel was constructed

between 1984 and 1989. In 1989, CERN started LEP, the world's highest energy electron-positron collider and, in 2000, LEP was terminated to liberate the tunnel for the LHC [55]. 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 accelarate protons to 7.5 TeV.

4.1.1 Overview

The LHC 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 [56] and CMS [57], both multi-purpose experiments designed to be sensitive to a wide range of SM processes and new physics searches; LHCb [58], a B-physics experiment; and ALICE [59], 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 located. [1] 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. This differs from a particle-antiparticle collider configuration, where both beams can share the same pipe with a common vacuum and magnet system. Because of the limited space in the tunnel, not

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

enough room was available for two separate rings of magnets. For this reason, the twin-bore magnet design proposed by Blewett in 1971 [60] 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

The LHC accelerator chain is shown in Fig. 4.1. 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 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 4 minutes and 20 seconds to fill the entire LHC ring. A total of 16 Radio Frequency (RF) cavities, delivering 2 MV longitudinal voltage at 400 MHz per cavity, are used to accelerate the protons [62]. Twenty minutes later, after passing through the RF cavities 10 million times and traveling a distance greater than twice the distance between the Earth and the Sun, the protons reach their maximum energy of 6.5 TeV, resulting in a center-of-mass energy at collision point of $\sqrt{s} = 13$ TeV. After collision, the beams are reconstituted, and the process continues for

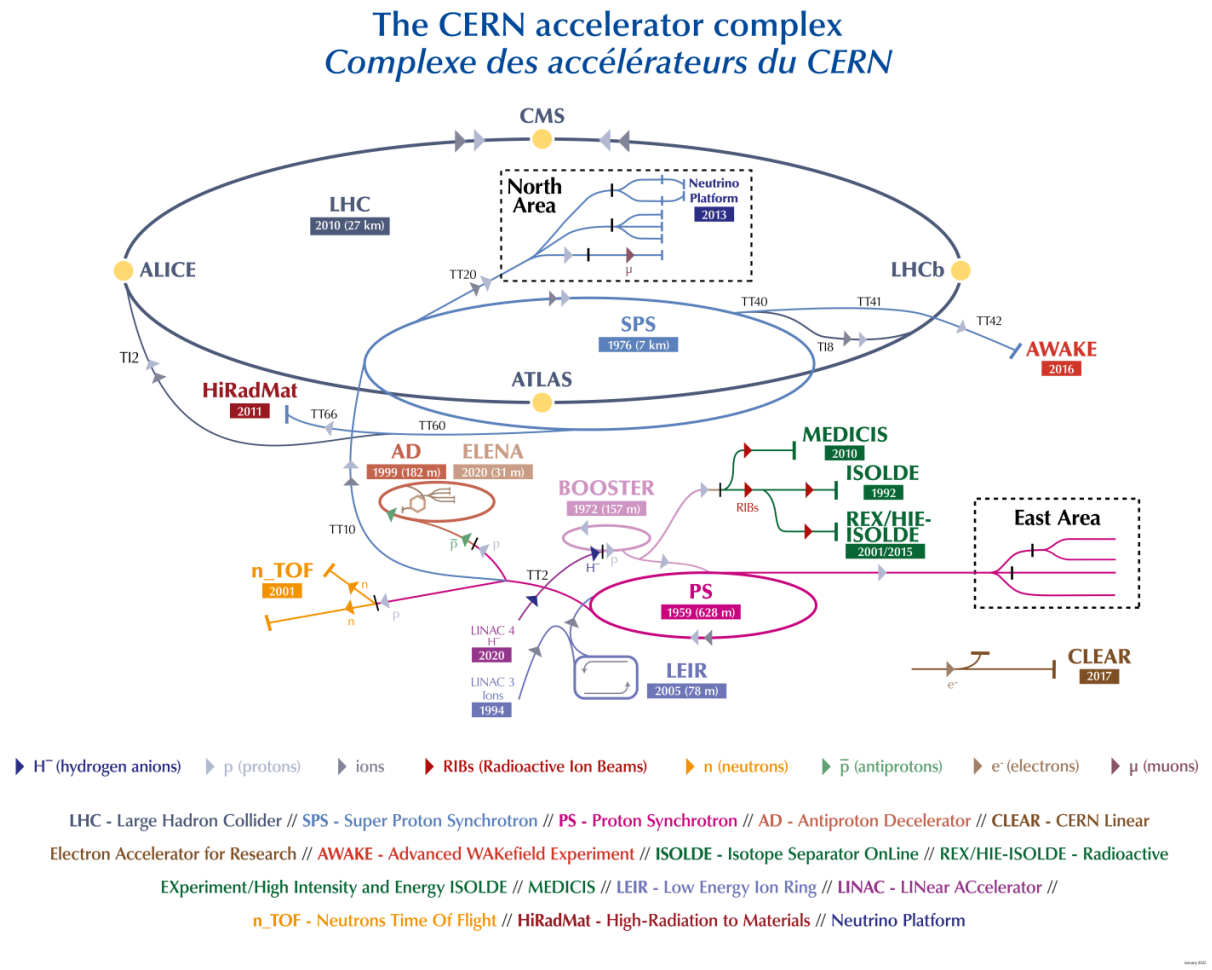


Figure 4.1: The LHC accelerator complex layout as of 2022 [61].

approximately 10 hours until the beam is depleted of protons and is ready to be dumped. At this point, protons exit the LHC rings and travel along a straight line until collision with a block of concrete and graphite.

The RF system and the beam bunch structure

The Radio Frequency (RF) cavities are straight metallic chambers containing a longitudinal electromagnetic field and housed in cylindrical refrigerators, called cryomodules, to maintain the cavities in a superconducting state. 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*: a proton with slightly higher energy will travel a longer path and hence will be delayed with respect to particle A, so that the next time it arrives at the gap it will see a decelerating voltage; for a proton with slightly lower energy the opposite will happen. [2] The result of this motion is that, instead of being evenly spread around the circular path of the accelerator, the particles get clumped 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 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 magnets system

Along the LHC rings, 10,000 superconducting magnets of about 50 different types are used to send the protons along the circular path without them losing their energy. A current of 11 kA passes through the superconducting cables made of niobium-titanium alloy of the

[2] In general, for a massive particle like the proton, an increase in momentum can cause an increase or decrease in the revolution frequency, according to the value of $\beta = \frac{v}{c}$. However, the transition at CERN occurs in the Proton Synchroton, while the LHC rings operate exclusively above transition, where the protons have $\beta \approx 1$ and the revolution frequency decreases as the momentum increases [63].

1232 main dipoles, each 15 metres long, to generate a nominal magnetic field of 8.33 Tesla, 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.

Particles in a bunch will occupy different positions on the transverse plane, defined as the plane perpendicular to the trajectory of the particles. 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 ends 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. When particles diverge too far from the central trajectory, the quadrupoles focus them back towards the center. These transverse oscillations on the horizontal and vertical planes are called betatron oscillations and the β-function is the envelope function within which the particles oscillate [63]. The quadrupoles are also used to squeeze the beam and increase the beam luminosity. In particular, eight sets of special low-β quadrupoles, called the *inner triplets*, are used at the IRs of the four experiments to make the beam even narrower, going from 0.2 millimetres across down to 16 micrometres.

4.1.3 LHC performance and operation

The bunch structure

The proton bunches have an elongated shape of 7.48 cm along the longitudinal direction due to the synchrotron motion, and of 16×16 μ m in the transverse plane due to the betatron oscillations. The bunches are spaced by 25 ns, or 7.5 m, giving a collision rate of 40 MHz (or 40 million collisions per second) and a maximum number of bunches in the ring of 3564.

However, not all bunches are filled with protons, as empty bunches are necessary to allow enough space for, e.g., new bunch insertion when non-useful 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 bunch crossing 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 bunch crossing can have either two bunches colliding, one bunch, or be empty of protons.

The cross section

The probability of a given collision event to occur is expressed by the cross section σ_{process}, which is fixed by nature at a given center-of-mass energy and can, in principle, be calculated at fixed order in QFT. The cross section is measured in units of squared-area. Because particle physicists are generally interested in rare events, σ is usually a very small number more conveniently expressed in barns, where $1b = 10^{-24}cm^2$. 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 given by

$$\frac{d\text{N}_{\text{process}}}{dt} = L_I \sigma_{\text{process}}, \tag{4.1}$$

where L_I is the instantaneous luminosity provided by the machine, which is approximately $L_I \approx 10^{34}\text{cm}^{-2}\text{s}^{-1}$ at the LHC. With about 30 million bunch crossings per second (set by

the number of filled bunches), the inelastic event rate is of 20 inelastic events per bunch crossing [3]. The cross section to get a Higgs boson via $qq \to Zh$ is approximately $50\,\mathrm{fb}^{-1}$. This means that there are $N = ((50 \times 10^{-15} \times 10^{-24}) \times 10^{34}) = 5 \times 10^{-4}$ events/second where a Higgs bosons is being produced or, in other words, a Higgs boson is produced every 33 minutes. This type of events is extremely rare when compared to the 11-14 order of magnitude larger number of events that will have occurred during the same amount of time. 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.

The luminosity

The luminosity is the number of collisions that can be produced in a detector per cm^2 and per second, and it is dependent only on the beam parameters. Assuming two identical Gaussian beams colliding, the instantaneous luminosity is:

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

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,

[3] This estimate was performed with the nominal parameters of the accelerator. As the accelerator complex continues to be upgraded, the number of simultaneous inelastic pp processes has consistently been increasing.

- σ_x and σ_y the RMS cross-sectional size of the bunch in the x and y directions. In terms of the β-function and emittance ϵ, $\sigma_i \approx \sqrt{\beta_i \epsilon}$,

- F is the geometric luminosity reduction factor due to the crossing angle at the IP caused by the beams traveling along different rings and therefore not being perfectly head-on [4].

The discovery reach of the LHC ultimately depends on the total integrated luminosity, $\mathrm{L} = \int \mathrm{L}_I dt$, related to the total number of events of a given process as $\mathrm{N}_{\mathrm{process}} = \mathrm{L}\sigma_{\mathrm{process}}$. The total integrated luminosity has units of [cm^{-2}] and, similarly to the cross section, is usually expressed in units of inverse barns. The knowledge of the luminosity is necessary to extract the visible cross section in any detector. The luminosity is measured by the experiment with specific detector sub-systems that are calibrated during special runs called van-der-Meer beam-separation scans [64].

Increasing the luminosity is the only way to increase the number of rare events produced and this requires high beam energies and intensities [55]. 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 probability of two protons colliding is related to the maximum number of particles per bunch and to 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

[4] The 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 full crossing angle at the IP, σ_z the RMS bunch length, and σ^* the transverse RMS beam size at the IP.

chamber. The nominal and ultimate values of the number of protons per bunch are 1.15×10^{11} and 1.70×10^{11}, respectively [65]. 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 more the particles are close together in distance and momentum. The emittance is determined solely by the initial conditions set by the injection chain. The β-function determines the beam envelope and is determined by the focusing forces of the quadrupoles and can be adjusted during a run. The β^* parameter is defined as the longitudinal distance from the interaction point where the transverse beam size grows twice as wide. The smaller β^*, the larger the luminosity.

Several LHC upgrades have already been performed and others are planned in the coming decade to bring the accelerator parameters to the ultimate design goals and beyond. However, with a higher collision rate comes 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 *pp* 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 [66], biasing and smearing the quantities reconstructed from the detector. An increase in pileup determines also an increase in the radiation levels that detectors and front-end electronics have to withstand and in the size of a collision event, stressing the trigger and data acquisition systems. For in-time pileup, one usually reports the average *pileup multiplicity* μ, or the average number of collisions per bunch crossing, which follows a Poisson distribution. The average pileup multiplicity was $< \mu >= 20$ in Run 1 and will reach $< \mu >= 200$ 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 coping with the increase 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 19^{th}, 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. This event determined the timeline of the following years. In particular, it was decided to run at $\sqrt{s} = 8$ TeV during the first run until the first Long Shutdown (LS) planned for 2013. The LHC plan is shown in Fig. 4.2. Years-long periods of consecutive data-taking, called *Runs*, are separated by major upgrades necessary to bring the collider to its ultimate performance:

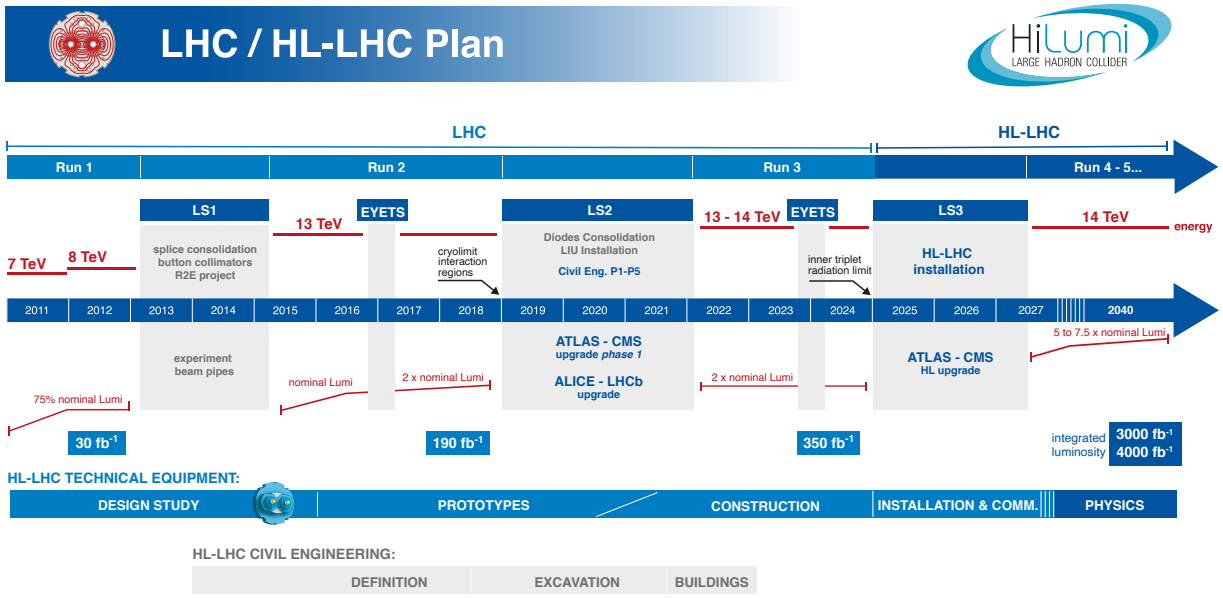


Figure 4.2: Snapshot of LHC schedule showing collision energy (upper line) and luminosity (bottom line) as of 2022 [67].

- 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 center-of-mass 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 saw an increase in energy from 13 TeV 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 50% increase in the collision rate, combined with luminosity leveling, will allow to run near the peak luminosity for a much longer fraction of the running time resulting in a year integrated luminosity of 250 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 [68] at the IRs. 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 $\sqrt{s} = 14$ TeV and deliver a levelled instantaneous luminosity of $5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ and an annual integrated luminosity of 250fb^{-1}.

The current goal is to reach a total integrated luminosity of 350fb^{-1} at the end of Run 3, largely beyond the original goal of 300fb^{-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 3000fb^{-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. The total integrated luminosity as a function of the runs and shutdowns is shown in Fig. 4.3.

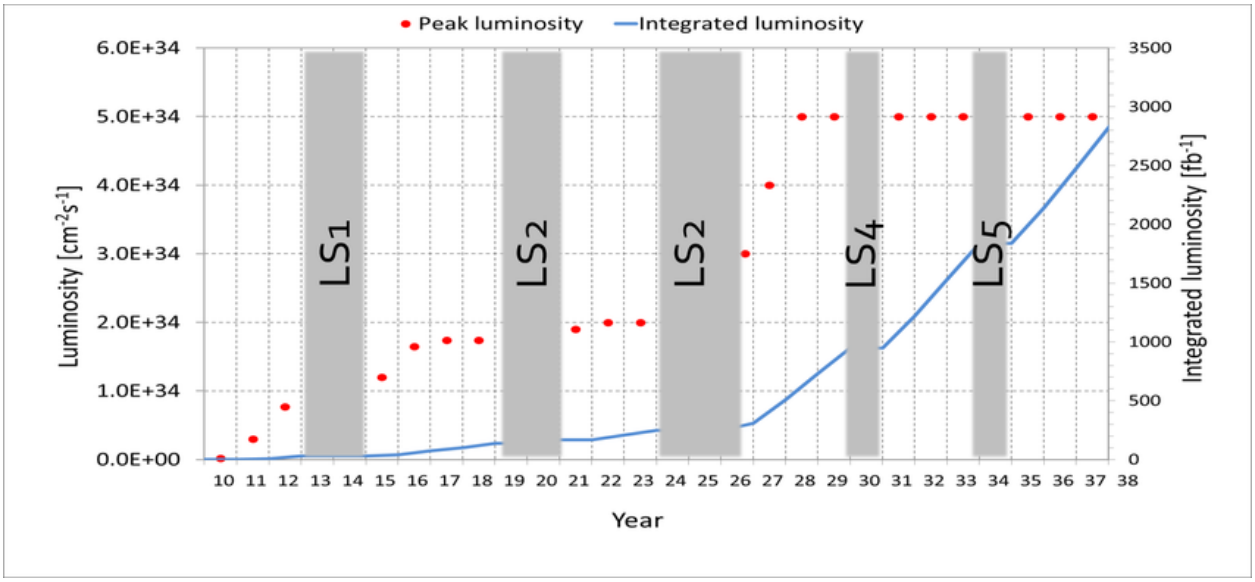


Figure 4.3: The LHC past and projected performance in terms of expected peak and integrated luminosity from 2010 through 2038, displaying Runs and Shutdowns periods. [69].

4.2 The ATLAS detector

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

4.2.1 Design goals

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. The following requirements were needed to maintain the physics reach of the experiment [56]:

- Fast, 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 signal 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 different sub-detectors, each sensitive to specific types of particles and phenomena, surround the interaction point. 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 detailed description of the detector as it was during Run 2, the period in which the data set used in this thesis was collected.

4.2.2 The ATLAS coordinate system

ATLAS uses a right-handed coordinate system, as sketched in 4.4. The interaction point (IP)

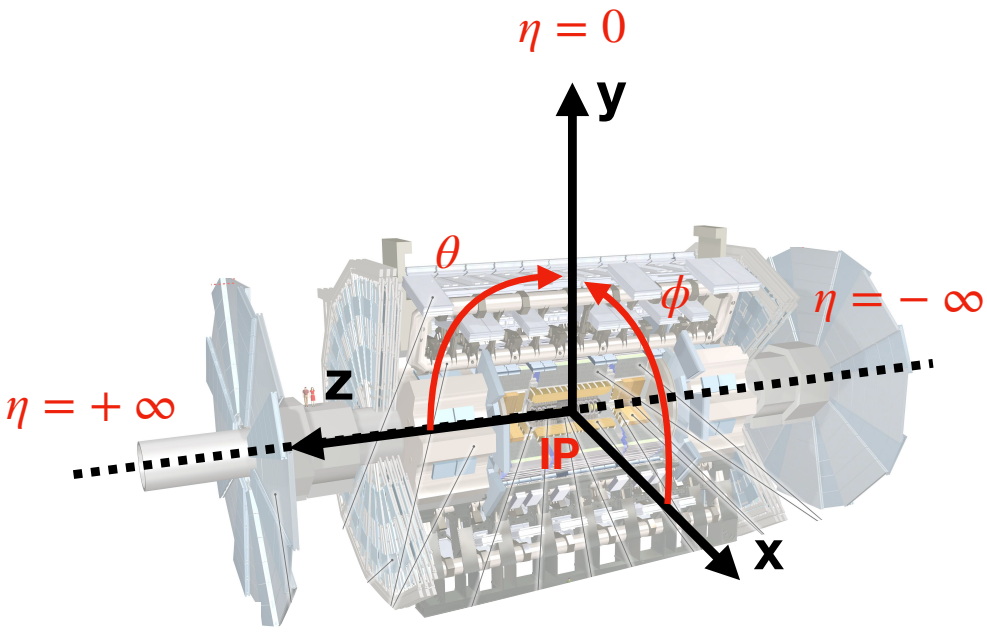


Figure 4.4: ATLAS coordinate system. Background taken from Ref. [56].

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 circumference traced by the collider,

while the positive y-axis points upward, towards the sky. The azimuthal angle ϕ is measured on the transverse plane around the beam axis, with $\phi = 0$ on the positive x-axis. The polar angle θ is measured on the z-y plane from the beam axis, starting from the positive z-axis.

The rapidity of a particle is defined as $y = 1/2\ln[(E + p_z)/(E - p_z)]$, where E and p_z are the energy and momentum of the particle along the z direction. Differences in rapidity are Lorentz invariant under boosts along the z (beam) direction. 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.

The ΔR distance between two objects i and j in the rapidity-azimuthal angle space is defined as $\Delta R_{i,j} = \sqrt{\Delta(y_i - y_j)^2 + \Delta(\phi_i - \phi_j)^2}$, which is also Lorentz invariant. Lorentz invariance under boosts in the z-direction is important in a pp collider where, due to the complex QCD structure of protons, the four-momentum is not conserved along the z component. On the other hand, the total four-momentum before and after the collision is, theoretically, zero, making the event reconstruction in the transverse plane essential to physics analyses.

4.2.3 The magnet system

The detector is fully immersed in a magnetic field provided by a system of four large superconducting magnets shown in Fig. 4.5. The field covers approximately 12,000 m^2 (22 m in diameter and 26 m in length) and has a total stored energy of 1.6 GJ. A thin superconducting solenoid [70] aligned with the beam axis surrounds the ID cavity and provides an axial magnetic field of 2 T at the center. The layout 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 field of 0.5 T and 1 T, respectively. The entire

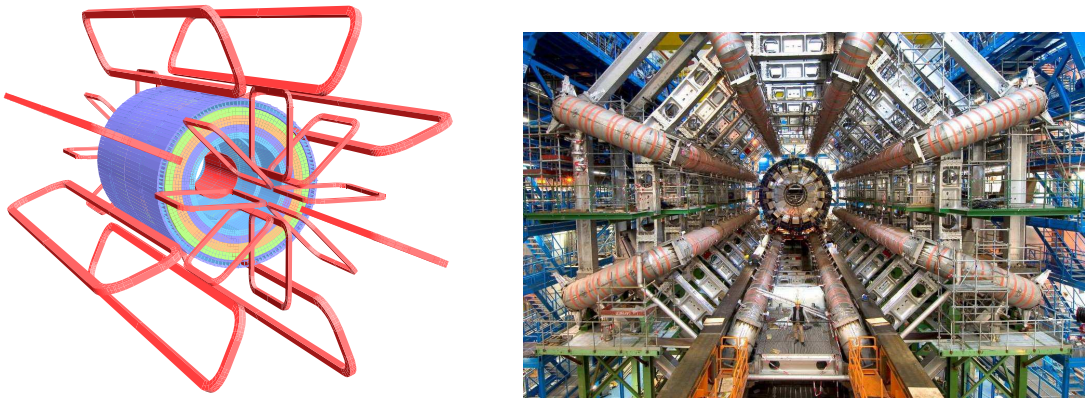


Figure 4.5: a) View of the ATLAS magnet system. The eight barrel toroids and each of the end-cap magnets are shown. The solenoid is also visible inside the tile calorimeter volume. b) Picture of the barrel toroid as installed in the underground cavern with its supporting structure. The barrel calorimeter and embedded solenoid are also visible on the other side awaiting to be put in position. The person standing in front of the structure gives the scale [56].

system is under vacuum and cooled down by the ATLAS cryogenics system. Each of the coils in the barrel toroid is housed in an individual cryostat, while only one cryostat houses an entire end-cap toroid. Detailed magnetic modelling provided highly precise description of the magnetic field. This resulted from the computation of the Biot-Savart contributions of all magnet windings and the perturbations induced by the magnetised calorimeter and other ferromagnetic structures in the detector volume. In particular, the inner detector 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. The rest of the solenoid and the muon chambers are equipped with 3D Hall cards to measure each field component.

4.2.4 The inner detector

The Inner Detector (ID) 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 com-

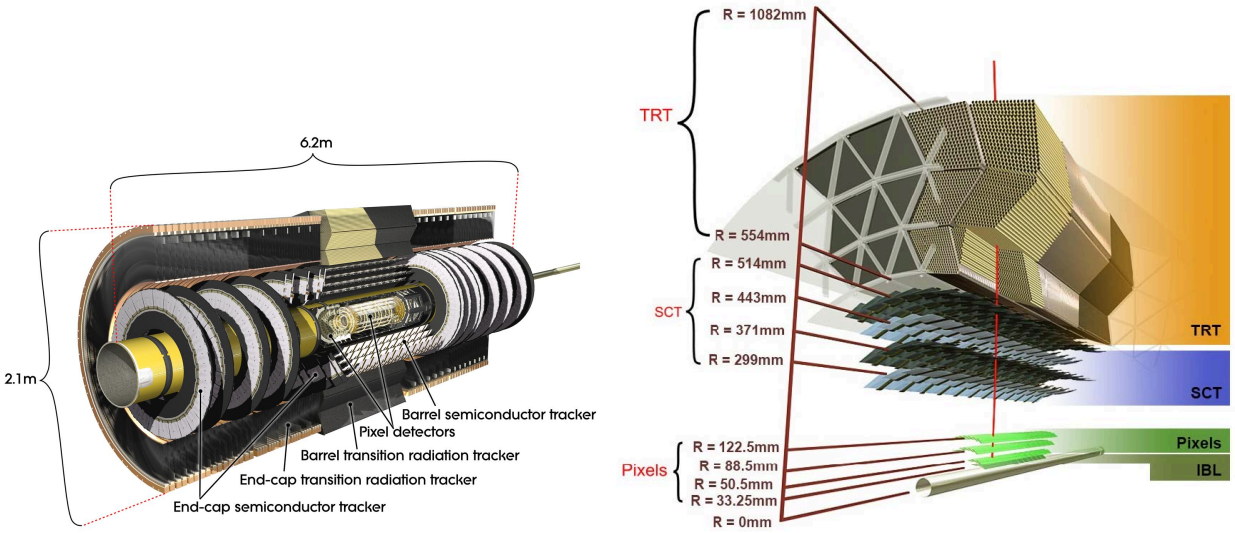


Figure 4.6: a) Schematic view of the ATLAS inner detector [71] b) Structural elements of the ID barrel region traversed by a charged track (in red) [72]. From the collision point, the track traverses the beryllium pipe, the three silicon pixel layers, four SCT barrel layers of silicon micro-strip sensors, and ~ 36 TRT straws.

plementary sub-components, 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 for Run 2 operations 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.

Pixel detector

The pixel system [73] is composed of three layers in the barrel region and three disks at each end-cap. 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 μm in the ϕ

direction and 400 μm in the z or r direction. The total number of pixels is 67 million in the barrel and 13 million in the end-caps, for a total of approximately 80 million readout channels. Intrinsic accuracies are $10\mu m$ on the $(R-\phi)$ plane and $115\mu m$ along the (z) direction. The pixel detector was designed to provide at least three points per charged track and covers the region $|\eta| < 2.5$.

The SemiConductor Tracker

The SCT [74] 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 ~ 16 μm in the $R-\phi$ plane and ~ 580 μm along the z direction.

Transition Radiation Tracker

The TRT [75] is the outer-most layer of the ID and provides continuous tracking coverage, with an average of 36 hits per track, within the $|\eta| < 2$ range. While the TRT provides only $R-\phi$ measurements and has a lower accuracy of 130 μm, this is compensated by the larger number of hits and longer track lengths, complementing the precision trackers. The TRT consists of about 300,000 straw tubes, 4 mm in diameter and made of Kapton, with a 31 μm diameter gold-plated tungsten wire. 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 creating an electron avalanche that produces a detectable signal. The straws are interleaved with transition radiation (TR) material (polypropylene) to enhance electron identification. TR consists of X-ray photons emitted by a particle passing through an inhomogeneous medium. The energy of the TR is

proportional to the relativistic factor $\gamma = E/m$ of the incident particle. A particle as light as the electron, will emit a TR significantly higher in energy than a pion or muon. An essential function of the TRT is indeed the discrimination of electrons from charged pions.

Insertable B-Layer

The first layer of the Pixel detector just around the beam pipe, called B-layer, was originally designed to be regularly replaced according to the radiation damage and increased peak luminosity. However, engineering changes that became necessary during ATLAS operation did not permit extraction of the layer from the ID anymore. 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 IBL [76, 77]. 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 $50 \times 250 \mu m^2$. The decrease in pixel size and the smaller radius from the collision, are crucial parameters for the improved performance of the ID. The new layer was installed in May 2014 during the LS1 before Run 2 and has been successfully operating since 2015.

The current ID was designed for 10 years of operation at nominal LHC parameters. In 2016, the LHC already exceeded the designed values, reaching unprecedented levels of pileup and radiation. Despite this, albeit minor upgrades, the ID performance remained adequate to sustain the harsher conditions. However, in order to maintain the physics performance at HL-LHC conditions, the ID will be fully replaced by a new tracking system, the Inner Tracker (ITk) [78, 79], which will be installed during the Phase II upgrades.

4.2.5 The calorimeters

Calorimeters [80, 81] are detectors used to measure the energy of incident particles via their total absorption. They are particularly well-suited for a high-energy multi-purpose experiment, as their energy resolution improves with increasing energy like $1/\sqrt{E}$, in contrast to, for example, magnetic spectrometers, where the momentum resolution deteriorates linearly with increasing momentum. They are also sensitive to any type of particle, including neutral particles, and can provide indirect neutrino detection.

When particles are stopped in the detector volume, showers are initiated. The shower evolution is different according to whether the incident particle interacts electromagnetically or hadronically, requiring different types of detectors. Calorimeters are typically segmented transversely and longitudinally. This allows to also measure the position of particles and discriminate between different particle types according to the shape of the shower. Modern calorimeters are also used to measure the arrival time of particles and provide fast signals for triggering. Triggering and particle identification are particularly relevant for this work.

Electromagnetic showers

An EM calorimeter measures EM showers induced by electrons and photons. At energies above $\sim$ 100 MeV electrons lose their energy almost exclusively via bremsstrahlung, while photons lose it via electron-positron pair production. These showers tend to be more contained in width and length, requiring a smaller detector volume, and their depth and width can be expressed in terms of the radiation length X_0 of the detector material [5]. For an elec-

[5] The radiation length for a material with atomic number Z and atomic weight A is given by [81]

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

tron, the average energy loss due to bremsstrahlung can be expressed as $-(dE/dx) = E/X_0$, while for a photon the probability of electron-positron pair creation is give by $7/9X_0$. The EM showers are quite narrow, of the order of centimeters, and the cell size has to be of the same order to allow 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. The longitudinal development of hadronic showers can be parametrized by the interaction length λ. 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 much greater 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 broadens the shower shape. These inelastic processes produce particles in the MeV scale, with the soft spectrum dominated by photons and neutrons, while the energetic component by charged and neutral pions, kaons, and nucleons.

One third of the pions produced are neutral pions, which quickly decay into two high energy photons before they have a chance to interact hadronically, initiating electromagnetic sub-cascades within the hadronic shower. The total fraction of energy in a hadronic shower coming 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 increases 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, as well as muons produced by pions and kaons decays, can escape from the calorimeter. The lower the electromagnetic fraction of the shower, the larger the fraction of "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 [81].

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* in which 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, which can result is a lower energy resolution. This type of calorimeter is usually built in alternating layers of heavy absorbing material, such as lead, and layers of active material. This makes sampling calorimeters more easily segmented longitudinally and radially, resulting in better space resolution and particle identification.

Sampling calorimeters are classified according to the type of active material. The shower generation starts in the absorber, where 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 only when the final particles have an energy smaller than what would

be needed to produce further particles. The deposited energy can be measured by collecting either the light produced in a scintillating material or the charge produced by ionization.

A scintillation sampling calorimeter uses organic (usually plastic) scintillators arranged in fibers or plates. Their main advantages are their low cost, the ductility in terms of what geometries and segmentations can be achieved, and their fast response. Their drawbacks are the aging and radiation damage of the optical readout, and non uniformities in the light collection chain.

In a noble-liquid calorimeter, where the active medium is a noble element such as Argon (Ar) or Xenon (Xe), half of the energy of the incident particle is converted into ionization energy and half into scintillation light, the latter produced by the recombination of electron-ion pairs. Usually, only the ionization energy is collected, but it is sufficient to retain a good resolution. The advantages of these calorimeters are their uniformity, thanks to the liquid being homogeneously distributed, and their radiation hardness. The drawbacks are the requirement of cryogenic equipment and a purification system. Liquid Argon (LAr) is the most commonly used noble liquid in this type of detectors because of its high purity and low cost.

The ATLAS Calorimeter

The ATLAS calorimeter [82] surrounds the ID and is a hybrid system consisting of a an EM and a hadronic calorimeter. The EM calorimeter is closest to the beamline and uses lead-LAr technology, representing approximately one interaction length λ. The surrounding hadronic calorimeter is a scintillator tile-Fe hadronic calorimeter (HCal) and covers approximately eight λ. LAr technology is used also for hadronic calorimetry in the end-caps and in the forward calorimeters. A schematic view of the calorimeter is shown in Fig. 4.7. Three cryostat

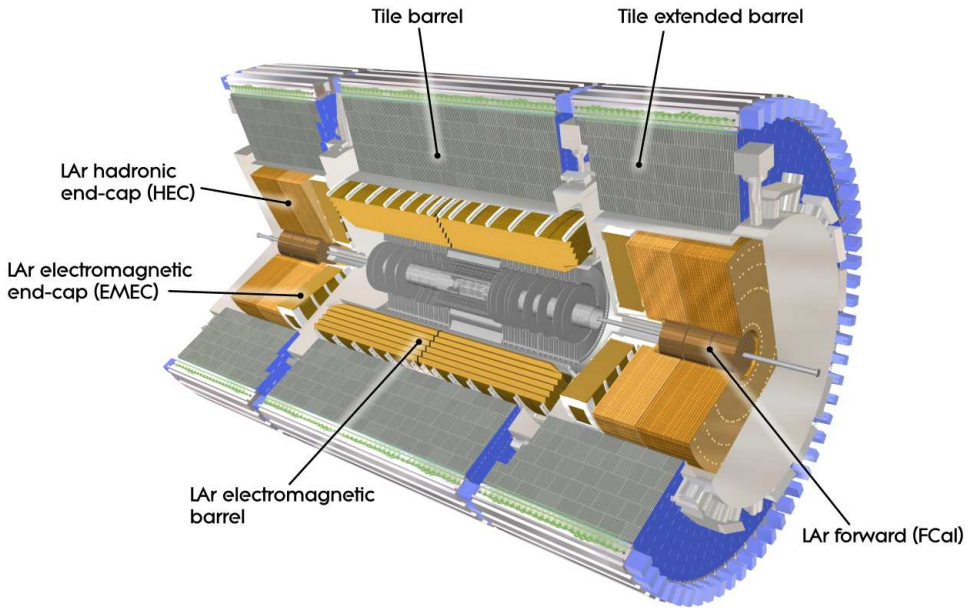


Figure 4.7: Schematic view of the ATLAS calorimeter [56].

systems surround the calorimeter components that use LAr technology: one cryostat houses the barrel EM calorimeter, while the two end-cap cryostats contain the EM end-caps, the HCal end-caps, and the forward calorimeters. All cryostats are built in aluminium alloy, 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 $1.15m < r < 2.25m$. The superconducting solenoid shares the same insulation vacuum as the liquid argon vessel. The end-cap cryostat has length 3.17 m and a radius of 2.25 m, the same as the barrel. The building block of a calorimeter is a *cell*, which measures the total integrated energy deposited in its volume.

The electromagnetic calorimeter

The electromagnetic (EM) calorimeter is a high-granularity LAr sampling calorimeter, covering the pseudorapidity range $|\eta| < 3.2$ and providing excellent energy and position resolution. Inside the LAr vessel, the calorimeter consists of two half cylinders joined at $z = 0$ and covering together the region $|\eta| < 1.475$. Each of the two end-cap EM calorimeters

(EMEC), one on each side of the barrel, consists of a 63 cm thick wheel covering the region $1.375 < |\eta| < 3.2$. Each EMEC wheel is partitioned into two co-axial wheels that are joined at $|\eta| = 2.5$, matching the acceptance of the inner detector. The absorber plates are made of lead with a thickness of 1.1-2.2 mm, according to the pseudorapidity region. These are interleaved with thin layers of liquid argon 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 fine segmentation in η.

The $\eta \times \phi$ granularity of the different layers is given in Tab. 4.1 and a sketch of a module in the barrel EM calorimeter is shown in Fig. 4.8. Precision physics studies were considered possible only within the region $|\eta| < 2.5$. This, together with the effort to keep the number of channels as low as possible to limit the costs, motivated the use of coarser granularity for higher η ranges. In particular, the inner end-cap wheel was designed with only two sampling layers. On the other hand, in the precision region of $|\eta| < 2.5$, both the barrel and the end-caps are composed of three layers of different resolutions and radiation lengths according to their physics requirements. The first sampling layer was optimized for π^0 rejection to have finer dimension in the pseudorapidity direction. To limit the number of channels, the granularity in the ϕ direction was reduced, resulting in thin strip-shaped cells. The second layer, which is expected to collect the largest fraction of the EM shower, is segmented in cells of size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. The last layer is expected to collect only the tail of the EM shower, so a coarser granularity of $\Delta\eta \times \Delta\phi = 0.05 \times 0.025$ is used. The depths of the first and second sampling layers were chosen to be $6X_0$ and $24X_0$, respectively. The third sampling varies in depth between 2 to 12 X_0.

η range	0-1.4	1.4-1.8	1.8-2.0	2.0-2.5	2.5-3.2
Pre-sampler	0.025 × 0.1	0.025 x 0.1			
Sampling 1	0.003 x 0.1	0.003 x 0.1	0.004 x 0.1	0.006 x 0.1	0.1 x 0.1
Sampling 2	0.025 x 0.025	0.025 x 0.025	0.025 x 0.025	0.025 x 0.025	0.1 x 0.1
Sampling 3	0.050 x 0.025	0.050 x 0.025	0.050 x 0.025	0.050 x 0.025	
Trigger	0.1 x 0.1	0.1 x 0.1	0.1 x 0.1	0.1 x 0.1	0.2 x 0.2

Table 4.1: Granularity of the EM calorimeter in $\eta \times \phi$ for different η ranges [82].

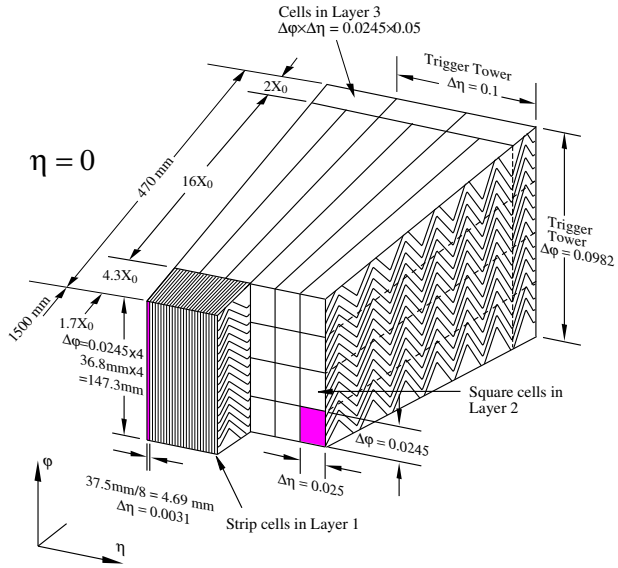


Figure 4.8: Sketch of a module in the barrel EM calorimeter, showing the $\eta \times \phi$ granularity for the cells and trigger towers in each layer [56].

The pre-samplers

Pre-samplers are used to provide shower sampling and compensate for the energy loss caused by the amount of material in front of the EM calorimeter. A thin 11 mm liquid-argon layer is positioned in front of the first active layer of the barrel EM calorimeter. A second 5 mm liquid-argon pre-sampler is positioned in front of the EMEC in the transition region $1.5 < |\eta| < 1.8$, between the barrel and the end-cap, to improve the energy measurement. Around $|\eta| = 1.4$, where the energy loss is significant, a scintillator layer, positioned between the two cryostats, is used to further recover the jet energy measurement.

The hadronic calorimeters

The HCal surrounds the EM calorimeter. It is composed of a scintillator-tile calorimeter, covering the range $|\eta| < 1.7$, while LAr calorimetry is used in the end-caps for the range $1.5 < |\eta| < 3.2$.

The tile calorimeter (TileCal) consists of a central barrel of 5.8 m in length covering up to $|\eta| < 1$, and two extended barrels of 2.6 m in length and inner and outer radii of 2.28 m and 4.25 m. It has a radial depth of approximately 7.4 λ. 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 ultra-violet scintillation light. Wavelength-shifting optical fibres, in contact with the tile edges, collect the scintillation light, convert it to a longer wavelength, and transmit it to the photomultiplier tube's located at the outer edge of each module. The grouping of the 540,000 readout fibres into bundles provides a three-dimensional cell structure with three layers in depth, of 1.5, 4.1 and 1.8 λ 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.

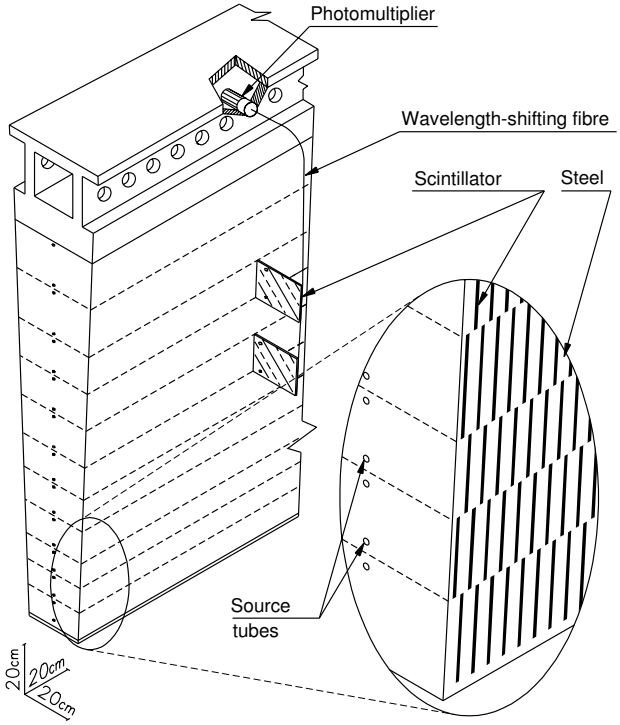


Figure 4.9: Schematic view of the tile calorimeter showing the various components of the optical readout: tiles, fibres, and photomultiplier tubes [56].

The hadronic end-cap calorimeter (HEC) is a LAr sampling calorimeter which uses copper as the absorber material. Each HEC is designed as two wheels, with the outer wheel having 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 forward calorimeter (FCal) provide both electromagnetic and hadronic energy measurements and extend the detector acceptance up to $|\eta| = 4.9$. The FCal is the innermost layer of the end-cap detectors, positioned along the beam axis at about 4.7 m from the interaction point. The very high levels of radiation make this a particularly challenging detector. In order to reduce the neutron albedo in the ID cavity, the front face of the detector is recessed

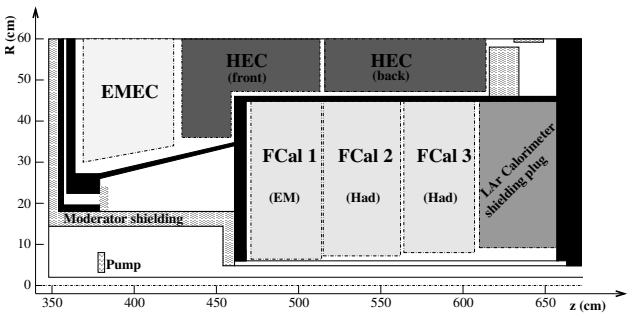


Figure 4.10: Schematic view of the FCal calorimeter inside the end-cap cryostat. The three FCal modules are visible, together with the shielding layers and the cryostat walls in black [56].

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 interaction point, uses copper as the absorber to optimize detection of EM radiation; the two outer layers use tungsten for hadronic calorimetry. Each section is composed of regularly spaced concentric tubes and rods positioned longitudinally along the beam axis, with the LAr active layer positioned between the tube and the rod.

Calorimeter read-out

A *cell* measures the total integrated energy deposited in its volume.

The dynamic range for the energy of the cells goes from $\sim$ 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:

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

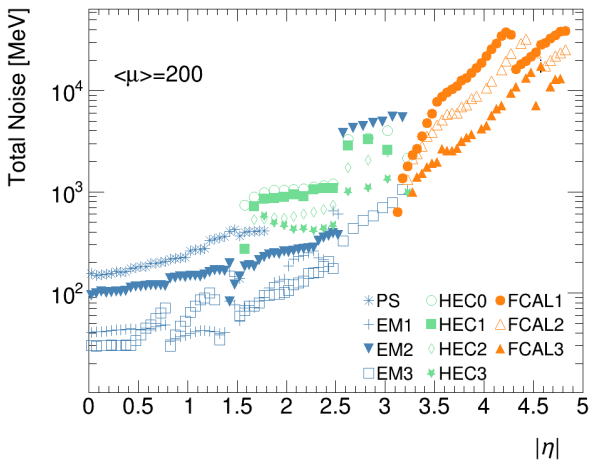


Figure 4.11: The total energy-equivalent cell noise at the EM scale as a function of η for the different detector sampling layers for a HL-LHC simulation with $\mu = 200$.

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. Its effect is not homogeneous in the detector volume, as shown in Fig. 4.11 for a $< \mu >= 200$ simulation. The majority of the soft particle energy flow is absorbed by the LAr and forward calorimeters, while the tile calorimeter has very 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. On the other hand, the back-end systems, located in the main services cavern (USA15) 70 m away from the detector, are made of commercial components.

4.2.6 The muon spectrometer

The muon spectrometer [83] detects the momentum and trajectory of charged particles escaping the hadronic calorimeter. Muons lose little energy via electromagnetic interactions and behave like minimum ionizing particles in a wide range of energies. This allows them to leave a characteristic signature in the detector that makes them easily distinguishable from other backgrounds and which made muons essential pieces of many analyses. However, it also makes their detection harder as their momentum has 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. As previously mentioned, 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 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.

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 of the spectrometer volume, covering the region $|\eta| < 2.7$ and with an average resolution of 80 μm per tube and 35 μm per chamber in the bending plane η.

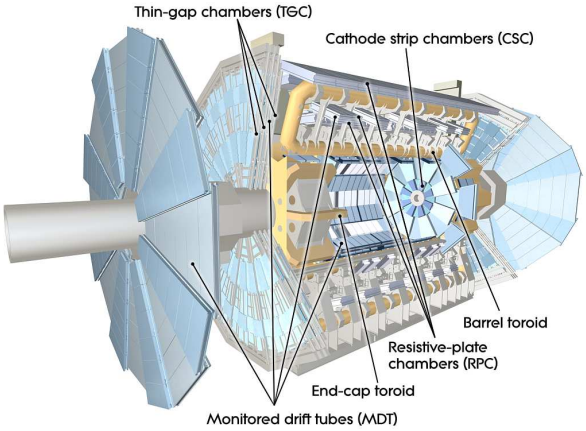


Figure 4.12: Schematic view of the muon spectrometer [56].

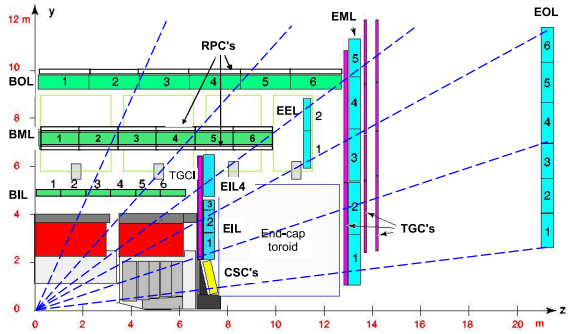


Figure 4.13: Cross-section of the muon spectrometer along a plane containing the beam axis (bending plane). The straight lines represent infinite momentum muons, which traverse three muon stations [56].

In the region $2 < |\eta| < 2.7$ the inner-most layer is made of Cathode Strip Chambers (CSCs), a finer granularity detector required by the higher background rates in the forward 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 chambers have a resolution of 40 μm in the bending plane and of 5 mm in the transverse plane.

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 is used to monitor the relative positions and possible deformations of the MDT chambers, while approximately 1800 Hall sensors are used to monitor the magnetic field throughout the spectrometer volume.

The muon triggers

The precision-tracking system is complemented by fast trigger chambers that deliver track information to the Level-1 muon trigger within a few tens of nanoseconds from the passage

of a particle. The trigger chambers look for high transverse momentum muon tracks. Reconstructed tracks are required to originate from approximately the interaction point and to pass certain p_T thresholds. Additionally, the trigger provides bunch-crossing identification and measurements of both coordinates, complementing the MDT measurement in the non-bending plane.

In the barrel ($|\eta| \leq 1.05$), Resistive Plate Chambers (RPC's) are operated in avalanche mode. The technology uses two parallel electrode-plates at a distance of 2 mm filled with gas in between. An electric field is applied between the plates. When a particle traverses the gas, it ionizes the molecules and causes an avalanche collected on the anode plate, with the gas gain proportional to the distance of the first electron cluster from the anode. The plates are disposed into three concentric cylindrical layers. Two layers are closer to the beam axis and are used to efficiently trigger on low p_T muons (6-9 GeV), while a third layer further out is used for high p_T muon measurements (9-35 GeV). Each plate is made of two detector layers, providing two measurements in η and ϕ.

The muon momenta increase significantly in the end-cap regions for a fixed p_T, posing increased resolution requirements. A higher resolution is called also by the fact that the trigger layers in the end-caps are outside the magnetic field and are closer together than in the barrel. For this reason, in the region ($1.05 \leq |\eta| \leq 2.4$) the trigger is composed of four layers of Thin-Gap Chambers (TGC), which are multi-wire chambers operated in saturated mode that provide good time resolution and high rate capability.

4.2.7 The forward detectors

In addition to the main detector systems, ATLAS is equipped with a set of smaller subdetectors located in the very forward region $|\eta| > 5$ on both sides of the IP.

Moving away from the interaction point, the first system is the LUCID detector. Located at ±17 m, LUCID is a Cerenkov integrating detector primarily dedicated to online luminosity monitoring. The relative luminosity measurement is based on the principle that the number of particles detected by LUCID is proportional to the number of simultaneous pp interactions in a bunch-crossing.

Next is the Zero-Degree Calorimeter (ZDC), 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, located at ±240 m, is used to measure the absolute luminosity, which can be extracted from the measurement of the elastic-scattering amplitude at small angles (3 μ rad) obtained during special runs with high-beta (β^*) optics and reduced beam emittance.

The AFP (ATLAS Forward Proton) detector was installed in 2017 at ±204m and ±217 m from the IP to measure diffractive protons scattered at small angles (100 μrad), where one or both protons remain intact.

4.3 The ATLAS Trigger

The ATLAS Trigger and Data Acquisition (TDAQ) 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 impossible, but it is also not desirable, as the events that are interesting for physics analyses are orders of magnitude

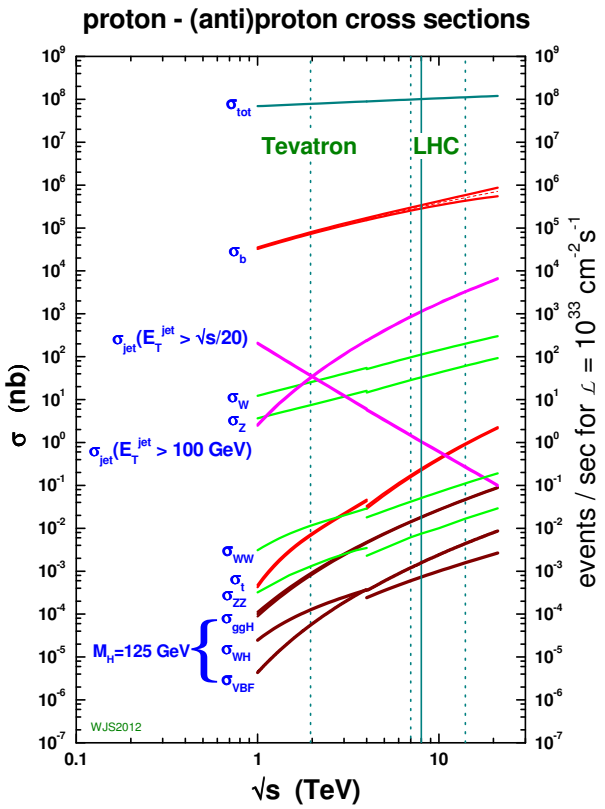


Figure 4.14: Standard Model cross sections as a function of collider energy [84].

rarer than the large background of QCD jet production and pileup, as shown in Fig. 4.14. 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 flows 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 using more complex algorithms run 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). [6]. 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 TDAQ system used during Run 2 is shown in Fig. 4.15. The level-1 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 Level-1

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 also identifies Regions of Interest (RoI) in η and ϕ that will 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 has to be 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

[6] During 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 EventFilter (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 overall simplification of the hardware and software [85].

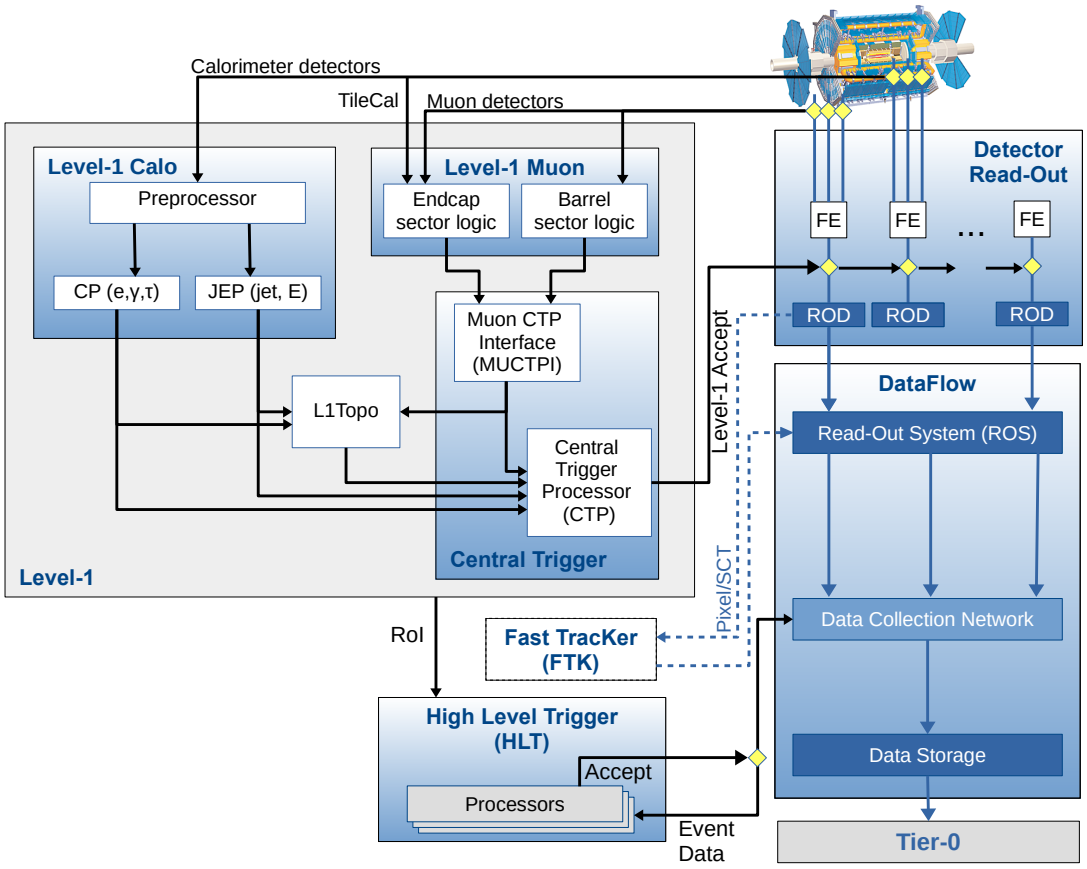


Figure 4.15: The ATLAS TDAQ system used during Run 2, showing the relevant L1 and HLT trigger components, as well as the detector read-out and data flow to permanent storage on L1 and HLT accepts, respectively [86]. Note that the Fast Tracker project was canceled and should be ignore in this figure.

made, is required to be less than 2.5μs. Custom-built electronics are needed to satisfy these extreme requirements.

The L1 trigger receives reduced-granularity data from the calorimeter and muon detectors, with the two detector systems handled by separate trigger components.

L1Muon

The L1 muon (L1Muon) [87] trigger uses the hits from the RPCs and the TGCs muon triggers to apply coincidences requirements and identify high p_T muon candidates. The results from L1Muon are sent to the Central Trigger Processor via the L1Muon Central Trigger Processor Interface (MUCTPI).

L1Calo

The L1 Calorimeter trigger (L1Calo) [88] 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 parts of the end-caps and in the FCal. 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 photomultiplier 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 digitizes the analogue calorimeter signals, identifies the bunch-crossing they originated from, and

performs a series of operations to clean and calibrate the signals before they are sent to the L1Calo algorithms. The data is then transmitted in parallel to the Cluster Processor and the Jet/Energy-sum Processor (JEP). Both Cluster Processor and JEP modules use sliding window algorithms to identify energy depositions of interest. See Ref. [88] for 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 Cluster Processor algorithms are run for every possible combination of overlapping 4×4 trigger-towers windows (0.4×0.4 in $\Delta\eta \times \Delta\phi$). 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. Therefore, 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. The different sizes are sensitive to different signatures: smaller window are better to discriminate

nearby small-radius jets, while larger sizes are more efficient for individual energetic large-radius jets. As the windows are overlapping, the algorithm allows for a jet to 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. Eight programmable jet thresholds can be implemented by the JEP, with each threshold specifying the minimum E_T requirement and the window size. The Common Merged Module (CMM) merges the results from the Cluster Processor and JEP modules and sends the information to the CTP. In Run 2, the data transmission rate allows the transmission of up to four jet or five EM/tau trigger objects, where one object is described by the E_T, η and ϕ coordinates, and the isolation threshold if used.

L1Topo

The L1 Topological (L1Topo) [89] processor was added to the L1 trigger chain during the LS1 in order to cope with the increased event rates by providing more sophisticated topological selections. It consists of two Field Programmable Gate Array (FPGA)-based modules that read trigger objects from the L1Calo and MUCTPI systems. The two modules are provided with the same data, and execute parallel and independent algorithms. They can receive a maximum of 120 electrons/photons, 120 taus, 64 jets, 32 muons, and 1 missing transverse momentum. To reduce the combinatorics, part of the computational time is dedicated to produce reduced lists of sorted trigger objects. 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}. It also provides the

flexibility of using central and/or forward objects, muon information, and requirements on triggers from adjacent bunch crossings to target exotic signatures such as long-lived particles. The L1Topo decisions are transmitted to the CTP after ∼200 ns.

Central Trigger Processor

The L1 trigger decision is formed by the Central Trigger Processor (CTP). The CTP receives inputs from the L1Calo, L1Muon through the MUCTPI, and L1Topo, as well as some of the detector subsystems. The trigger decision is implemented as a logical combination of the L1 outputs according to the *trigger menu*. The CTP is also responsible for applying prescales 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 regions-of-interest (RoIs) are sent to the HLT.

4.3.2 High-Level Trigger

The HLT has the goal of reducing the event rate from 100 kHz down to 1kHz. The HLT has access to the full granularity calorimeter information, data from the muon spectrometer precision chambers, and tracking information from the ID, which is not available at L1. The processing sequence is typically divided 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 [90] and are run on a farm of more than 40,000 Processing Units (PUs), which are continuously replaced with newer hardware throughout operations. Some of the algorithms use the L1 RoIs as seeds for feature extraction algorithms, requiring event data only around

the RoI, while other algorithms 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 [91] with a radius parameter R of 0.4 or 1.0, and the inputs are topological energy clusters produced by running the topo-clustering algorithm [92] on the full set of calorimeter cells calibrated at the EM scale. These offline algorithms will be discussed in Chap. 5. A detailed description of the HLT algorithms can be found in Ref. [85]. 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 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. Each chain targets a specific physics signature and will be used by a physics analysis to recover events with the desired topology. Trigger names are usually given by the name of the trigger level (`L1` or `HLT`, with the latter usually omitted), followed by the object multiplicity, the particle type (`g` for photon, `j` for jet, `xe` for E_T^{miss}), and the p_T threshold. Some triggers are pre-scaled in order to adjust the rate of accepted events: a pre-scale value of n means that an event that would pass the given trigger is retained with a probability of $1/n$. The list of trigger chains forms the *trigger menu*. Every chain has an assigned event rate and the threshold requirements of the L1 and HLT triggers are set to keep the expected rate within this budget.

The most significant constraints on the trigger menu design during Run 2 were the limits on the L1 rate of 100 kHz, set by the detector read-out, and on the HLT output rate of 1kHz.

A share of the rate budget is assigned to each physics signature according to the physics goals of the collaboration and determines how low the relevant thresholds can be. The design of the Run 2 trigger menu was shaped by the goal of maintaining the un-pre-scaled single-electron and single-muon trigger p_T thresholds around 25 GeV, in order to preserve the trigger efficiency for 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. For instance, the physics trigger menu and operations for 2015 data-taking can be found in Ref. [85].

Different types of triggers and trigger menus exist. The *primary* triggers are the ones used to select events of interest for physics analyses and are usually un-pre-scaled. 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 are used for detector calibration and store only partial event information, which allows them to operate at high rates, and *support* triggers, used for monitoring and usually operated with pre-scaled factors. For trigger algorithm development and rate predictions, a special menu is used called *minimum bias* [93]. A data set collected by a trigger that fires at random would provide the probability of the selection to accept events in a truly unbiased sample. 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. A minimum bias sample re-introduces some bias for these harder events with respect to a zero-bias sample, while still not favoring any particular signature. This is obtained by using a collection of several different L1 trigger items targeting various signature types. Because the correlation between the triggers is preserved, a re-weighting of

the events allows to recover a zero-bias sample. The resulting samples includes a mixture of soft and hard processes, with soft events dominating.

4.3.4 The Phase I trigger upgrade

The TDAQ Phase I upgrades [94] were installed before the start of Run 3 in order to cope with the concomitant LHC upgrade during the 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. In fact, the peak instantaneous luminosity during Run 3 will be limited at $\mathcal{L} = 2\times10^{34}\text{cm}^{-2}\text{s}^{-}1$, mainly because of heating limitations of the inner-triplet magnets. However, the improvements will still provide 60% more intense beams already in Run 3 and luminosity leveling will allow to remain at peak luminosity of $2.4 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ for up to 10 hours during an LHC fill, increasing the average pileup to $< \mu > \approx$ 60-70[]. 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. In particular, the DAQ system had to handle a 30% larger event size at L1, while the latency and output rate of the L1 trigger are fixed by the original specifications of the detector to 100 kHz. This means that the HLT system had to target an output event rate 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 [72]. The Run 3 TDAQ system is shown in Fig. 4.16.

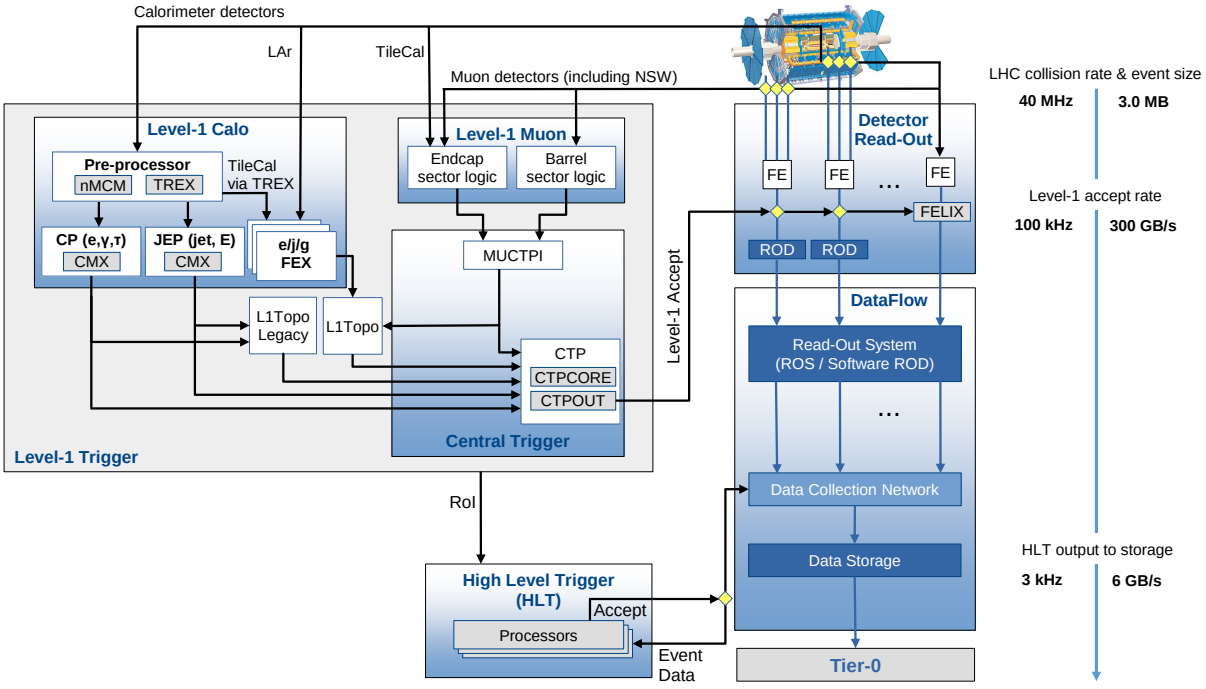


Figure 4.16: Schematic view of the Trigger and Data Acquisition system at the beginning of Run 3 [72].

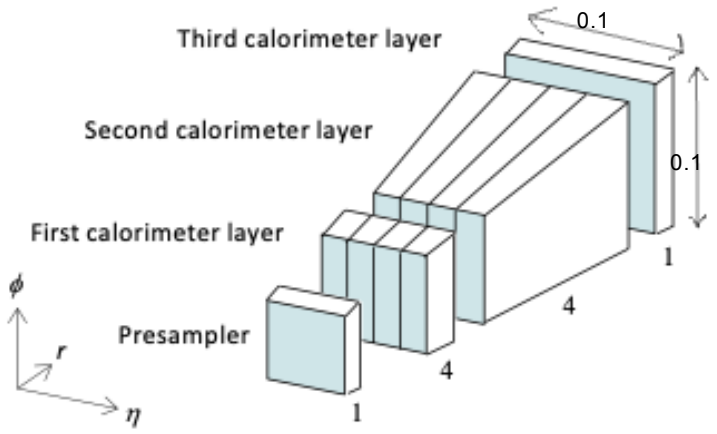


Figure 4.17: Example of an EM barrel 0.1×0.1 trigger tower divided into ten Super Cells for Run 3 [72].

L1Calo

The Run 3 L1Calo system has new electromagnetic and jet Feature EXtraction (FEXs) algorithms running on FPGA modules. It also has access to a finer granularity calorimeter information, as in place of the 0.1×0.1 trigger towers of Run 1 and 2, the LAr processing system now sends the information in the form of *Super Cells* containing sums of four or eight calorimeter cells. Fig. 4.17 shows an example of a trigger tower 0.1×0.1 in the EM Barrel detector that now contains ten Super Cells.

Different FEX algorithms are used to reconstruct different trigger objects (TOBs). The electron feature extractor (eFEX) module runs algorithms for 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 of 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 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 have access to 0.1×0.1 towers, still an improvement with respect to the 0.2×0.2 resolution from 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. [72].

The jFEX small-radius jet algorithm

The calorimeter inputs to the jFEX 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 FPGA 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. 4.18. The search window has size of 5×5 trigger towers, or 0.5×0.5 in $\eta \times \phi$ space. The seeds are constructed as the sum of 3×3 tower blocks around each tower in the search window. The energies of the seeds found in the search window are compared using comparative operators, which take care of the possibility of comparing equal digital values. Once the maximum energy seed is found, the tower at the center of the seed is chosen as the center of the jet. The already computed energy sum inside the seed (shown in red in Fig. 4.18), is added to the energy ring including all the towers within a radius of $0.2 \leq R < 0.4$. The final jet consists of 45 towers forming an approximately round shape of $R = 0.4$.

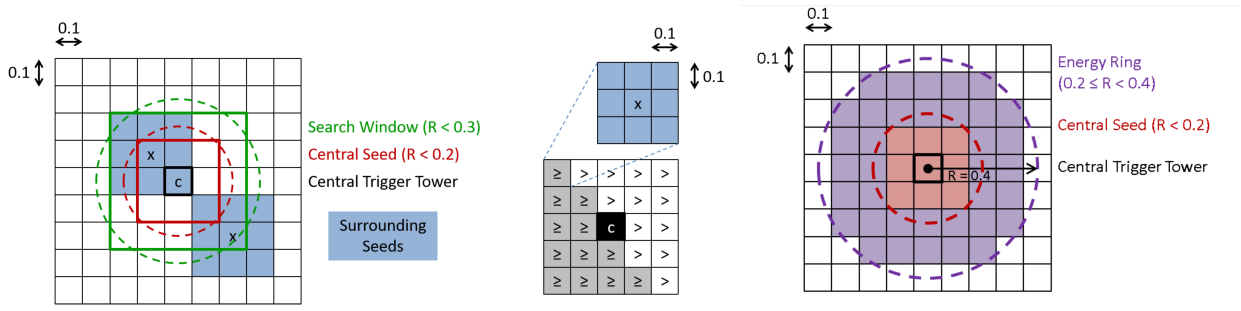


Figure 4.18: The jFEX small-radius jet algorithm. From left to right: the seed finding process with identification of local maxima; comparative operators used to identify a local maxima; the final small-R jet centered on the trigger tower at the center of the seed and built with all the towers within $R < 0.4$ [72].

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 4000 fb^{-1} of data, ten times the entire data set collected up until Run 3. This will allow the ATLAS and CMS experiments to substantially extend their 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. In the ID, the silicon strip and pixel sub-systems will be replaced by the ITk, which will provide an acceptance up to $|\eta| = 4.0$. The LAr and tile calorimeters readout electronics will be upgraded to improve the current limitations on the L1 trigger latency and output rate of 100 kHz. Similarly, the MDT front-end electronics will be replaced to handle the higher rates and also to provide

MDT hit information to the first step of the trigger chain. The rest of the MS upgrade will focus on improving the performance of the trigger chambers, by upgrading the electronics of the RPC and TGC chambers and adding new RPC detectors to increase the solid angle coverage.

The Phase II TDAQ upgrade 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. This 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. The diagram in Fig. 4.19 shows how ATLAS representative physics goals are connected to the required trigger objects and to the trigger systems necessary to achieve these goals.

The architecture of the Phase II TDAQ upgrade [10] is shown in Fig. 4.20. 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). A 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 in Run 3). With a predicted event size of 6 MB, the total output bandwidth will be 60 GB/s.

The Level-0 Trigger

The L0 trigger is composed of the L0Calo, L0Muon, MUCTPI, and the Central Trigger Processor (CTP), inherited from the current trigger system, and a new addition, the Global Trigger system.

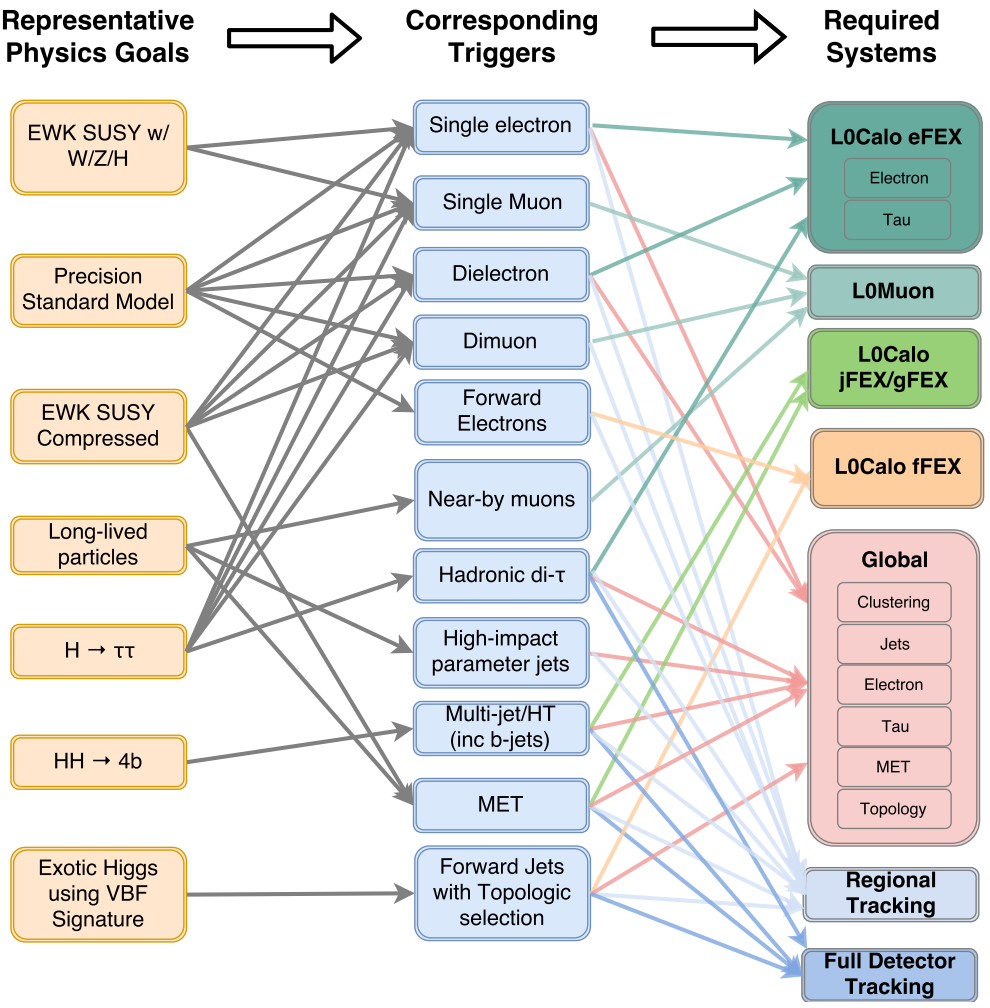


Figure 4.19: Diagram showing the relationship between ATLAS physics goals, required triggers, to the related trigger components of the Phase II trigger system [10].

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 Phase I L1Calo FEX algorithms, as described in the previous section, with the addition of a forward FEX (fFEX) for reconstruction of forward electromagnetic 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 of precision MDT momentum measurements and signals from the RPC inner stations in the trigger decision. The MUCTPI calculates

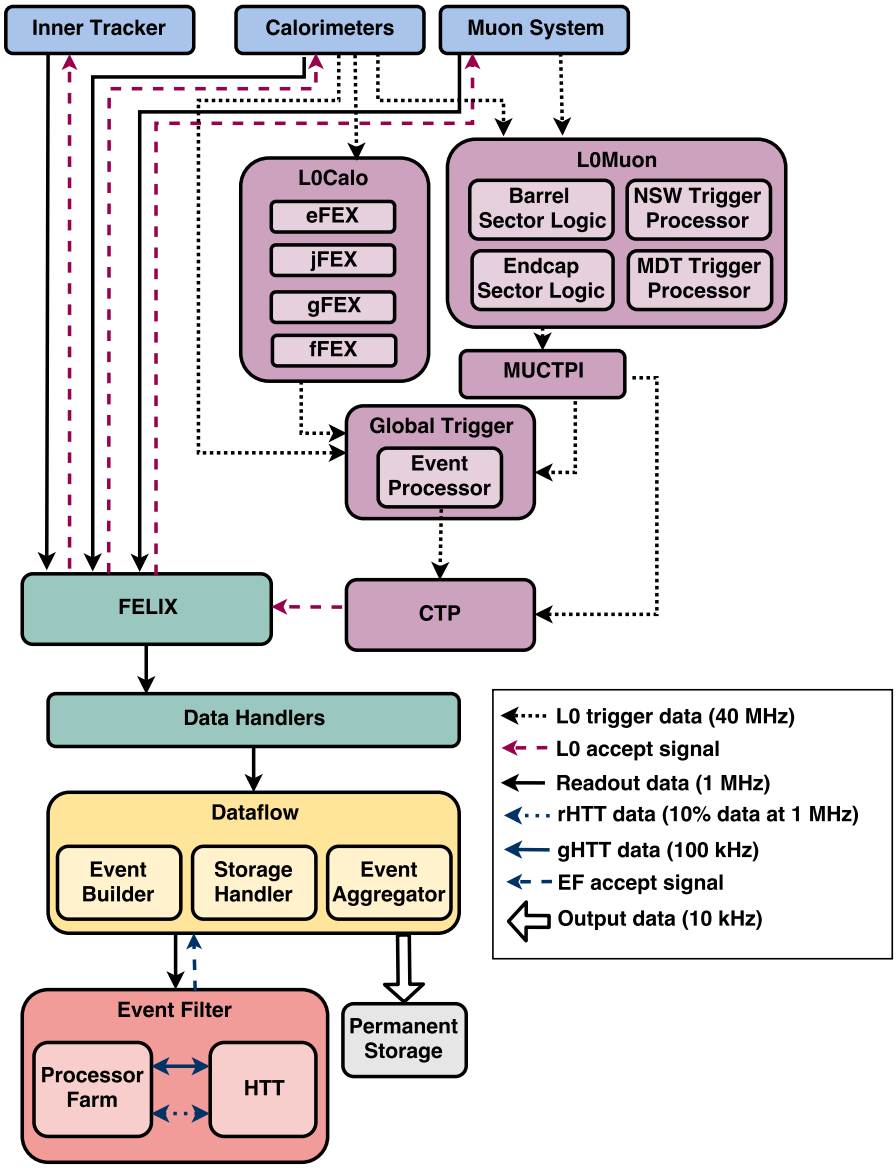


Figure 4.20: Design of the TDAQ Phase II upgrade with its three main systems: Level-0 Trigger, Data Acquisition system, and Event Filter. The black dotted lines indicate the data flow at 40 MHz from the detector systems to the Level-0 trigger system which must produce a trigger decision within 10 μs. 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 [10].

multiplicities of high energy muons, checks for double-counting of muon candidates, and interfaces the L0Muon with the Global Trigger (GT) and the CTP.

The GT will be an entirely new addition to the trigger system. It will have access to full granularity calorimeter data, as well as TOBs from the L0Calo and L0Muon, to run offline-like algorithms on custom FPGA hardware. The refined TOBs will then be available as input to topological algorithms, as the GT will replace and extend the functionalities of the L1Topo system. The GT will produce new TOBs and trigger conditions that are 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 is composed of a large CPU-based processing farm and an HTT co-processor designed to provide fast hardware-based track reconstruction. A fast initial rejection is provided by regional tracking based on TOBs received from the L0 trigger. This first selection will bring the event rate from 1 MHz down to 400 kHz. Next, software-based reconstruction will achieve further rejection bringing the rate down to 100 kHz. Around 100 kHz, the HTT will provide global tracking for a more refined track reconstruction useful for b-jet tagging , E_T^{miss} soft term calculation, and pileup suppression.

4.4 ATLAS Event reconstruction

[96] The reconstructed final state of a collision in ATLAS includes electrons, photons, muons, τ-leptons, jets, and missing transverse energy E_T^{miss}. Except for muons, reconstruction of all the other objects requires calorimeter information. Fig. 4.21 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 degree of curvature of the tracks provides charge and momentum information [7]. Neutral particles like photons and neutrons do not interact with the ID. Electromagnetic interacting particles, like electrons and photons, are stopped in the electromagnetic (EM) 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 have a very small probability of interacting with matter and escape the detector volume undetected. Their presence in the event is inferred from a momentum imbalance on the transverse plane, by reconstructing the total missing transverse energy E_T^{miss}. Because of conservation of momentum on the transverse plane, a non-zero E_T^{miss} indicates that some object(s) is unaccounted for. In the following we will briefly review the reconstruction of the physics objects relevant for this thesis. Jets will be treated separately and more extensively in the next section. The algorithms discussed here are developed by the ATLAS Combined Performance groups, which provide working points (WPs), calibrations, and general recommendation for all physics analyses.

4.4.1 Tracks and vertices

The goal of track reconstruction [98], or *tracking*, is to reconstruct the trajectory of charged particles from the electronic signals, or *hits*, left in the ID. Tracking is a pattern recognition

[7] A particle of charge q moving with momentum p perpendicular to a magnetic field with strength B follows a circular trajectory of radius $r = \frac{p}{qB}$

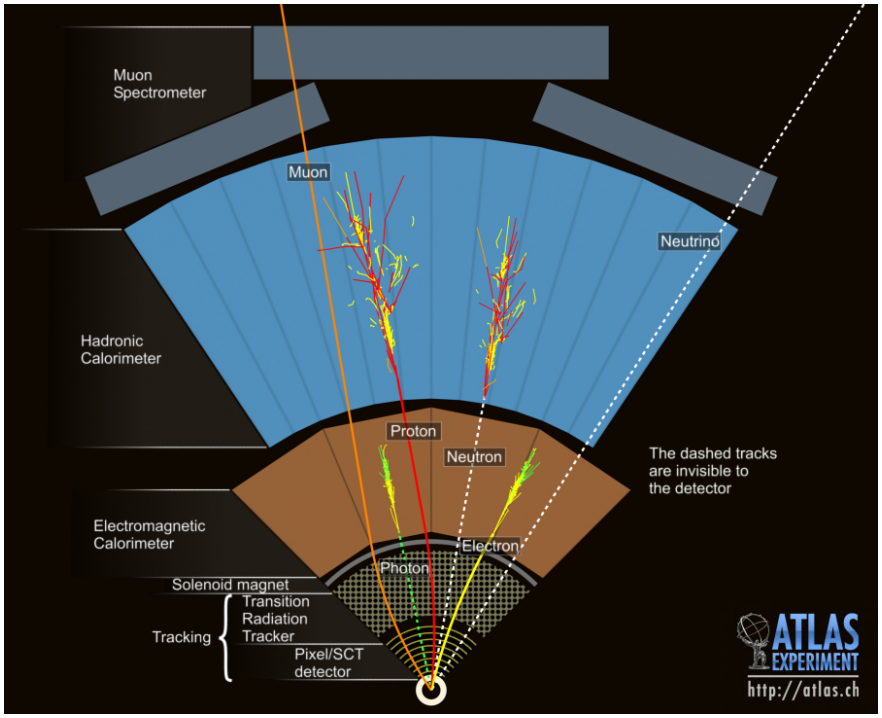


Figure 4.21: Cross-section of the ATLAS detector with simulated particles trajectories. [97]

task made more difficult by the busy environment of the ID, including in-time and out-of-time pileup, and collimated tracks. Track reconstruction starts from seeds made of tracks with three hits recorded in the Pixel detector or the SCT detector. The seeds are then extended to include further hits to create track candidates. An ambiguity resolution step is then applied, to remove overlaps or wrongly assigned hits. Finally a χ^2-based track fit is performed and only tracks with $p_T > 400$ MeV which pass 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* (PVs) [99] in a given collision, by iteratively associating the reconstructed tracks with $p_T > 500\,\text{MeV}$. Primary vertex candidates are required to have at least two reconstructed tracks with $p_T > 500\,\text{MeV}$ and

to be compatible with the interaction region of the proton beams [100]. The number of reconstructed PVs is usually $N_{PV} > 1$. The *hard-scatter* vertex is the hardest primary vertex, defined as the vertex with the largest sum of p_T^2 of the tracks associated to it. The other PVs are assumed to be produced by in-time pileup.

4.4.2 Electrons

The ability to identify electrons originating from prompt decays of W and Z bosons is essential to the ATLAS physics programme, as the clean signature left in the detector provides a tool for efficient triggering and background rejection. Hence, several analyses, including the one discussed in this thesis, rely on the efficient identification of prompt electrons from misidentified hadrons, electrons produced by photon conversions, and non-isolated electrons from heavy-flavor decays.

Electrons and photons are stopped in the electromagnetic calorimeter (EMCal). The interaction with the lead absorber initiates EM showers, which ionize the LAr in the active gaps. The applied voltage causes a drift of the ionization electrons, which produces an electronic signal proportional to the energy deposited by the shower. The energy depositions in the calorimeter are reconstructed as topologically connected clusters of cells. Before entering the EMCal, electrons and photons travel through the ID, but only electrons leave tracks. A schematic view of the path of an electron traveling through the detector is shown in Fig. 4.22. Almost 40% of photons convert to electron-positron pairs (*converted photons*). An electron can lose significant amount of 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

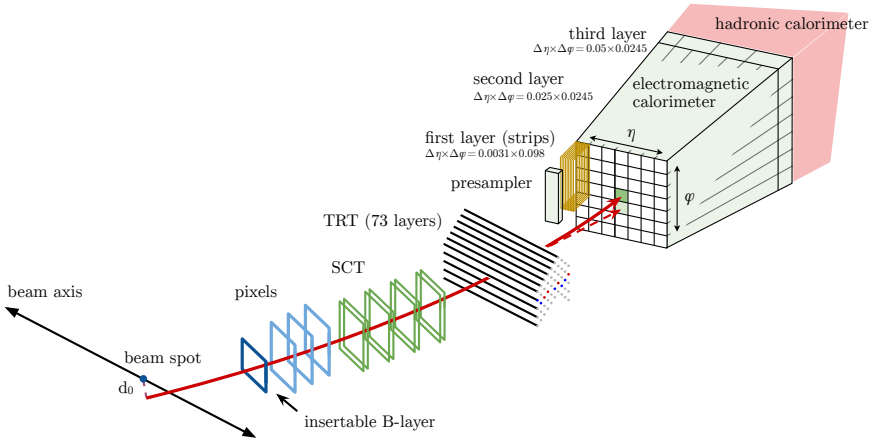


Figure 4.22: Illustration of the trajectory of an electron through the detector, shown by the solid red path. The dashed red line represents a photon emitted via bremsstrahlung radiation. [101]

they are contained in the collimated electromagnetic shower.

In 2015 and 2016 electrons and photons were reconstructed using a sliding window algorithm **??** seeded by calorimeter towers. However, since 2016 a new algorithm based on topological clusters was introduced. The new algorithm **??** produces variable-size clusters, as opposed to fixed size towers, which 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. 4.4.5) that are primarily generated by showers in the EM calorimeter, by requiring the EM fraction $f_{EM} > 0.5$. From these EM clusters, a set is selected as seeds of possible electrons and photons, and superclusters are formed by associating nearby EM clusters that share the same originating 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 (and

photons) are further cleaned via quality criteria based on a likelihood discriminant. To cover the different signal efficiency and background rejection needs of ATLAS analyses, four sets of electron identification criteria have been defined, called *VeryLoose*, *Loose*, *Medium*, and *Tight*, with increasing background rejection power. Isolation requirements are also defined to suppress background from hadrons faking electrons.

4.4.3 Muons

Muon reconstruction [102] is based on complete 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 gives the final *combined muon*. Other reconstruction strategies are used in order to retain efficiency in regions outside of the ID or MS acceptance regions. *Inside-out* muons are reconstructed from extrapolating ID tracks into the MS, where it is 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. *Segment-tagged* muons are reconstructed only from ID tracks that satisfy tight requirements on hits in the MS, but only the ID information is used to obtain the muon parameters. *Muon spectrometer extrapolated* tracks are reconstructed using only MS tracks and are used to extend the acceptance outside of the $|\eta| < 2.5$ region covered by the ID. Lastly, *calorimeter-tagged* muons are used to compensate for the MS inefficiency mainly in the $|\eta| \sim 0$ gap region, and are reconstructed by matching ID tracks to calorimeter energy depositions in the last layer of the calorimeter consistent with a minimum-ionizing particle signature.

After reconstruction, identification criteria are applied to select the highest quality tracks. Different WPs are provided, tailored to the different selection efficiency and purity require-

ments of the physics analyses. Muon candidates are separated into prompt muons (muons originating from the interaction vertex), and non-prompt muons (muons originating from secondary decays). Of the latter category, a distinction is made between non-prompt muons originating from bottom and charm quark decays, which provide good quality tracks closer to the primary vertex, and decays from light hadrons, which typically results in lower quality tracks due to the change in direction caused by the in-flight decay. The prompt muon efficiency is defined as the probability that a prompt muon is reconstructed as a muon and passes the WP selection. The purity of the selection is given by one minus the probability of incorrectly identifying as prompt-muon a non-prompt muon originating from a light hadron decay. The WPs used in ATLAS are, in order of decreasing efficiency and increasing purity: *Loose*, *Medium*, and *Tight*. Special WPs are provided for analyses targeting more exotic regions of phase space, where muons can reach very and and very low p_T. These are the *Low* – p_T and *High* - p_T WPs, respectively.

4.4.4 Missing transverse momentum

Missing transverse momentum (E_T^{miss}) [103] is an important proxy to identify the production of stable weakly interacting particles in the hard scatter. Such particles escape the experimental volume without leaving any detectable signal behind and include neutrinos, as well as possible new BSM particles.

As the colliding partons travel along the z-axis, the initial transverse momentum components of the system are zero. Conservation of momentum on the transverse plane implies that the vectorial sum of the transverse momenta of all the objects emerging from the hard scatter $\mathbf{p}_T^{HS}$ should also produce zero x and y components. Indicating the contributions from all the observable electrons, photons, taus, muons, and jets, and the non-observable

(invisible) particles, $\mathbf{p}_T^{HS}$ can be calculated as,

$$\mathbf{0} = \mathbf{p}_T^{HS} = \underbrace{\sum \mathbf{p}_T^e + \sum \mathbf{p}_T^\gamma + \sum \mathbf{p}_T^\tau + \sum \mathbf{p}_T^\mu + \sum \mathbf{p}_T^{jet}}_{\mathbf{p}_T^{\text{obs}}(\text{observable})} + \underbrace{\sum \mathbf{p}_T^\nu}_{\mathbf{p}_T^{\text{inv}}(\text{not observable})} \quad . \quad (4.4)$$

It follows that, in an ideal detector, any deviation from zero indicates the presence of a particle (or particles) that eluded detection with transverse momentum $\mathbf{p}_T^{\text{inv}} = -\mathbf{p}_T^{\text{obs}}$.

In practice, due to the 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 particles and jets that pass certain kinematic threshold and reconstruction quality criteria. In general $\boldsymbol{E}_T^{\text{hard}} < \boldsymbol{p}_T^{\text{obs}}$. To partially recover this loss, an additional soft-term $\boldsymbol{p}_T^{\text{soft}}$ is included, built from reconstructed charged-particle tracks coming from the hard-scatter vertex, but not associated to any hard object. [8] The missing transverse momentum observable $\mathbf{E}_T^{miss} = (E_x^{miss}, E_y^{miss})$ is then calculated as the negative vectorial sum of the hard and soft terms:

$$\mathbf{E}_T^{\text{miss}} = -(\mathbf{p}_T^{\text{hard}} + \mathbf{p}_T^{\text{soft}}), \quad (4.5)$$

From the transverse vector components one can calculate the magnitude and direction in the transverse plane in terms of the azimuthal angle ϕ as

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}, \quad (4.6)$$

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

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

As the hard objects are reconstructed and calibrated independently, it is possible that different objects share common energy contributions, such as topoclusters contributing both to a jet and to an electron. For this reason, a *signal ambiguity resolution* procedure is implemented in order to avoid double counting of energy.

Important quantities to estimate the event hadronic activity are also $\sum E_T$, the scalar sum of the transverse momenta of all the hard and soft contributions to the $E_T$miss calculation, as well as H_T, the scalar sum of the transverse momenta of all the reconstructed hard objects only.

In practice fake E_T^{miss} can arise due to limitations of the 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} having been produced in the event, is given by the significance $\mathcal{S}$. This is calculated with respect to the event activity as,

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

Another more recent object-based definition calculates the significance as a likelihood ratio to test the hypothesis $\mathbf{p}_T^{\text{inv}} = 0$ and $\mathbf{p}_T^{\text{inv}} \neq 0$. This is described in Ref. [104] and is the one used in the analysis discussed in this thesis to select events with true neutrinos.

4.4.5 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}. In general a cluster can contain the full or partial response to one or multiple signal particles. The clusters

are three-dimensional thanks to the longitudinal segmentation of the sampling layers, which permits to resolve signal-induced energy flow structures from the background noise.

The topoclustering algorithm **??** starts by evaluating the significance of each cell as,

$$\mathcal{S} = \frac{E^{EM}_{\text{cell}}}{\sigma^{EM}_{\text{noise,cell}}}. \tag{4.9}$$

where σ is given by Eq. (4.3). Both the cell energy and noise are evaluated at the EM 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. At this point the formation stops. The resulting clusters have a high $\mathcal{S}$ core, which differentiates them from background noise, while retaining the outer layer to not lose signals that are closer to the noise level. The output of the three steps of the clustering process is simulated for a di-jet event in the first FCal later in Fig. 4.23.

Calorimeter cells can have negative energy signals if induced by out-of-time pileup that occurred 100 ns before the event, due to the shaping of the calorimeter signal. 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 also negative energy cells in the clustering

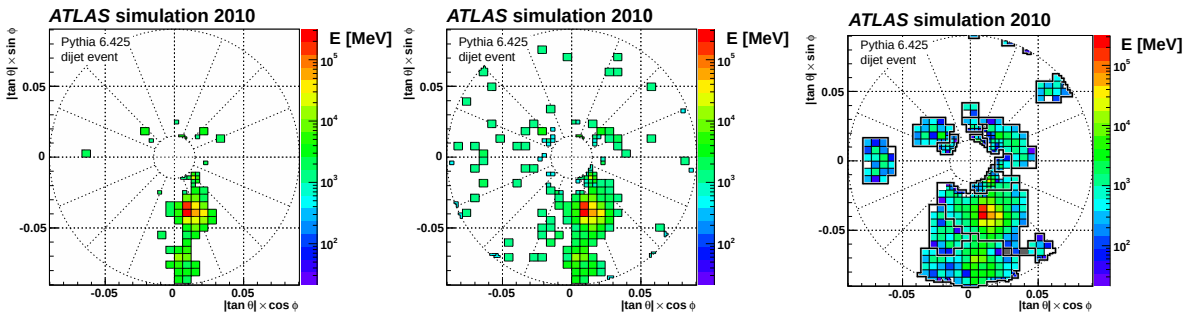


Figure 4.23: Stages of topological clustering in the FCal0 layer simulated for a di-jet event with at least one jet entering the FCal calorimeter and no simulated pileup. From left to right: all seed cells with $|\mathcal{S}| > 4$ starting a protocluster; all neighboring cells with $|\mathcal{S}| > 2$ are added recursively to the protocluster; all neighboring cells with $|\mathcal{S}| > 0$ are included. Topocluster fragments visible in this module but not associated to a seed, are seeded in other surrounding calorimeter layers. [92]

process, as these positive and negative noise fluctuations 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 other algorithms such as 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 cells four-vectors. Including the negative energy cells in this calculation would distort the calculation, to the point of projecting clusters on the opposite side of the detector. However, not including the negative energy cells 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. [92]. Once the basic kinematic variables ($\eta_{\text{clus}}, \phi_{\text{clus}}, E^{EM}_{\text{clus}}$) are calculated, the final 4 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. The EM scale correctly reconstructs electromagnetic (γ/e) energy depositions, but does not correctly represent the

hadronic depositions, as it does not account for the non-compensating calorimeter response to hadrons. The topoclusters need therefore to be calibrated to properly represent the hadronic scale. Calibration is also important to compensate for the inefficiency due to the loss of low $\mathcal{S}$ signal clusters because of the clustering strategy, which is dependent on the pileup conditions. The calibration is referred to as "Local hadronic Cell Weighting" (LCW) and consists in a series of corrections to iteratively reweigh the cell energy. It is performed using simulation of neutral and charged pions, representative of electromagnetic and hadronic showers, respectively. The calibration exploits information from cluster moments to identify the nature of the shower generating the signal and applying corrections to the cells' energy according to the probability of each cell of originating from a hadronic shower. The final calibrated cluster energy is given by $E^{\mathrm{LCW}}_{\mathrm{clus}} = \sum_{i \in \mathrm{cluster}} w^{\mathrm{LCW}}_{\mathrm{cell},i} E^{\mathrm{EM}}_{\mathrm{cell},i}$, with $E^{\mathrm{LCW}}_{\mathrm{clus}} \geq E^{\mathrm{EM}}_{\mathrm{clus}}$ and the cluster (η, ϕ) directions recalculated accordingly.

4.4.6 *b*-tagging

The identification of jets[9] containing bottom or charm quarks is an important step in LHC physics, 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 the b-flavoured mesons is of the order of $1.5 \times 10^{-12} s$ ($< c\tau > \sim 4.5$ mm), which corresponds to a mean flight length of $< l >= \gamma\beta c\tau$ before decaying. At the LHC energies, this is of the order of a centimeter [105], a sizable distance observable in the ID as a displaced vertex: a certain number of tracks points to a secondary vertex, with a large impact parameter. The secondary decay system of the B meson is usually well described by

[9] Jets are discussed in detail in Chap. 5

the decay of the b quark inside the hadron (spectator model), and proceeds in most cases via $b \to cW^-$, with the virtual W decaying either into a pair of leptons $l\bar{\nu}$ or into a pair of quarks, which then hadronizes [106]. Thus, B meson decays into at least one c-flavoured hadron (a D meson), which then decays further also with an appreciable lifetime, producing a characteristic topological configuration with two secondary vertices.

The tagging of b-jets relies on the track reconstruction of the displaced B meson decay. Because of the need of ID tracks, b-tagging algorithms can only be used for jets within the ID acceptance $|\eta| < 2.5$. ATLAS adopted a two stage approach for the identification of b-jets. In the first stage, a series of low-level algorithms [107] 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`[108] algorithm [10] , a high-level tagger which includes as input also the output probabilities from the `RNNIP` algorithm [109]. 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 given by

$$D_{\text{DL1}r} = \ln\left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{\text{light}}}\right), \tag{4.10}$$

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

[10] Historically, two high-level taggers were available: the *MV2* boosted decision tree classifier, and the *DL*1 deep feed-forward neural network. The DL1 algorithm was introduced at the beginning of Run 2 and has now evolved into the `DL1r`. The `DL1r` algorithm achieves the best tagging performance and is the current recommendation for physics analyses.

Chapter 5

Collider physics

Thanks to the unprecedented center-of-mass (CoM) energies of the LHC, the ATLAS experiment can probe the SM over scales ranging from 10 GeV up to 10 TeV. The experimental reach of collider physics is vast and relies on two fundamental principles [96]: 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 result in the production of heavier particles, according to Einstein's equation $E = mc^2$ [1]. 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.

A representative example of a *pp* collision at the LHC is shown in Fig. 5.1. 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

[1] A third important principle is that high energy collisions should allow to probe the conditions in the early universe according to Boltzmann relation between energy and temperature $E = k_B T$ [96].

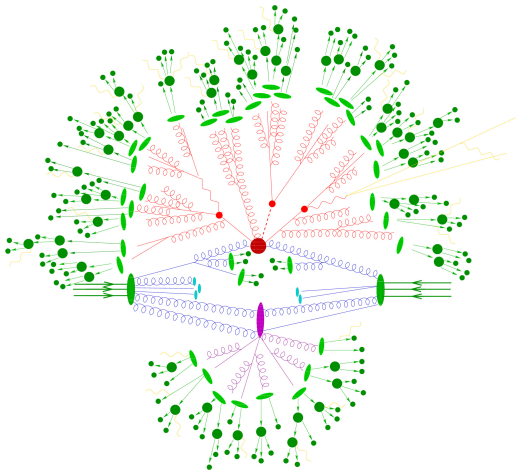


Figure 5.1: Example of a proton-proton collision. In black are the incoming colliding protons whose partons are described by parton distribution functions (PDFs). The initial state radiation of quarks and gluons (in blue) interacts with a fraction of the energy of the incoming protons. In red is the hard interaction, with additional hard QCD emissions. After parton showering and hadronization the intermediate hadrons (light green) are formed. These further decay into the final state stable particles (dark green). In purple is a secondary interaction characterizing the underlying event. Additional radiated photons are shown in yellow and the beam remnants in blue-green. [110].

to study. However, on top of this hard process several effects related to QCD dress up the event.

The partons confined in the incoming protons, as well as the hadronization of the final state partons into the confined hadrons detected in the experimental apparatus, are at a much lower scale than the hard scattering process, of the order of 1 GeV, where perturbation theory is not valid. Any process induced by the colliding protons of the hard scatter of interest but not involving the hard scatter is referred to as the underlying event (UE). This includes multiple secondary scatterings and beam-beam remnants.

The partons confined in the incoming protons are at a much lower scale than the hard scattering process, of the order of 1 GeV, where QCD is non perturbative. The primary partons, 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 colour singlet states, with unstable hadrons decaying further. The final stable hadrons (with lifetimes $\tau > 10ps$) 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.

Typically, the final state of such collisions involves hundreds of particles and the process to go from the colliding partons to the particles observed in the detector spans a wide range of scales. Exact perturbative methods cannot be applied at the low scales that enter the problem, where the validity of QCD as a perturbative theory comes short and confinement becomes important. These processes cannot be calculated theoretically, and have to be modelled and fit to data. The high final state multiplicities make matrix element computations in the perturbative regime often too complex to be calculated exactly, while the phase space integrals needed to compute expectation values of observables have prohibitively high dimensions. As we will see, 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. In the renormalized theory, observable predictions are expressed in terms of the renormalized coupling $\alpha_s(\mu_R^2)$, where μ_R is the renormalization scale, often taken to be the scale of the momentum transfer Q of the process. The running of the strong coupling is governed by the renormalization group equation,

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

where $b_0 = (33 - 2n_f)(12\pi)$ and n_f is the number of quark flavours relevant at the given scale.

The renormalization group equation 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 $\alpha_s(M_Z^2) = 0.12$ (measured at the reference scale $Q^2 = M_Z^2$). Eq. (5.1) is solved as

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

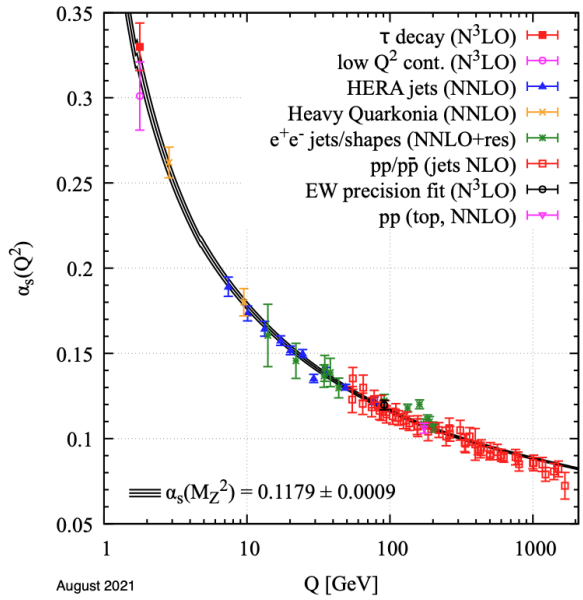


Figure 5.2: Measurements of α_s as a function of the energy scale Q. The order in perturbative QCD used in the extraction of α_s is indicated in parenthesis. [111].

The negative sign in Eq. (5.2), combined with the fact that $b_0 > 0$, causes the coupling to effectively decrease with increasing energy, as shown in Fig. 5.2. For large momentum transfers and small distances, QCD becomes almost a free theory, a phenomenon known as *asymptotic freedom*. In this regime, where $\alpha_s << 1$, perturbation theory is applicable and QCD is well behaved. On the other hand, at small momentum transfers or large distances, the coupling diverges and QCD becomes non perturbative. The scale at which QCD becomes non-perturbative is called Λ_{QCD} and it is known from experiment to be $\approx 200\,\text{MeV}$.

The fact that the coupling diverges at large distances prevents quarks from ever being observed alone, but only as colour singlet bound-states, mesons or baryons. This phenomenon is called *confinement*. As two quarks are pulled apart, the potential energy between them increases without limit, up to the point that enough energy is available to produce a new quark anti-quark pair. For the purpose of LHC physics, confinement plays a fundamental role in modelling the transition from free quarks and gluons in the final state of a hard

scatter to the bound-state hadrons actually observed in the detectors.

5.1.2 The hard scatter cross section

In high energy proton collisions one is faced with the fact that hadrons are composite. The quark model describes the hadron as a sea of strongly interacting quarks and gluons, where $q\bar{q}$ pairs 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 only the *valence quarks*, which can be described, at first approximation, as the quarks whose net number is non-zero. The calculation of the cross-section of a pp collision has to take into account this structure.

When a hard perturbative probe interacts with a proton, there are two scales at play. One is the soft long-distance physics of the proton structure and the second is the high energy short-distance physics of the hard process. While the latter is calculable in perturbative QCD, the first momentum scale 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 [2]. The total cross section for a process $q\bar{q} \to f$ in a hadron-hadron collision can be expressed as [112]

$$\sigma = \int dx_1 f_{q_i/p_1}(x_1, \mu_F^2) \int dx_2 f_{q_j/p_2}(x_2, \mu_F^2) \hat{\sigma}_{q_i q_j \to f}(x_1 p_1, x_2 p_2, \mu_F^2), \tag{5.3}$$

where p_1 and p_2 are the colliding hadrons, μ_F is the factorization scale at which the hadrons are probed, $\hat{\sigma}_{q_i q_j \to f}$ is the parton-level cross section for the production of the final state f through the initial partons q_i and q_j, and the functions $f_{q/p}(x_n)$ are the PDFs. The

[2] The factorization theorem was initially formulated for deep inelastic scattering, but it can be equally valid in pp and $p\bar{p}$ collisions [112].

PDFs represent (at first approximation) the number density of quarks of type q_i carrying a fraction x_n of the momentum of the n^{th} proton. There is some arbitrariness in choosing a value for μ_F, but for matrix elements characterized by a well defined scale, such as the invariant mass scale $\hat{s}$ in Drell-Yan production, usually one fixes $\mu_F = \sqrt{\hat{s}}$ [113].

5.1.3 Parton distribution functions

The factorization theorem can be intuitively understood by the fact that the hard interaction occurs over a much shorter time scale than the fluctuations inside the proton structure, so that from the point of view of the hard scatter, the quark sea appears frozen. The PDFs are therefore decoupled from the short distance physics and their shape can be treated as universal, or process-independent. Therefore, although not calculable *a priori*, the PDFs 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 evolution [3]. Fig. 5.3 shows the PDFs behavior for gluons and the different quark flavors inside a proton as a function x for $Q = 2$ GeV and $Q = 100$ GeV, where μ is taken to be equal to Q. While at low μ the valence up and down quarks carry a significant portion of the proton momentum (shown by the clear peaks at high x values), at high μ the gluon and sea-quark distributions have much more significant contributions, 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. 4.14, cross sections for gluon-initiated processes have a steeper slope of increase with increasing center-of-mass energy than quark initiated ones [96]. It should be noted that different collaborations use different functional forms for the PDFs and may also constraint the fit using different datasets.

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

In the context of ATLAS physics, PDFs modelling contributes significantly to the systematic uncertainties of many analyses.

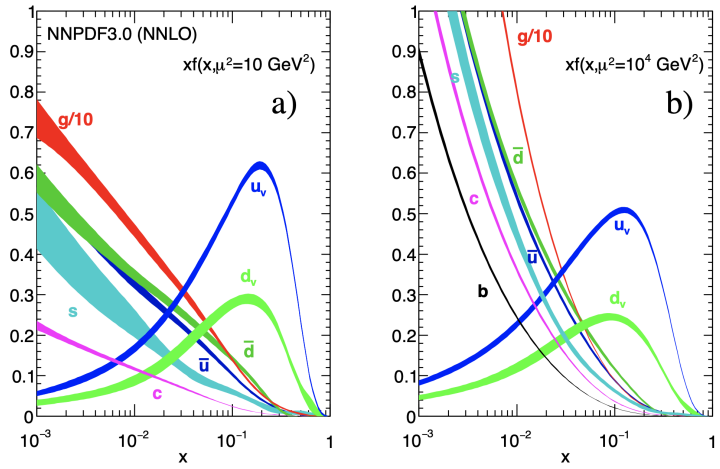


Figure 5.3: Parton distribution functions obtained by NNLO NNPDF3.0 global analysis illustrating the gluon and quark flavor contributions to the proton composition as a function of x at $Q = \mu_F = 10$ GeV (a) and $Q = \mu_F = 100$ GeV (b). Note that a factor of 0.1 is applied to the gluon PDF. [29].

5.1.4 Matrix element

The cross-section of an interaction is calculated using two main ingredients: the amplitude, or matrix element, $\mathcal{M}$ and the phase space available for the interaction to occur. The matrix element represents the probability amplitude for the transition from an initial state i to a final state f to occur and can be calculated with Feynman diagrammatic techniques. The phase space integral represents the kinematics available to the participating particles. It is sometimes referred to as the *density of final states*, which intuitively conveys the idea that the more final states (with each final state defined by different energy, momenta, and mass sharing of the final state particles) are accessible, the larger the contribution of the phase space term to the cross section, and therefore the more likely the given process. In practice, the hard-scatter cross-section $\hat{\sigma}_{i \to f}$ is obtained by taking the absolute value squared of the

matrix element, summing over all possible polarizations and colour states, and integrating over the phase space.

The cross section can be calculated at fixed order in perturbation theory as

$$\sigma = \sigma^0 + \alpha_s \cdot \sigma^1 + \alpha_s^2 \cdot \sigma^2 + O(\alpha_s^3), \tag{5.4}$$

where one approximates the series up to a given order n in the strong coupling α_s, with the assumption that the contribution from the omitted higher orders should be small. The σ coefficients are the hard parton-level cross sections calculated at a given order and the higher the order that is included, the higher the precision of the calculation. The leading order (LO) contribution σ_0 is the Born level cross section, σ_1 is the next-to-leading order (NLO) correction, σ_2 the next-to-next-to-leading order (NNLO), and so on. Each power of α_s corresponds to the addition of new diagrams including an extra real or virtual emission, starting from the LO diagram with no emissions.

Specifically, if one defines $\mathcal{M}^l_{F+k}$ the amplitude for the LO process F produced in association with k final state partons ($F + k$ jets) and with l additional loops, the order of the perturbative expansion is given by $n = k + l$.

The cross section calculated at NLO accuracy contains the contribution from diagrams with one gluon emission or one gluon loop. These diagrams introduce different types of divergences, which have to be regulated to preserve unitarity. Diagrams with loop corrections introduce ultraviolet (UV) divergences in the limit of the loop propagator momentum $q \to \infty$. 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. One common method is the $\bar{M}S$ renormalization scheme, which assumes a

cutoff scale up to which perturbation theory is valid, hence removing the problematic region of phase space and making the integral finite.

Both virtual and real-emission diagrams exhibit infrared and collinear divergences in certain corners of the phase space, where the emitted gluon with momentum k is soft ($k \to 0$) or collinear ($\theta \to 0(\pi)$). However, as proven by the KLN theorem, order-by-order unitarity implies that the 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. Therefore, as long as both contributions are included, the calculation of the $k + l = n^{th}$ order is finite [113]. 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 observed cross section. However, it is desirable not to have theoretical calculations based on experiment-specific and potentially variable parameters. It is therefore desirable to study observables for which the KLN theorem holds, called *IRC-safe* observables.

5.1.5 Showering and hadronization

A colour-charged particle produced in the hard scatter will start at a high factorization scale and move towards a lower scale via bremsstrahlung radiation, i.e. via the quark and gluon splittings $q \to qg$, $g \to gg$, and $g \to q\bar{q}$. Because the emission rate for these processes diverges in the collinear (opening angle $\theta_{qg/gg} \to 0$) and soft (gluon momentum $k_g \to 0$) limit, these will be 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 tells us that these particles have to undergo some transition from free coloured partons to colour singlet hadrons, with unstable primary hadrons subsequently decaying

further. The process is called hadronization.

Although the quark and gluon emissions occur in the perturbative regime, the high parton multiplicities would require matrix elements 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. For these reasons, event generator packages use 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 to the hard scattering process.

Parton shower generators are based on the principle that at first approximation the functions governing quark and gluon splittings are universal and can be fully factorized from the hard scatter cross section calculation into what are called *fragmentation functions*. These are the equivalent of PDFs for the final state and parametrize the probability of finding a hadron h within the fragmentation products of parton k, carrying a fraction z of the parton momentum [114]. Similarly to the PDFs, they depend on a factorization scale μ_F and, although they cannot be determined from first principles, they can be fitted to data. Parton showering programs rely also on the fact that, at first approximation, the parton decays can be considered independent, except for being ordered in a decreasing sequence of scales: a parton shower will start with high virtuality and large angle emissions and move towards small virtuality and small angles [115].

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 very 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 generally combined to provide an exact description of the hard scattering process of interest, including any hard jet emission, and a probabilistic description of the soft emissions and hadronization, as discussed later.

5.1.6 Soft hadron-hadron physics

Due to the composite nature of protons, each pp collision contains several few-GeV parton-parton collisions. These are called *multi-parton interactions* (MPIs). In addition, each colliding parton may also leave behind a *beam remnant*, which does not take part in the initial state radiation or hard-scattering process, but still remains colour connected to the rest of the event. The UE describes any process that accompanies a hard inelastic scatter, such as beam-beam remnants and MPIs. To first approximation the UE radiation is soft and evenly distributed in rapidity, appearing as a background noise in the detector. While the processes that make up the UE are similar to the soft interactions that dominate a Minimum Bias sample[4], the two are not the same thing. The definition of the UE requires a hard scatter to have occurred, i.e. the UE is the background to a hard-scattering event. While the UE is directly correlated to the final state via colour connections, its contribution is understood only phenomenologically from data due to its non-perturbative nature. In particular, the UE is related to the *pedestal effect* observed in distributions of hard jets kinematics: here a constant E_T plateau is observed, substantially higher than the one observed in a minimum bias event. The larger activity of the UE, as compared to Minimum Bias, is explained by a

[4] *Minimum bias* is a general term used to describe a data sample collected by an experiment using a "minimum bias trigger" (see Sec. 4.3.3). The definition of such a trigger is experiment-related, but it generally indicates a loose trigger that does not pose any strong requirement on hard activity. The resulting samples includes a mixture of soft and hard processes, with soft events dominating.

trigger induced bias. The trigger selection of a hard jet biases the event selection towards more central collisions, which are associated to a larger number of MPIs and increased activity.

5.1.7 Monte Carlo event generators

Monte Carlo (MC) simulations are an essential part of the ATLAS physics program, as they allow to study the feasibility and sensitivity of new analysis methods, isolate specific signal or background components with targeted phase space selections, perform calibrations between data and MC and, most importantly, provide the distributions for the background-only hypothesis in any statistical fit. A MC sample is a set of events corresponding to a given process. Each event represents the same hard interaction, but the distributions of the kinematics of the final state objects will vary according to the true probability distributions. The MC generation path is generally composed of different independent steps carried out by different MC simulation programs.

Parton-level

Matrix element generators are used to simulate pp collision events at the parton-level. One of the most widely used is Madgraph (Matrix element generator). 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 remains often 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 then computed using numerical integration and 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. They also allow for parameter tuning. These softwares 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, thereby breaking the string into two separate colour singlet pieces. At the end of this fragmentation process the colour connected partons are combined to create hadrons.

The most common method to simulate pp collisions is to combine LO matrix element predictions (e.g. Z+jet, Z+2-jet, etc.) with parton showers. Another possibility is to start from NLO matrix element calculations before interfacing with a parton shower generator. Such approaches are used by MC@NLO and POWHEG (Positive Weight Hardest Emission Generator). 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 take to avoid double counting

in overlapping regions of phase space, a process called *merging*, and is today only possible for certain processes. In ATLAS, production at NLO accuracy is performed for W, Z, and *top*-quark productions, where the loop corrections provide significant accuracy improvements. The generation of BSM signal samples is usually performed at LO in order to reduce the model dependence.

The stable hadrons at this point are referred to as *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` [116], which simulates their interaction with the different detector materials. The simulated detector outputs are then processed by a digitization step which 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 between a real data and a MC event is that the MC simulation retains the *truth information* about the hard process, including particle types, momenta, and decay chains [96]. At this point, the same reconstruction algorithms are run on data and MC events. The objects output by the reconstruction step are said to be at the *reco level*.

In the context of ATLAS performance studies it is important to remember the distinction between *parton level*, *hadron-level*, and *reco level*. Reconstruction algorithms can be fed input objects from any of these stages, according to the need. In particular, in the case of jet reconstruction, a *reco jet* is a jet produced by running a jet algorithm on the *reco level*

objects, while a *truth jet* is the output of a jet algorithm run on *hadron level* objects.

5.1.8 Jets

A *jet* is a collimated spray of particles resulting from hard quarks and gluons produced in the hard scatter. As we saw in the previous section, the final state of a hard scatter can be described in simulation by three equivalent levels of description: the final state partons of the hard process (*parton level*), the final stable hadrons before interaction with the detector systems (*hadron level* or *particle level*), and the observable energy depositions in the detector (*detector level* or *reco level*). A jet algorithm is an algorithm that takes a list of input objects - be them particles, hadrons, or energy deposits - and returns a list of new objects called jets.The fragmentation process that relates these three levels is very complex and the final composition of the jets — in terms of type, multiplicity and energy of the particles inside the jet — varies between events. Nonetheless, the final direction of the jet defined by its total momentum $\mathbf{p} = \sum \mathbf{p_i}$ is generally a good representation of the original direction of the parent parton [12]. Therefore, in principle there is a close correspondence between these three levels of description, as represented in Fig. 5.4. This makes jets important proxies to study the partonic dynamics of the collision and generally important tools in high energy particle colliders.

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 and often dependent on the physics one wants to study. As an example, consider a simple di-jet event $q\bar{q} \to q\bar{q}$, where two quarks are produced in the hard scatter accompanied only by soft and collinear emissions. The event will have two cone-like energy depositions in the detector associated to the two quarks and can be reconstructed in a straightforward

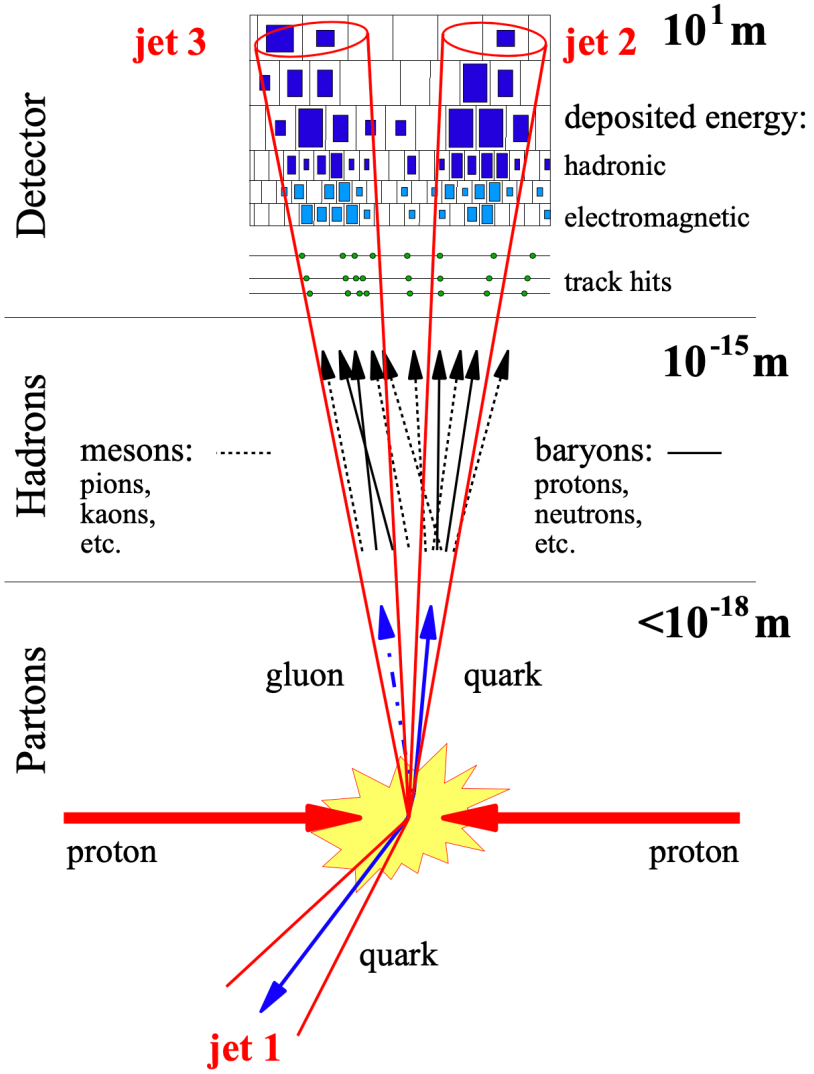


Figure 5.4: Illustration of the correspondence between a jet and the possible different groups of objects associated to it: the partons produced in the final state, the hadrons produced after showering and hadronization, the energy depositions in the calorimeter. [114]

way. In contrast, consider now the case where one of the two quarks emitted a hard wide-angle gluon. There is somewhat an ambiguity on whether this should be considered a two- or a three-jet event. This decision relies on how hard or separated in angle one believes an emission has to be for it to be considered a separate jet. This decision is made via the choice of jet algorithm and is dependent on what physics question one wishes to study [112]. 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, identification, and pileup suppression are particularly relevant for the work presented in this thesis and will be discussed in detail in the following sections.

5.2 Jet reconstruction algorithms

The reconstruction of jets depends on the *jet definition* and the jet inputs. The particles input to the algorithms are described by their 4-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 the constituents.

All jet algorithms can be classified according to two broad categories [112]:

- *Cone algorithms* rely on an event-level (top-down) approach where the jets are viewed as dominant directions of energy flow.

- *Sequential-recombination algorithms* are based on a bottom-up approach, where the closest particles are combined 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, which is less sensitive to non-perturbative effects and easier to calibrate. However, this changed with the development of the anti-k$_T$ algorithm [91], as it provided both an IRC safe and soft-resilient shape option.

5.2.1 Infrared-collinear safety

Ideally the set of hard jets reconstructed in an event should be insensitive to soft and collinear emissions characterizing the showering process. Experimentally, below a certain scale one has no way of distinguishing a parton from a parton plus a collinear or soft emission. In this way, detectors regularize the IRC radiation. 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.

An observable is *infra-red safe* if it is calculable in fixed order perturbation theory. This is true if two conditions are met

1. Safety against soft radiation: the observable does not depend on the number of soft particles in the final state.

2. Safety against collinear radiation: any number of collinear splittings should not change the value of the observable.

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 $\vec{p}_i \to \vec{p}_j + \vec{p}_k$, whenever $\vec{p}_j$ and $\vec{p}_k$ are parallel (collinear) or one of them is small (soft).

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 branchings. 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 flows [112]. In order to reduce the computational time, most implementations are *seeded* algorithms, where the seeds are selected to be representative of the energy flows in the event. A proto-jet is built around the seed, whose constituents are selected by drawing a cone of radius R around it[5]. The four-momentum of the jet is calculated from the kinematics of 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 obvious 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 ATLAS iterative cone with split-merge (ICSM) algorithm, implement a split-merge approach: if two overlapping jets share more than a

[5]As defined in Sec. 4.2.2, 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 $\Delta R_{\text{seed,object}} < R$.

fraction f of their energy, the jets are merged, otherwise the constituents are split among the two jets. The other type of algorithms builds 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. And example of this type of algorithm is the 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 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 will be harder than the two jets alone. While before we had two hard jets, now we have just one. In 2007 an IRC safe cone algorithm was developed called SISCone (Seedless Infrared Safe Cone) [117]. 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 all the different cone algorithms that have been used in the past see Ref. [118].

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 measure between particles and

iteratively combine the closest pair of particles until not particles are left. The inter-particle distance is given by,

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta R_{ij}^2}{R^2}, \qquad d_{iB} = k_{ti}^{2p}, \tag{5.5}$$

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 [119, 120], which combines soft and collinear particles first; $p = 0$ for the Cambridge/Aachen algorithm [121, 122], which clusters particles together only based on angular proximity; and $p = -2$ for the anti-k_t algorithm, which preferentially combines hard particles [91].

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 1.

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 [123]. 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 [6], 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 is produced in the event. If d_{iB} is the minimum distance, this will produced a new jet with $p_T \to 0$. If instead the particle is closest to another particle, the k_t distance will be dominated by the $1/k_{ti}^2$ term, so that $d_{ij} \to \infty$ and the soft particle will be clustered last. Similarly, a particle originating from a collinear emission will have $\Delta R_{ij} \to 0$, 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 H, 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

[6] This is based on the proof that, if a particle gets combined with another and the particle has the smallest k_t distance, then its partner is necessarily its geometrical nearest neighbor

close by particles get assigned to the jet or not.

Conversely to the k_t algorithm, anti-k_t clusters first hard objects that are close together, which ensures that the jet grows around a hard core. As new particles are added to the proto-jet, the jet axis can move slightly but, in the absence of other close by 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 $k_{t1} >> k_{t2}$, 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 between the jets will be a straight line equidistant between the two.

An illustration of the jet areas produced by the IRC-safe algorithms discussed in this section is shown in Fig. 5.5.

5.3 Jets in ATLAS

Being a hadron-hadron collider experiment, jets are ubiquitous at the LHC and are essential components to many SM measurements and searches for new physics. 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 is made of neutral hadrons [124]. Jets interact therefore with the inner detector (ID), before being stopped in the calorimeter, where they leave a cone-like energy deposition.

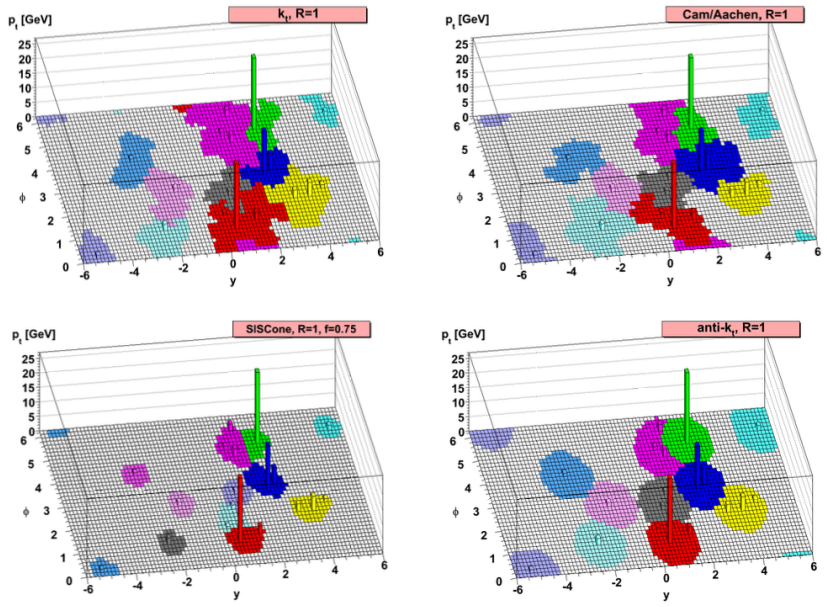


Figure 5.5: Illustration of the active areas of the jets output by the main IRC-safe jet reconstruction algorithms [91]: k_t(top-left), Cambridge/Aachen (top-right), SISCone (bottom-left), anti-k_t (bottom-left).

The standard for jet reconstruction in ATLAS is the anti-k$_T$ algorithm with radius parameter of $R = 0.4$ for small-R jets and of $R = 1.0$ 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-vector objects, which can be charged particle tracks from the ID or energy deposits in the calorimeter, or a combination of the two. Stable particles from MC generators can also be used for MC studies. Jets produced with different inputs are referred to as *jet collections*.

An important part for a meaningful physics interpretation is the calibration of the jets. This is necessary because of the non-compensating nature of the hadronic calorimeter, the need to suppress electronic noise, tracking inefficiencies, dead material in front of calorimeters, the influence of pile-up and other effects. Measurements of the jet energy scale and energy resolution, as well as associated systematic uncertainties, are essential components for

sensitive results. The Jet Energy Resolution (JER) is derived as the ratio E_{reco}/E_{true} and the jet energy scale (JES) is derived in bins of E_{true} and η. The result of the jet energy scale calibration restores the jet energy to the true energy at particle level. In-situ calibrations are also derived by comparing the jet response in data and MC simulation, and the ratio is used as an additional correction in data. This is done to account for potential differences caused by an imperfect simulation. The calibration chain is different for small-R [125] and large-R [126] 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, the calibration of large-R jets includes both energy and mass corrections.

Different techniques for pileup mitigation can also be used, both directly on the set of input objects (input-object level), as well as on the reconstructed jets (jet-level) to remove soft radiation. Lastly, an essential step in most physics analyses is the identification of the true particle a given jet originated from, referred to as *jet tagging*. Several algorithms have been developed in ATLAS in the context of heavy flavor and boosted large-R jet identification.

5.3.1 Jet algorithm

In ATLAS, jets are reconstructed using the anti-k$_t$ jet algorithm, as implemented in the `FASTJET` package [127]. 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 [96],

$$\Delta R \approx \frac{2m}{p_T}. \tag{5.6}$$

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 are collimated enough that they cannot be resolved as separated jets anymore. *Large-R jets* with $R = 1.0$ are used to reconstruct boosted decays of massive particles such as top-quark and W, Z, and Higgs bosons.

5.3.2 Jet inputs and jet collections

EMTopo

During Run 2 the standard inputs for jet reconstruction were topological clusters (see Sec. 4.4.5). According to the topoclusters energy scale, EM scale or LCW calibrated, the corresponding jet collection is referred to as *EMTopo* or *LCTopo*. This was possible thanks to the very good ATLAS calorimetry, which provides clusters with excellent 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 by ATLAS at the end of Run 2.

Particle Flow

The Particle Flow (PFlow) algorithm [124] relies on tracking information to improve the

performance for the reconstruction of low p_T charged particles. Tracks provide in fact a superior momentum resolution for low p_T particles and better angular resolution that allows to recover low p_T charged particles swept outside of 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 uses the track-cluster system to obtain the best position and energy measurement according to the resolution of the tracker and calorimeter in the given energy regime. A subtraction step leaves energy-subtracted modified clusters associated to the track-cluster system that could be originating from an overlapping shower. The final input objects (PFOs) consist of tracks, as well as the remaining modified clusters and the clusters not matched to any track, which 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 in Ref. [125].

Track-CaloClusters

The *Track-CaloClusters* (TCCs) [128] algorithm focuses on combining the tracker information on the angular energy flow, together with the calorimeter energy information, to resolve distinct particles associated to a single topocluster. This method improves the identification of substructure in large-R jets. However, TCC jets suffer of 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 [129], 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

Jet inputs can also consist of simply inner detector tracks, producing *track jets*. These are primarily used as proxies for b-jets contained in large-R jets. At the beginning of Run 2 the standard radius parameter for track jets was set to $R = 0.2$. This was later changed to a variable radius parameter $R(p_T) = \frac{30GeV}{p_T}$, 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. (5.6). The algorithm has two additional parameters, R_{min} and R_{max} [130], to set the lower and upper limits on the jet size. The resulting jets are referred to as *variable radius track jets*.

Truth jets

Truth jets, take as input generator level stable particles (hadron level). Truth jets are very important for performance studies, such as algorithm development or calibration. For these studies it is also often important to know the true parton a reco jet originated from. In order to reconstruct this from the final simulated event, *truth matching* is usually done: the reco-jet is matched to the closest truth jet, and the truth jet is matched to the closest parton, where matching usually consists in a ΔR requirement. The label of the reco jet will be the type of the matched particle.

5.4 Boosted jet tagging

Several BSM models predict new heavy resonances with large couplings to the SM bosons as a way to address naturalness issues (see Chap. 3). These resonances are expected to appear at masses of the order of 1 TeV and decay into highly Lorentz boosted SM bosons. Since in more than 60% of the cases $W/Z/H$ 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 in Chap. 7.

A tagger attempts to identify the true particle a jet originated from. 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 $1/\epsilon_\mathrm{b}$[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 bosons, the radiation pattern is generally isolated from the rest of the event. Some ambiguity remains related to how much of the radiation originating from the jet is contained in the jet. Containment definitions used to select training samples typically involve the use of truth information from the parton-level of the MC simulation. This is more complicated for jets originating from coloured particles, for which a formal separation of the decay from the rest of the event is not possible [131].

The reconstructed mass of a jet is one of the most important discriminators between

[7] In other fields background rejection is more commonly defined as $1 - \epsilon_\mathrm{b}$.

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 the parton fragmentation and hadronization process leads to characteristic hadron multiplicities and kinematic distributions inside jets of different origin. Jet substructure [132, 133] is a relatively new field that aims at exploiting the radiation pattern inside jets as a tool to identify boosted particle decays, as well as to perform precision tests of QCD. Several substructure observables rely on the identification of the dispersion or of the clustering around directions of energy flows in the jet radiation pattern. Some techniques, like N-subjettines, rely on the identification of hard subjects as explicit axes. Similarly, declustering techniques, which walk in reverse the jet clustering history, can be used to identify hard subjets. Other jet-shapes methods, like energy correlation functions, angularities, planar flow, and Fox-Wolfram moments, can be used to 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 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,

$$\sqrt{d_{ij}} = \min(p_{T,i}, p_{T,j} \times \Delta R_{ij}). \tag{5.7}$$

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 subjects. The variables $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ are helpful in distinguishing the two- and three-prong decays of heavy particles, which show a more symmetric energy sharing between the subjets than the splittings that originate from QCD radiation in q/g-jets.

N-subjettiness

The N-subjettiness [134] observables τ_N are also obtained by reclustering the jet constituents using the k_t algorithm in order to identify the N hardest subjets. The variable τ_N is then calculated as,

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}\}, \tag{5.8}$$

where d_0 is the normalization factor $d_0 = \sum_k p_{T,k} R_0$. The result can be interpreted as a metric of how good the hypothesis that the jet has N hard subjets is. The subjets identify axes of energy flow in the jet. 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 $\tau_N >> 0$, as a larger number of constituents will be at a larger distance from the identified subjet directions. For instance, a jet originating from a W decay, with two subjets, will have $\tau_1 >> 0$ and $\tau_2 \approx 0$. On the other hand, a QCD jet containing two hard quarks can have $\tau_2 \approx 0$. However, a QCD jet is characterized by a diffuse spray of wide angle radiation so that, even in the presence of two hard subjets, these are accompanied by more radiation, determining a correlation between τ_2 and τ_1. For this reason, the N-subjettiness ratio $\tau_{21} = \tau_2/\tau_1$ is usually used to discriminated between a two-

hard-prongs decay, like a W-decay, and a q/g-initiated jet. Similarly, the ratio $\tau_{32} = \tau_3/\tau_2$ is used to identify 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 the finding of subjets. For a hadron collider, the N-point ECF is defined as

$$\mathrm{ECF}(N,\beta) = \sum_{i_1<i_2<\ldots<i_N\in J} \left(\prod_{a=1}^{N} p_{T,i_a}\right) \left(\prod_{b=1}^{N-1}\prod_{c=b+1}^{N} \Delta R_{i_b i_c}\right)^{\beta}, \tag{5.9}$$

The one-,two- and three-point ECFs for a jet J at a hadron collider are given by

$$ECF(1,\beta) = \sum_{i\in J} p_{T,i}, \tag{5.10}$$

$$ECF(2,\beta) = \sum_{i<j\in J} p_{T,i}p_{T,j}(R_{ij})^{\beta}, \tag{5.11}$$

$$ECF(3,\beta) = \sum_{i<j<k\in J} p_{T,i}p_{T,j}p_{T,k}(R_{ij}R_{ik}R_{jk})^{\beta}, \tag{5.12}$$

$$ECF(4,\beta) = \sum_{i<j<k<l\in J} p_{T,i}p_{T,j}p_{T,k}p_{T,l}(R_{ij}R_{ik}R_{il}R_{jk}R_{jl}R_{kl})^{\beta}, \tag{5.13}$$

where the sums runs over the jet constituents. 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 $ECF(N+1) << ECF(N)$. It follows that the ratio $r_N = ECF(N+1)/ECF(N)$ behaves very similarly to the N-subjettiness τ_N, while energy correlation double ratios $\frac{r_N}{r_{N-1}}$ behave like N-subjettiness ratios.

Sometimes one refers to the normalized n-point energy correlation functions, defined as,

$$e_n^\beta = \frac{ECF(n,\beta)}{(ECF(i,\beta))^n}. \tag{5.14}$$

Two important dimensionless ratios of ECFs are,

$$C_2^\beta = \frac{ECF(3,\beta)ECF(1,\beta)}{ECF(2,\beta)^2} = \frac{e_3^\beta}{(e_2^\beta)^2}, \tag{5.15}$$

$$D_2^\beta = \frac{ECF(3,\beta)ECF(1,\beta)^3}{ECF(2,\beta)^3} = \frac{e_3^\beta}{(e_2^\beta)^3}. \tag{5.16}$$

proposed in Refs. [135] and [136], respectively. 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 having predominantly low values of C_2 and D_2, and the latter larger values.

Grooming algorithms

Grooming techniques aim at cleaning the jet of soft and wide angle radiation in order to bring out the hard radiation pattern inside the jet. The overall effect is that of reducing the sensitivity to both pileup and the details of fragmentation. In the context of jet tagging, grooming has proven a useful tool to identify jet substructure and it is often used in combination, or as input, to jet tagging algorithms. In practice, 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 jet and is optimized in order to improve the resolution of the jet kinematics and properties. Several grooming algorithms are used in ATLAS, including trimming, pruning, split-filtering, and soft-drop [137]. 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. The soft-drop algorithm is relevant for this thesis and is the one described here.

In the first step, the anti-k_t $R = 1.0$ jets are reclustered with the C/A algorithm, which produces a pair-wise clustering tree ordered by increasing angular separation. The 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 following condition is evaluated:

$$\frac{\min(p_{T,i}, p_{T,j})}{p_{T,i} + p_{T,j}} > z_{\text{cut}} \left(\frac{\Delta R_{ij}}{R_0} \right)^{\beta}, \tag{5.17}$$

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 $i + j$ jet is taken as the final result, otherwise only the more massive subjet is kept and the procedure is repeated. The soft threshold z_{cut} determines the choice of keeping the hard structure and exclude soft emissions, starting from wide angle radiation. The angular exponent β weights the soft threshold according to the angular separation between the two subjets and determines how aggressive the grooming is. In ATLAS the recommendation is $\beta = 1.0$ and $z_{\text{cut}} = 0.1$.

5.5 Pileup suppression

Pileup refers to the simultaneous *pp* collisions that occur per bunch-crossing (see Sec. 4.1.3). The average pileup multiplicity was already $< N >= 20$ in Run 1, $< N >= 50$ at the end of Run 2, and is expected to go up to $< N >= 200$ 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 or thousands of soft particles on top of the decay products of a hard collision of interest. The

pileup transverse momentum density is generally uniformly distributed across the detector and varies event by event, and can significantly impact the reconstruction of final state objects. Mitigating the effect of this extra radiation is one of the main challenges for trigger and data analysis at the LHC. The material in this section comes mostly from Ref. [138].

In the context of jets, pileup contamination has two main consequences: an upward shift 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 ρ 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 across the detector volume [8]. The fluctuations are a form of noise that blurs the reconstructed quantities reducing their resolution.

One third effect not accounted here is the impact of the particles originating from pileup in the jet clustering procedure itself, as jets built with and without pileup will look slightly different due to different clustering histories. However, this effect produces a negligible average bias to the reconstructed $p_T$and its smearing effects scale as $\sqrt{\rho p_T}$ making them relevant only at large p_T[138].

All pileup mitigation techniques aim at reducing these effects, but different approaches exist 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 in recent years that combine previous techniques and extend them to constituent-by-constituent noise-suppression

[8]Pileup radiation is known to be non uniform across rapidity and aximuthal angle. For example, for a minimum bias collision the pileup energy deposits are significantly larger in the central rapidity region, than in the forward region [138].

methods. 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 separated the $y - \phi$ plane of an event in patches, 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 ρ, the pileup transverse momentum density in the event. 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 jet[9]. Computing ρ event-by-event and A jet-by-jet provides jets with better resolution compared to other methods that subtract average pileup contamination per-vertex or per-event, as averaging usually introduces extra resolution degradation. This correction leaves on average an unbiased transverse momentum of the jet and a small p_T degradation due to pileup fluctuations σ.

The area-median subtraction approach has proven to be a robust method, that leaves

[9]The *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 rapidity and azimuth. 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 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 $A(J) = \frac{\mathcal{N}_g(J)}{\nu_g}$. [138]

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 to not 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 double-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 and substructure information can be a useful tool. In this context, grooming techniques become useful.

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. The final result is essentially unbiased from pileup and the smearing effect is of $\sigma\sqrt{A_{\text{jet}}}$. 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. A groomer would therefore find the three hard prongs of the top decay and clean the jet of the noise. This produces a more 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 removed or with their energy adjusted, thereby adjusting the jet four momentum and substructure simultaneously. Of these methods, the most used algorithms at the moment are Soft-Killer (SK) [139] in ATLAS and PUPPI [140] in CMS. In the following, we will briefly describe the SK algorithm. These methods are independent of the jet clustering algorithm.

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 4.4.5, 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 first divided into patches of area A_i, the transverse-momentum flow density $\frac{p_{T,i}}{A_i}$ is

calculated for each patch i, and ρ is set as the median of all the patches:

$$\rho = \underset{i \in \text{patches}}{\text{median}} \left\{ \frac{p_{T,i}}{A_i} \right\}. \tag{5.18}$$

In the case of SK, the patches are taken to be squares of constant size $a \times a$, which significantly reduces computation time, without reducing effectiveness. 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 very fast to compute, as this is equivalent to taking the median of the p_T's of the leading particles in each patch:

$$p_T^{\text{cut}} = \underset{i \in \text{patches}}{\text{median}} \left\{ p_{T,i}^{\text{max}} \right\}. \tag{5.19}$$

The value of p_T^{cut} computed by the SK algorithm was computed in simulation as a function of $< N_{PV} >$ in Ref. [139] and was shown to barely exceed 2 GeV even at the HL-LHC $< N_{PV} >= 200$ pileup conditions.

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. (5.18) is then applied on the subtracted towers.

Voronoi and Constituent Subtraction

Voronoi (Vor) [141] and Constituent Subtraction (CS) [142] 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 then receives a pileup-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 that we saw above. After the energy has been corrected, different techniques can be used to remove pileup-like topoclusters, with the simplest one being removing any topocluster with $p_T^{\text{corr}} < 0$ GeV.

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 the particle p_T is applied:

$$\text{If } p_{T,i} \geq p_{T,k}^g\text{:} \quad p_{T,i} \to p_{T,i} - p_{T,k}^g, \tag{5.20}$$

$$p_{T,k}^g \to 0\,\text{GeV} \tag{5.21}$$

$$\text{otherwise:} \quad p_{T,k}^g \to p_{T,k}^g, -p_{T,i}, \tag{5.22}$$

$$p_{T,i}^g \to 0\,\text{GeV}. \tag{5.23}$$

Fixing the maximum ΔR values to be considered allows to tune the maximum jet area. This differs from the Voronoi area that only depends on the neighboring topoclusters. The original method similarly provides also a mass correction, but this is not applicable when using topoclusters, as they are massless. Both Voronoi and the CS method can be combined with SK to provide $p_T$corrected topoclusters before running the SK algorithm.

Chapter 6

Concepts of statistics and machine learning

6.1 Statistical inference

Consider a set of independent and identically distributed (i.i.d.) data $X = (x_1, ..., x_n)$ that is assumed to be sampled from a probability density function (PDF) $p(x|\theta)$, dependent on the 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. $\mathcal{A}$ represents a possible outcome of a measurement assumed to be repeatable and, given N measurements, the probability of $\mathcal{A}$ is given by

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

From the 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 Estimation (MLE). One first builds the likelihood function

$$L(\theta) = \prod_{i=1}^{N} p(x_i; \theta) \tag{6.1}$$

which attains higher values for the choices θ that are closer to the true distribution $p(x_i; \theta^{\text{true}})$, and vice versa. In this case, the true parameter θ is assumed to be fixed, but unknown. The best estimate of θ is then found by maximizing the likelihood or, as it is often the case, by minimizing the negative log-likelihood (NLL)

$$\hat{\theta}_{ML} = \arg\max_{\theta} \, p(X|\theta) = \arg\max_{\theta} \, \prod_{i=1}^{n} p(x_i|\theta) = \arg\max \sum_{i=1}^{n} \log \; p(x_i|\theta). \tag{6.2}$$

The parameter estimate $\hat{\theta}_{ML}$ is a function of the data and therefore is also a random variable. In other words, one can define a bias and variance for the estimator, representing how the estimator $\hat{\theta}$ 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, however, 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 ML estimator prone to overfit and be onverconfident in its prediction 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,

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

where the denominator can generally be regarded as simply a normalization factor. In this case, θ is itself a random variable and the posterior distribution $p(\theta|X)$ naturally expresses both aleatoric and epistemic uncertainty over θ. For a point estimate, one can take the maximum of the posterior (MAP)

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

The objective function retains the log-likelihood term from MLE, 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 $\hat{\theta}_{MAP} \to \hat{\theta}_{ML}$.

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 $\mathcal{A}$ provides an approximation $\hat{f} = \mathcal{A}(D)$. The quality of the prediction $\hat{y} = \hat{f}(\mathbf{x})$ is measured by a loss function $J = J(y, \hat{f}(\mathbf{x}))$ and the

learning – or training – is formulated as an optimization problem over the model parameters to minimize the loss function.

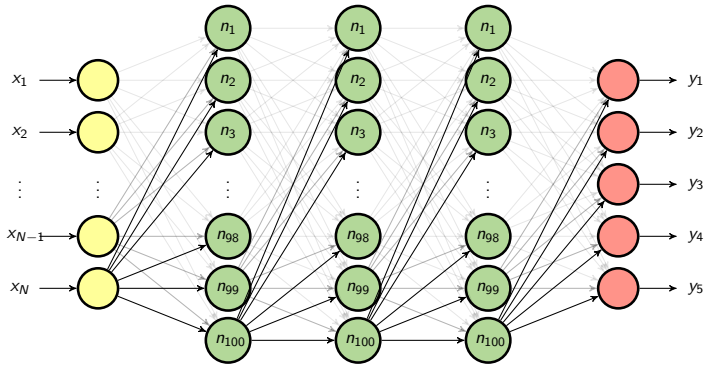


Figure 6.1: Architecture of a deep neural network with N input features, 3 hidden layers of 100 neurons each, and 5 output nodes.

A simple feed forward neural network [143] 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 trasformation $A(\mathbf{x}) = W^T\mathbf{x} + \mathbf{b}$ followed by a nonlinear transformation by a monotonic activation function $h(\mathbf{x}) = s(A(\mathbf{x}))$. This operaton is then cascaded over all the L hidden layers $\hat{f}(\mathbf{x}) = h_L(...(h_1(\mathbf{x})))$. The architecture – the choice of the hyper-parameters – is fixed before training, at which point the neural network output depends only on the weights $\mathbf{w}$, $\hat{f} = NN_{\mathbf{w}}$[1]. The training then consists in finding the optimal weights.

Most neural networks are trained using maximum likelihood. In order to do this, one has to specify a probability model $p(y|\mathbf{x}, \mathbf{w})$. Once the probabilistic model $p(y|\mathbf{x}, \mathbf{w})$ is chosen, the cost function is set to be the NLL $-\log p(\mathbf{y}; \mathbf{x}, \mathbf{w})$ and the optimization procedure consists in finding the parameters $\mathbf{w}$ that minimize this loss

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

[1]The term weights will include both weights and biases from now on

Regularization is often implemented by adding a prior over the weights and finding the maximum a posteriori (MAP).

$$\begin{aligned}\mathbf{w}_{MAP} &= \arg\max_{\mathbf{w}} \log p(\mathbf{w}|D) \\ &= \arg\max_{\mathbf{w}} \log p(D|\mathbf{w}) + \log p(\mathbf{w}).\end{aligned}$$

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

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

which corresponds to a Gaussian of mean $\hat{y} = NN_{\mathbf{w}}(\mathbf{x})$ and

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

Defining the cost function as the NLL, the optimal parameters are found by minimizing the loss, which, removing the terms not dependend on $\mathbf{w}$, results in,

$$\arg\min_{\mathbf{w}} - \log p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \beta) = -\frac{\beta}{2} \sum_{n=1}^{N} ||y_n - \hat{y}_n)||^2 \tag{6.5}$$

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 $\mathcal{N}(\mathbf{w}; 0, \frac{1}{\lambda} I^2)$, then the log-prior term in the cost function corresponds to the weight decay penalty $\lambda \mathbf{w}^T \mathbf{w}$. Similarly, it can be shown that for a classification task, where Y is a set of classes and $p(y|\mathbf{x}, \mathbf{w})$ is a

categorical distribution, minimizing the NLL results in the cross-entropy loss function.

In practice, the training is divided into two steps. During the *forward pass*, the network reads the input $\mathbf{x}$, produces the output $\hat{y}(\mathbf{x})$ 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. To compute the gradient the *backdrop* algorithm is generally used. Stochastic gradient descent is then implemented to update the model parameters in the direction of decreasing loss. The estimate of the gradient is computed on a mini-batch of m examples as

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

Here L is the per-example loss. The update then follows the estimated gradient downhill

$$\mathbf{w}^{new} = \mathbf{w} - \alpha \mathbf{g}$$

where α is the learning rate, a hyper-parameter 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 before. In order to make conclusions regarding an excess over 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

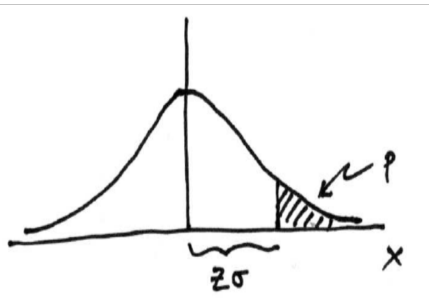
The concepts of significance and p-value are related and are often used in evaluating how well a given hypothesis describes the data. Suppose the background only null-hypothesis H_0 and the new-physics-hypothesis H_1 predict two different PDFs $f(\vec{x}|H_0)$ and $f(\vec{x}|H_1)$ for a set of observations $\vec{x} = (x_1 1, \ldots, x_N)$. 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 propability, 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 ther 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 Z[2] 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:

[2]The significance Z in particle physics is also often called S, which is the terminology used in the rest of this work.

$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - \Phi(Z),$$
$$Z = \Phi^{-1}(1-p) \tag{6.6}$$



where $\Phi(Z)$ is the inverse of the cumulative distribution (quantile) 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 $Z = 3$, and to report a "discovery" is $p < 2.8 \times 10^{-7}$, or a significance of $Z = 5$. On the other hand, to exclude a signal hypothesis a p-value of 0.05 (i.e. 95% confidence level) is generally used, which corresponds to a significance of $Z = 1.64$.

Likelihood parameters

The data is usually a set of independent and identically (i.i.d.) distributed measurements $\boldsymbol{x}$ and the hypothesis is expressed as a PDF, with each hypothesis predicting a different PDF $f(\boldsymbol{x})$. Often a continuous set of hypotheses is considered $f(\boldsymbol{x}; \mu)$, 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 $\sigma_s = \mu \sigma_s^{\mathrm{MC}}$. 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 $L(f(\boldsymbol{x}; \mu))$, giving the probability of the data given the hypothesis $f(\boldsymbol{x}; \mu)$. 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, which parametrize the effect of the uncertainty on the relation between $\boldsymbol{x}$ and μ. 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 sample and a variable of interest x, e.g. the invariant mass distribution of the reconstructed signal resonance. The signal and background samples will have different PDFs of the variable x, $f_s(x;\boldsymbol{\theta})$ and $f_s(x;\boldsymbol{\theta})$.

If one constructs a histogram $\boldsymbol{n} = (n_1, n_2, \cdots, n_N)$ with N bins of the variable x, the expectation value of the number of events in a given bin i is given by $E[n_i] = \mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})$, where $b_i = b_{tot} \int_{\text{bin } i} f_b(x;\boldsymbol{\theta_b})$, 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 $L(\mu, \boldsymbol{\theta})$ can then be expressed as,

$$L(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{N} \text{Pois}\left(n_i | \mu s_i\left(\boldsymbol{\theta}\right) + b_i(\boldsymbol{\theta})\right) \prod_{\boldsymbol{\theta}_k}^{\boldsymbol{\theta}} \mathcal{N}(\boldsymbol{\theta}_k | \boldsymbol{\theta}_k^0, \sigma_k). \tag{6.7}$$

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

Profile likelihood ratio test

In order to test a hypothised value of μ one needs a test statistics q_μ [144]. This is generally obtained from the profile likelihood ratio $\lambda(\mu)$, given by

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} \tag{6.8}$$

The numerator is the *profile likelihood (ML) function.* The quantity $\hat{\hat{\boldsymbol{\theta}}}$ represents the ML estimate of $\boldsymbol{\theta}$ conditional on the specified value of μ, and thus depends on μ. The denominator is the maximized unconditional likelihood function, where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are set to their ML estimators. It represents the global maximum, so that the ratio is always ≤ 1, with λ closer to 1 implying a better agreement between data and the hypothesed value of μ. Generally one assumes that the presence of a signal could only increase the observed number of events, so one defines

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}, & \hat{\mu} \geq 0, \\ \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(0, \hat{\hat{\boldsymbol{\theta}}}(0))}, & \hat{\mu} < 0, \end{cases} \tag{6.9}$$

where it is assumed that the best level of agreement between a value of μ and data for an observed $\hat{\mu} = 0$ 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 aproximation, the

PDFs assume the shape of χ^2_1 functions, such that the p-value can be expressed in terms of the cumulative distribution of a standard Gaussian as

$$p_\mu = \int_{q_\mu,obs}^{\infty} f(q_\mu|\mu)dq_\mu = 1 - F(q_\mu|\mu) = 1 - \Phi(\sqrt{q_\mu}). \tag{6.10}$$

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

$$p_{\mu,95\%} = 1 - \Phi(\sqrt{q_\mu}) = 1 - \Phi(1.64) = 0.05. \tag{6.11}$$

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

$$q_{\mu=0} = \begin{cases} -2\ln\lambda(0), \hat{\mu} & \geq \mu, \\ 0, \hat{\mu} & < \mu \end{cases} \tag{6.12}$$

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 then be calculated as

$$p_0 = \int_{q_0,obs}^{\infty} f(q_0|0)dq_0. \tag{6.13}$$

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

$$q_{\mu=0} = \begin{cases} -2\ln\lambda(\mu), & \hat{\mu} \leq \mu, \\ 0, & \hat{\mu} > \mu, \end{cases} \tag{6.14}$$

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

$$p_\mu = \int_{q_\mu,obs}^{\infty} f(q_{mu}|\mu)dq_\mu. \tag{6.15}$$

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

$$\mu_{up} = \hat{\mu} + \sigma\Phi^{-1}(1-\alpha), \tag{6.16}$$

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 their four-momenta from the four-vectors of their decay products and they should appear as a narrow resonance in 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 searches.

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 [42, 43] and little Higgs [41] 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. 3.0.3). A second class of models, the two-Higgs-doublet models (2HDMs) [4], predict the simplest extension of the SM scalar sector, by including an additional scalar SU(2) doublet, resulting in five physical scalars (see Sec. 3.0.3). In addition, Randall-Sundrum (RS) models with warped extra dimensions [145] or by 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 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 $3.2\,\text{fb}^{-1}$ [146], $36.1\,\text{fb}^{-1}$ [147], and $139\,\text{fb}^{-1}$ [148] dataset. The VV searches were performed using the

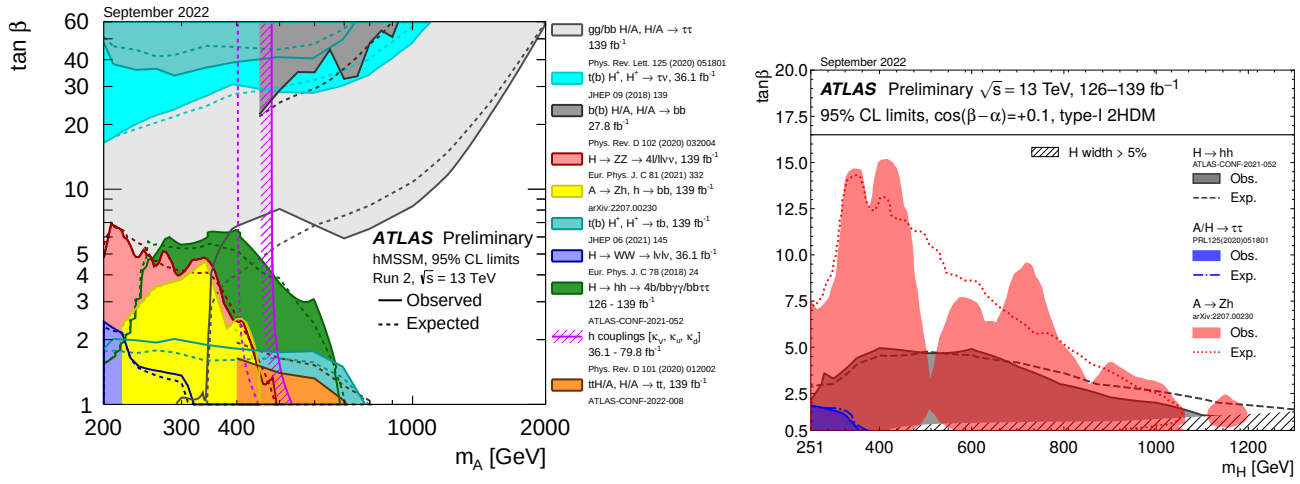
$36.1\,\text{fb}^{-1}$ [149, 150] and the $139\,\text{fb}^{-1}$ [151] dataset. Similar searches in semi-leptonic final states have been performed in CMS as well, with the latest analyses using the $137\,\text{fb}^{-1}$ dataset being a search for a new resonance decaying to $WZ/WW/WH$ [152] and to ZH final state [153]. ATLAS has also performed searches for the same process in different final states, including two fully hadronic searches based on an integrated luminosity of $139\,\text{fb}^{-1}$ in the VH [154] and VV [155] final state. CMS has performed a fully-hadronic VH search [156] and a ZH search in final states with two taus and two light leptons [157] with the $35\,\text{fb}^{-1}$ dataset.

Statistical combinations of the available searches for different processes and in the different final states have also been performed with $36.1\,\text{fb}^{-1}$ analyses [158] and the on-going effort with the $136\,\text{fb}^{-1}$ searches [159], 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 [160] and with the $35.9\,\text{fb}^{-1}$ dataset [161].

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 $36.1\,\text{fb}^{-1}$ publication with a local (global)

significance of 3.6 (2.4) standard deviations and a disagreement around a mass of 400 GeV was also observed in the $A \to t\bar{t}$ CMS search with 35.9 fb^{-1} [162] with a local (global) significance of 3.5 (1.9) and a local excess above 2σ is observed in the $A \to \tau\tau$ ATLAS search with 139 fb^{-1}. It should be noted that the excess is around the threshold of 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 bosons 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.



(a) (b)

Figure 7.1: Left: Regions of the $[m_A, tan\beta]$ phase space excluded for a type of MSSM via direct searches and 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 $[m_H, tan\beta]$ (in this benchmark scenario $m_H = m_A = m_{H\pm}$) phase space excluded for Type I 2HDM by direct searches for $cos(\beta - \alpha) = 0.1$, comparing observed (filled) and expected (lines) limits. All limits are quoted at 95% confidence level. [163]

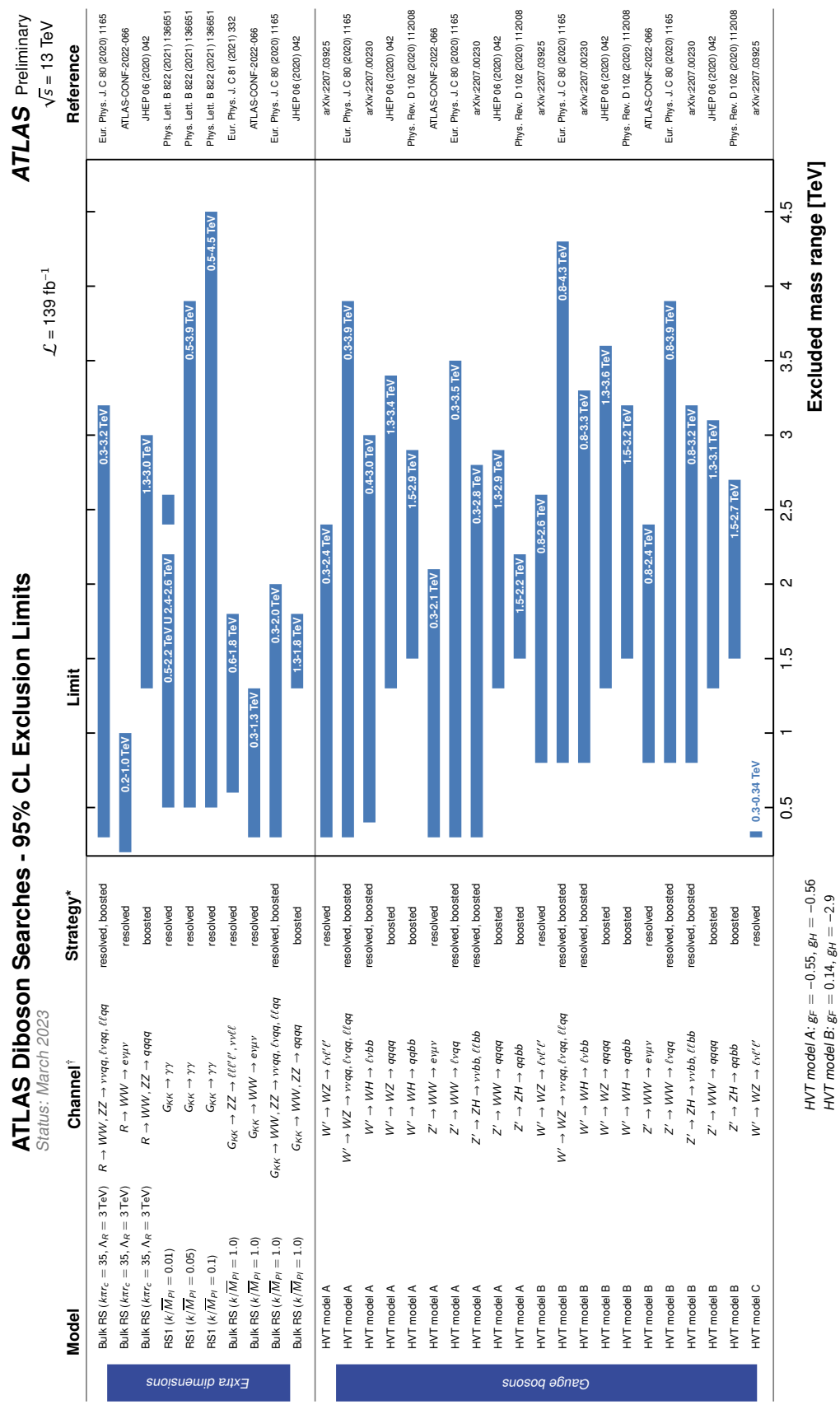


Figure 7.2: Summary of mass exclusion limits at 95% confidence level from ATLAS diboson searches with the full Run dataset. [**?**].

7.2 Analysis overview

In this work the on-going search for a new heavy resonance decaying through VV and VH processes in the semi-leptonic final state is presented. The leptonic decay of one vector boson proceeds as $Z \to \nu\bar{\nu}$, $W \to l^{\pm}\overset{(-)}{\nu}$, or $Z \to l^{+}l^{-}$, where l refers to a light charged lepton (electron or muon). The hadronic decay of vector bosons proceeds as $W \to qq$ and $Z \to q\bar{q}$, while the analysis targets only Higgs boson decays to a pair of b-quarks in order to capitalize on the large BR of the $H \to b\bar{b}$ decay channel ($\sim 57\%$)[1].

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 $A \to ZH$. 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 $139\,\text{fb}^{-1}$ dataset collected up until the end of Run 2. This dataset has already been analyzed in the latest VV and VH publications, but a new revision was motivated by several developments, including improved b-tagging algorithm, improved jet collections, and a new optimized event selection, including 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,

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

[1] From now on, the references to particle/anti-particle state will be omitted, e.g. we will only refer to the W leptonic decay as $W \to l\nu$.

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 significance 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 physics[2]. 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 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 potential 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 of the invariant mass distribution in all the signal regions (SR) and the background-dominated control regions (CRs). This is done under the background-only and signal-plus-background hypothesis, to

[2]The 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.

quantify the probability of potential excesses to be due to the presence of new physics phenomena.

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 then 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* (VR) 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 are also sometimes 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 studies.

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 non-linear decision boundaries. Consider a neural network (NN) with one output node providing a probability for a classification task. This is a new one-dimensional test statistic $t(x)$, similar to the standard one-dimensional cuts. Just like the parameters of the hard cuts, the weights of the NN are optimized so that $t(x)$ gives the maximum separation between the PDF of $t(x)$ for the signal and background hypotheses. However, $t(x)$ is much more refined test statistic, as it encapsulates information about the N dimensional input variables and in the feature space the cut on $t(x)$ is in fact a non-linear decision boundary. The ability to map a high dimensional input space to few dimensions is in fact one of the advantages of the NN. In addition, a NN can learn any type of correlation between the input variables, so that the decision boundary provides both the optimal signal efficiency *and* signal purity. This motivated a more global machine learning approach in the analysis, with 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 repur-

posed from previous publications. The vector-boson fusion (VBF) and gluon-gluon fusion (ggF)/Drell-Yan (DY) production modes are characterized by different 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 with large dijet invariant mass. In the previous VV search [151], the VBF and ggF/DY final regions were made orthogonal via a Recursive Neural Network (RNN) [164], 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.

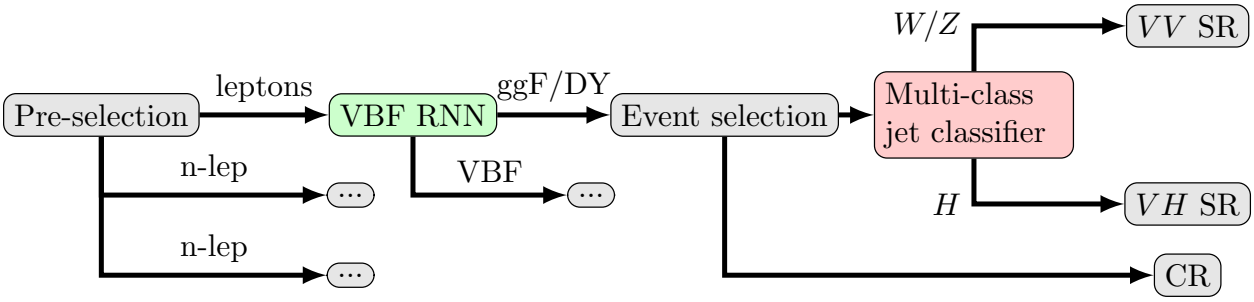


Figure 7.3: Analysis workflow displaying where the VBF-RNN and the MCT are applied.

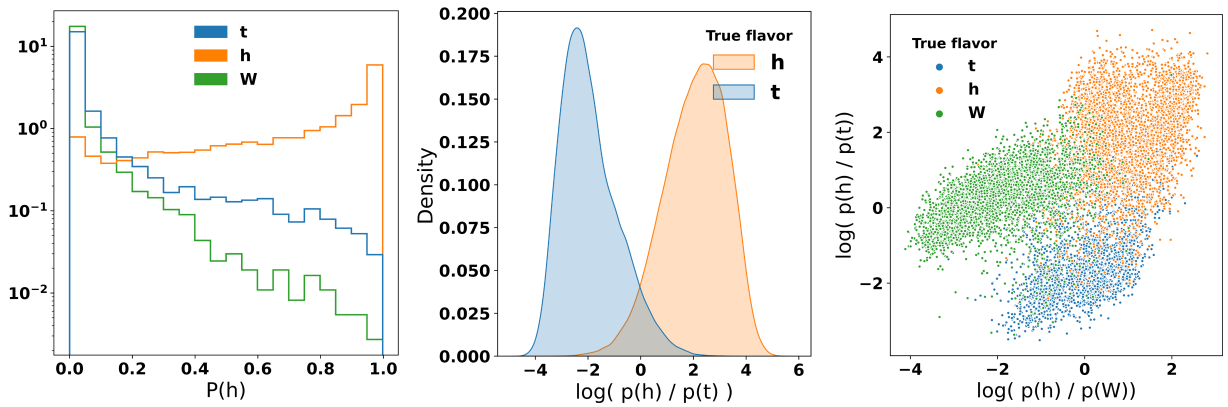
7.2.3 The multi-class tagger

This thesis presents the development of a new five-class 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 (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 or the W/Z tagger [165], 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, which opens the possibility for simultaneous scoring, i.e. the ability to ask the question "is the jet more top-like or more V-like?." This question is answered by likelihood ratios between different class probabilities, 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 flavour tagging (see Sec. 4.4.6), but it had not been explored at the beginning of this work by ATLAS in boosted jet tagging. A multi-class approach has already been used by CMS for boosted jet tagging [166].

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



(a) (b) (c)

Figure 7.4: Examples of multi-class classification using likelihood ratios.

requires thhe orthogonalization of the final regions. Since the VV and VH processes differ only by the hadronic decay part, the multi-class tagger provides an optimal way to solve this issue, considering also that no Higgs vs. vector boson classifier existed in ATLAS at this time.

The next sections will discuss the development of the MCT for large-R jet classification, its extension to the resolved jets topology, and its use 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.

7.3 Signal and background processes

The search targets different signal interpretations, with different production modes. Both DY (or quark-antiquark annihilation) and VBF processes are considered in the search for new HVT bosons. In the search for a pseudo-scalar A, ggF and b-quark associated produc-

tion (bb) 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.

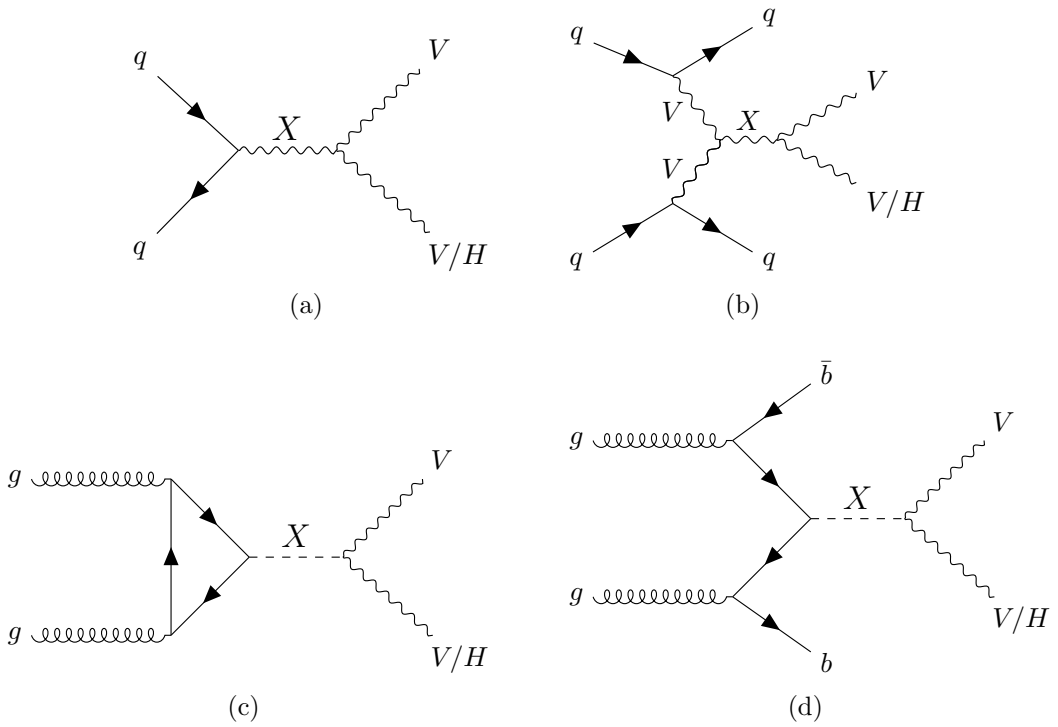


Figure 7.5: Representative Feynman diagrams for Drell-Yan, vector-boson fusion, gluon-gluon fusion, and b-quarks associated production of a new heavy resonance X. When multiple options are possible for quarks flavour and for vector boson charge, these are left unspecified.

In the HVT (see Sec. 3.0.3) signal interpretation as a new electrically charged W' or electrically neutral Z' vector boson, the possible decay modes are $W' \to WZ$, $W' \to WH$, $Z' \to WW$, and $Z' \to ZH$. 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 $g_H = g_V c_H$, 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 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 new heavy resonances, which can be used to constraint the coupling g_F and g_H. The search for a new HVT boson is performed in the mass range between 300 GeV and 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. 3.0.4). The resonance can decay to a ZH final state. The search is performed in the mass range between 220 GeV to 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) predicted by certain Randall-Sundrum models [167, 168]. This corresponds to a neutral scalar resonance. The other is a neutral spin-2 graviton (G_{KK}) [145, 169] of the first Kaluza-Klein (KK) excitation in a bulk Randall-Sundrum model. Both are predicted to have dominant BRs to WW or ZZ final states and can be produced via ggF or VBF processes. The 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

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

Table 7.1: Channels used in the search for W' and Z' HVT bosons, pseudoscalar A, and Radion/Graviton.

given in Tab. 7.1.

Several SM background processes can have similar final states as the signals: W and Z boson production in association with jets (referred to as W+jets and Z+jets, or collectively as V+jets); top quark production, with the main contribution coming from top-quark pair production, denoted as $t\bar{t}$, but including also single-top-quark production; non-resonant di-boson production (WW, WZ, or ZZ) with semi-leptonic decays; and multiijet 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 $\sqrt{s} = 13\,\text{TeV}$ and are passed through the full GEANT4-based [116] ATLAS detector simulation. All generated samples include the simulation of in-time and out-of-time pileup by overlaying 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. Simulated events are corrected for differences with data in the energy scale and efficiency of reconstruction, identification and triggering of leptons and jets, as well as in tagging efficiencies of heavy-flavor jets. This correction is applied as a multiplicative factor to the event weight.

The QCD-multijet 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 anlaysis is able to select a phase space with negligible multijet contamination.

The HVT Z' and W' production via quark-antiquark annihilation was modelled at LO accuracy in QCD with `MadGraph5 2.3.3` (MG5) [170] generator, using the NNPDF2.3 LO PDF set [171], interfaced with Pythia 8.186 [172] for modeling of the parton shower with the A14 set of tuned parameters [173]. 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 $\mathcal{B}(h \to b\bar{b}) = 0.569$, $\mathcal{B}(h \to \bar{c}) = 0.0287$, and $m_H = 125\,\text{GeV}$ were assumed. Another set of samples is generated for model C for VBF production only.

The 2HDM ggA signal samples were generated with the narrow width approximation using the `MadGraph5 2.3.3` generator at LO accuracy in QCD, with the `2HDM_GF FeynRules` model [174] and the NNPDF2.3 LO PDF set. The 2HDM bbA process was generated using the four-flavour scheme at next-to-leading order (NLO) with massive b-quarks with

MadGraph5_aMC@NLO2.2.3 and the NNPDF2.3 NLO PDF set. Shower modeling was performed with Pythia 8.186 with A14 tuning. Resonance masses in the range between 220 GeV and 2 TeV were simulated for each signal process. The signal samples were generated only for Higgs decays to a pair of b-quarks and leptonic decays of the Z boson.

Signal samples for the RS graviton were produced with MadGraph5-2.2.2 (MG5) interfaced to Pythia 8.186 using the NNPDF2.3 LO PDF. The radion samples are generated with MadGraph5-2.6.0 (MG5) interfaced to Pythia 8.212. For each scenario, different samples are produced for masses ranging from 300 GeV to 6 TeV.

A summary of the MC generators used to produce the various background processes is given in Tab. 7.2.

Process	Generator	σ_{prod} calc. order in QCD
	Vector boson + jets	
$W \to l\nu$	Sherpa2.2.1	NNLO
$Z \to ll$	Sherpa2.2.1	NNLO
$Z \to \nu\nu$	Sherpa2.2.1	NNLO
	Top quark	
$t\bar{t}$	Powheg+Pythia8	NNLO+NNLL
single top (s-channel)	Powheg+Pythia8	NLO
single top (t-channel)	Powheg+Pythia8	NLO
Wt-channel	Powheg+Pythia8	NLO
$t\bar{t}+h$	MadGraph5_aMC@NLO + Pythia8	NLO (QCD) and NLO (EW)
$t\bar{t}+V$	MadGraph5_aMC@NLO + Pythia8	NLO
	Diboson	
$qg/q\bar{q} \to WW \to \ell\nu qq$	Sherpa2.2.1	NLO
$qg/q\bar{q} \to WZ \to \ell\ell qq$	Sherpa2.2.1	NLO
$qg/q\bar{q} \to WZ \to \nu\nu qq$	Sherpa2.2.1	NLO
$qg/q\bar{q} \to WZ \to \ell\nu q\bar{q}$	Sherpa2.2.1	NLO
$qg/q\bar{q} \to ZZ \to \ell\ell q\bar{q}$	Sherpa2.2.1	NLO
$qg/q\bar{q} \to ZZ \to \nu\nu q\bar{q}$	Sherpa2.2.1	NLO
$gg \to WW \to \ell\nu qq$	Sherpa2.2.2	NLO
$gg \to ZZ \to \ell\ell q\bar{q}$	Sherpa2.2.2	NLO
$gg \to ZZ \to \nu\nu q\bar{q}$	Sherpa2.2.2	NLO
$qg/q\bar{q} \to \ell\ell\nu\nu$	Sherpa2.2.2	NLO
	SM Higgs	
$qq \to Wh \to \ell\nu bb$	Powheg+Pythia8	NNLO (QCD) and NLO (EW)
$qq \to Zh \to \nu\nu bb$	Powheg+Pythia8	NNLO (QCD) and NLO (EW)
$qq \to Zh \to \ell\ell bb$	Powheg+Pythia8	NNLO (QCD) and NLO (EW)
$gg \to Zh \to \nu\nu bb$	Powheg+Pythia8	NLO+NLL
$gg \to Zh \to \ell\ell bb$	Powheg+Pythia8	NLO+NLL

Table 7.2: Summary of the Monte-Carlo generators used to produce the various background processes. [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 $\sqrt{s} = 13\,\text{TeV}$. 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 unprescaled 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 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 [85], 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 are 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.

The full list of triggers used in each run is shown in Tab. 7.4. Possible differences between data and MC are taken into account by applying scale factors provided by the ATLAS Combined Performance groups.

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.

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,

Data-taking period	$e\nu qq$ 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) ($L < 1.0 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$)	HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e140_lhloose_nod0 HLT_e300_etcut	HLT_mu26_ivarmedium OR HLT_mu50	HLT_xe90_mht_L1XE50
2016b (run ≥ 302919) ($L < 1.7 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$)	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. [From analysis ATLAS internal note]

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 $p_T > 5\,\mathrm{GeV}$.

Electrons

Electrons are reconstructed from topological clusters in the electromagnetic (EM) calorimeter matched to tracks in the ID. All electron candidates are required to have $p_T > 7\,\mathrm{GeV}$ 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 in 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 $W \to e\nu$. This relies on tight identification and isolation criteria, and requires electrons with $p_T > 30\,\mathrm{GeV}$. The `Loose` criteria is used to select $Z \to e^+e^-$ electrons in the 2-lepton channel, with looser identification and isolation working points, and no isolation

requirement for electrons with $p_T > 100\,\text{GeV}$.

Muons

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

Small-R jet

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 $p_T > 20\,\text{GeV}$. Forward jets are reconstructed in the forward region ($2.5 < |\eta| < 4.5$) and are required to have $p_T > 30\,\text{GeV}$.

To reduce contamination from jets originating from pileup vertices, jets with $p_T < 60\,\text{GeV}$ and $|\eta| < 2.4$ are further required to pass a jet-vertex tagging [175] selection.

Large-R jet

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. 4.4.5) 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 obtain an event-dependent $p_T$cut to remove low $p_T$constituents. Further pileup suppression is obtained by applying the Soft-drop algorithm, with parameters $\beta = 1.0$ and $Z_{cut} = 0.1$, on the set of reconstructed jets, which removes constituents associated with soft and wide-angle radiation.

Variable radius track jet

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 $R_{max} = 0.4$ and $R_{min} = 0.02$. All VR jets are required to have $p_T > 10GeV$, $|\eta| < 2.5$, and number of associated tracks 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 overlapping track jets used for b-tagging ($p_T > 5GeV$ and nTrk > 1) and VR track jets selected by the analysis are removed, according to the recommendation of the ATLAS Combined Performance group.

Flavour-tagging

The DL1r flavour-tagging algorithm is used to tag signal jets and VR track jets. A cut on the DL1r score as defined in 4.10 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 85% WP cut is used on VR track jets. 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 [176] are reconstructed using $R = 0.4$ calorimeter jets. They are required to have one or three associated tracks, $p_T > 20\,\text{GeV}$ 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 $p_T > 50\,\text{GeV}$. 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. According to the number of reconstructed charged leptons, the events are separated into lepton channels targeting the corresponding leptonic decays: $Z \to \nu\nu$ in the 0-lepton channel, $W \to e\nu$ or $W \to \mu\nu$ in the 1-lepton channel, and $Z \to ee$ or $Z \to \mu\mu$ in the 2-lepton channel.

The next step is to reconstruct the hadronic candidate. According to the signal hypoth-

esis, 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 optimized 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 heavy resonance decay from the reconstructed hadronic and leptonic final states.

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

Merged regime

The $H/W/Z$ candidate is taken as the leading large-R jet in the event. Events are required to have the leading large-R jet with $p_T > 250\,\text{GeV}$ ($p_T > 200\,\text{GeV}$) 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 [165]. 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).

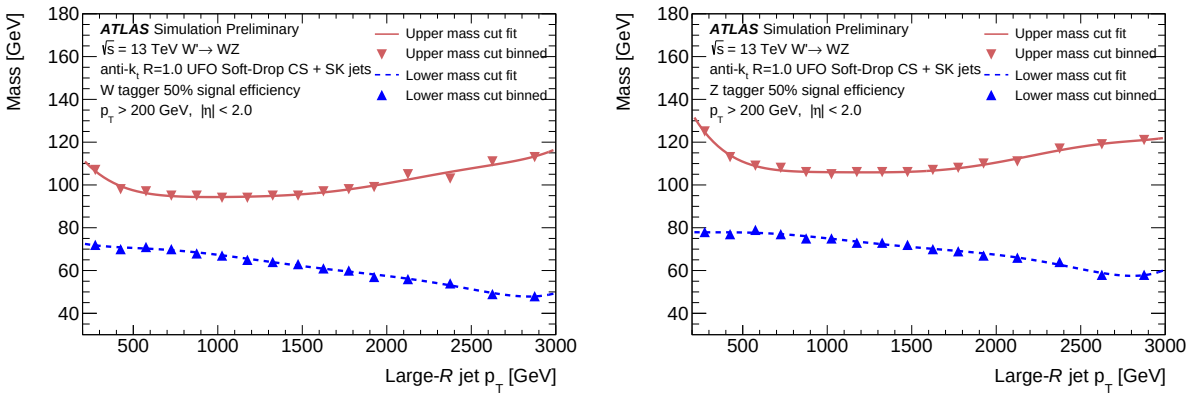


Figure 7.6: Upper and lower cut values on $m(J)$ for the cut-based W tagger (left) and Z tagger (right) in bins of large-R jet p_T [165].

7.6.2 0-lepton channel

The 0-lepton selection targets $Z \to \nu\nu$ 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. 4.4.4, is required to be $\mathcal{S} > 10$.
- The reconstructed $\mathbf{E}_T^{miss}$ is required to be isolated by requiring $\min[\Delta\phi(jet, \mathbf{E}_T^{miss})] > \frac{\pi}{9}$, where ϕ is the angle between $\mathbf{E}_T^{miss}$ and the nearest small-R jet.
- The track-based missing transverse momentum p_T^{miss} is required to be above 80 GeV.

The presence of non-zero E_T^{miss} does not allow the complete reconstruction of the $Z \to \nu\nu$ 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 $(E_T^{miss}, E_x^{miss}, E_y^{miss}, 0)$ with the four-vector representing the $H/W/Z$ candidate without the longitudinal components $(\sqrt{p_x^2 + p_y^2 + m_2}, p_x, p_y, 0)$, where p_x, p_y, m are the kinematics of the reconstructed $H/W/Z$ in the resolved or merged regime.

7.6.3 1-lepton channel

The 1-lepton channel targets $W \to l\nu$ 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 $E_T^{\text{miss}} > 60(100)\,\text{GeV}$ 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 relation $m_W^2 = (p_l + p_\nu)^2$. The result is given by,

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

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

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

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 $\Delta(l, j_1/j_2) > 1.0$ 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.

The hypothetical signal is reconstructed from the four-momentum sum of the reconstructed leptonically and hadronically decaying bosons. The final discriminant is given by the invariant mass of the reconstructed resonance. As an example, for the $X \to WH$ resonance system the invariant mass is calculated as,

$$m_{Wh} = \sqrt{(E_W + E_H)^2 - (\vec{p}_W + \vec{p}_H)^2}. \tag{7.4}$$

7.6.4 2-lepton channel

The 2-lepton channel selects events compatible with a $Z \to ll$ decay. Events are required to have exactly two isolated `Loose` leptons of the same flavour (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 $p_T > 27\,\text{GeV}$ and the sub-leading lepton is required to have $p_T > 20(25)\,\text{GeV}$ 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 di-lepton system is required to be consistent with the Z boson mass in order to suppress backgrounds without a resonant di-lepton pair. Electron pairs are required to have $m_{ee} \in [83, 99]\,\text{GeV}$, 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:

$$(85.6 - 0.0117 \cdot p_{T,ll}) < m_{\mu\mu} < (94.0 + 0.0185 \cdot p_{T,ll})\,\text{GeV} \tag{7.5}$$

The same p_T balance requirement as in Eq. (7.3) is used in the region $m(VV/VH) < 320\,\text{GeV}$ to further suppress background:

$$\frac{\min\left(p_T(Z_{lep}), p_T(W/Z/H_{had})\right)}{m(VV/VH)} > 0.35 \tag{7.6}$$

At higher signal masses the background contamination is sufficiently low and 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 di-lepton system and the reconstructed hadronic decay. The final discriminant is given by

the invariant mass of the reconstructed resonance. In the $X \to ZH$ case, this is given by,

$$m_{Zh} = \sqrt{(E_H + E_Z)^2 - (\vec{p}_H + \vec{p}_Z)^2}, \tag{7.7}$$

7.7 Final regions definition

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.

Production modes

In a given analysis, the searches targeting ggF and DY production processes have similar event topologies and 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 significant 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 of 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 $H \to b\bar{b}$ decay and to the number of b-tagged jets.

In the VV analysis, for regions targeting $Z \to qq$ 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 $Z \to b\bar{b}$ 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 $V \to J$ candidate, determined via a multivariate W/Z classifier [165]. In summary, for each lepton channel the VV final regions are classified according to: merged and resolved reconstruction of $V \to qq$ decay, high and low purity of merged V reconstruction, 0/1 or 2 b-tagged jets in $Z \to qq$ decays, mass windows of $Z \to qq$ or $W \to qq$ 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

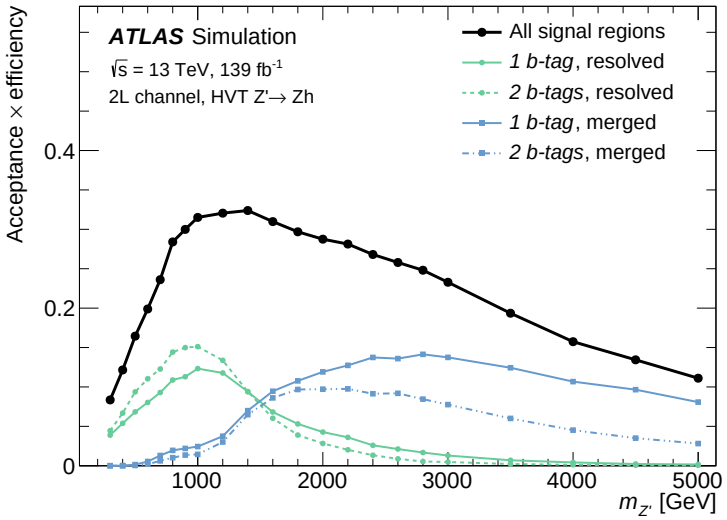


Figure 7.7: Acceptance × efficiency in the VH 2-lepton analysis for $Z' \to Zh$ signal as a function of the resonance mass in latest VH publication [148].

practice, each event is reconstructed with both strategies. According to the reconstruction strategy, different jet requirements are implemented. Most events are 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 an H boson, in the second. While the combination of the HVT signal regions was the motivation for using the MCT, it was found that applying it to all SRs and then including both VV and VH orthogonalized SR's in each fit, brought a higher sensitivity. For instance, in the search for a pseudoscalar A via ggF production, where historically only VH-like signal regions were

used, now the VV-like signal regions are included as well. The former represent "high purity" regions, while the latter recover part of the efficiency as the "low purity" region.

Control regions

Control regions (CRs) are used to constraint the normalization of the most dominant backgrounds in the final fit. Ideally, the CR should cover a phase space as close as possible to the signal region(s) while being orthogonal. All CRs are common to the VV and VH analysis.

The main control regions are obtained by inverting the mass window requirement of the given signal region. Events with mass values within $50\,\text{GeV} < m_{jj/J} < 200\,\text{GeV}$, but outside of VV and VH mass windows, are put in the *mass sideband* CRs. A special control regions 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 flavour requirement, requiring leptons of opposite flavor.

7.8 Merged MCT

The merged MCTis 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). In this section the training and testing performance of the MCTare presented, 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 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 and only tracks with $p_T > 10$ GeV and $|\eta| < 2.5$. The DL1r algorithm is used for each VR track jets 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, as well as the top quark sample, are obtained from simulations of heavy BSM resonances decays into boosted SM particles. Their truth labeling relies on truth matching [3] and further reconstruction quality criteria.

The QCD sample is obtained from multijet processes, where the jets are produced via

[3] All 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.

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. 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 $H \to b\bar{b}$ decays are generated with $G \to HH$ processes, where G is a Randall-Sundrum graviton. The events are simulated using PYTHIA8 with the ATLAS A14 tune and the `NNPDF2.3 LO PDF`. Only reconstructed Higgs candidates truth-matched to a true Higgs particle and with two ghost-associated b-hadrons are selected. No b-tagging requirement is required on the associated track jets.

A sample of hadronically decaying top quarks is obtained from simulated $Z' \to t\bar{t}$ decays, generated with `PYTHIA8` and `NNPDF2.3 LO 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 Split_{23}/1.e3 ¿ exp[3.3-6.98e-04*p_T/1.e3].

W and Z boson jets are obtained from $W' \to WZ$ decays, where only hadronically decaying W/Z are considered. The samples are generated with `PYTHIA8` and `NNPDF2.3 LO 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 the cut on the d_12 k$_t$-splitting scale:

$$\sqrt{d_{12}} > 55.25 \cdot \exp\left[\frac{-2.34 \cdot 10^{-3}}{GeV} p_T\right] \tag{7.8}$$

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-jets (See Sec. 5.3.2). The flavour 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. Kinematics and b-tagging scores are provided for up to three leading track jets. 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 high-level 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 the individual variables thanks to the ability of the network to learn correlations between the inputs. Classifiers using *high-*

level inputs have already been shown to bring an increased performance in the context of *b*-tagging, where the DL1r algorithm is 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 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 [177]. The original set was down-selected in order to remove significant redundant information by removing variables which did not affect the network performance. The final reduced list of inputs used for the training is shown in Tab. 7.5. Examples of input variables are shown in Fig. **??**.

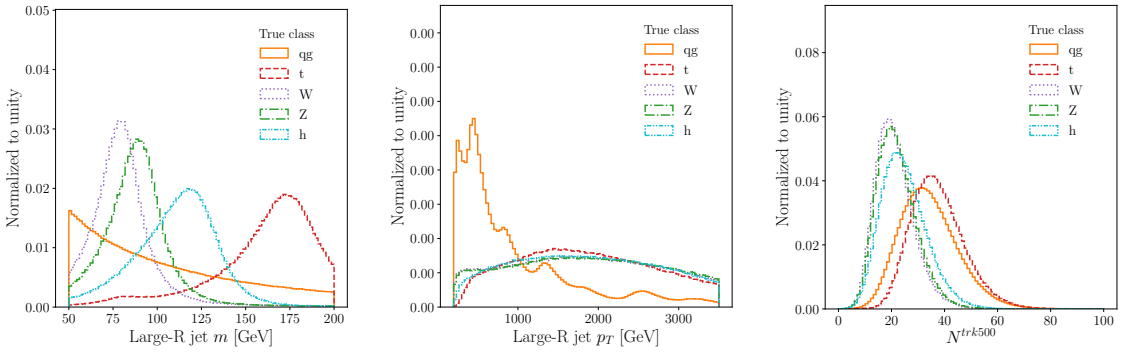
Type	Observable	Definition
Large radius jet	p_T	Transverse momentum
	m	Mass
	N^{Const}	Constituents multiplicity
	N^{trk500}	Tracks multiplicities $p_T > 500$ MeV
3 leading track jets	p_T	Transverse momentum
	DL1r pb	Bottom quark probability
	DL1r pc	Charm quark probability
	DL1r pu	Light quark probability
Substructure observables	τ_2	N-subjettiness
	τ_3	
	τ_{21}	
	τ_{32}	
	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

As expected, one of the most powerful observable is the jet mass, which for large-R jet originating from a heavy particle has a scale associated to the mass of the particle.

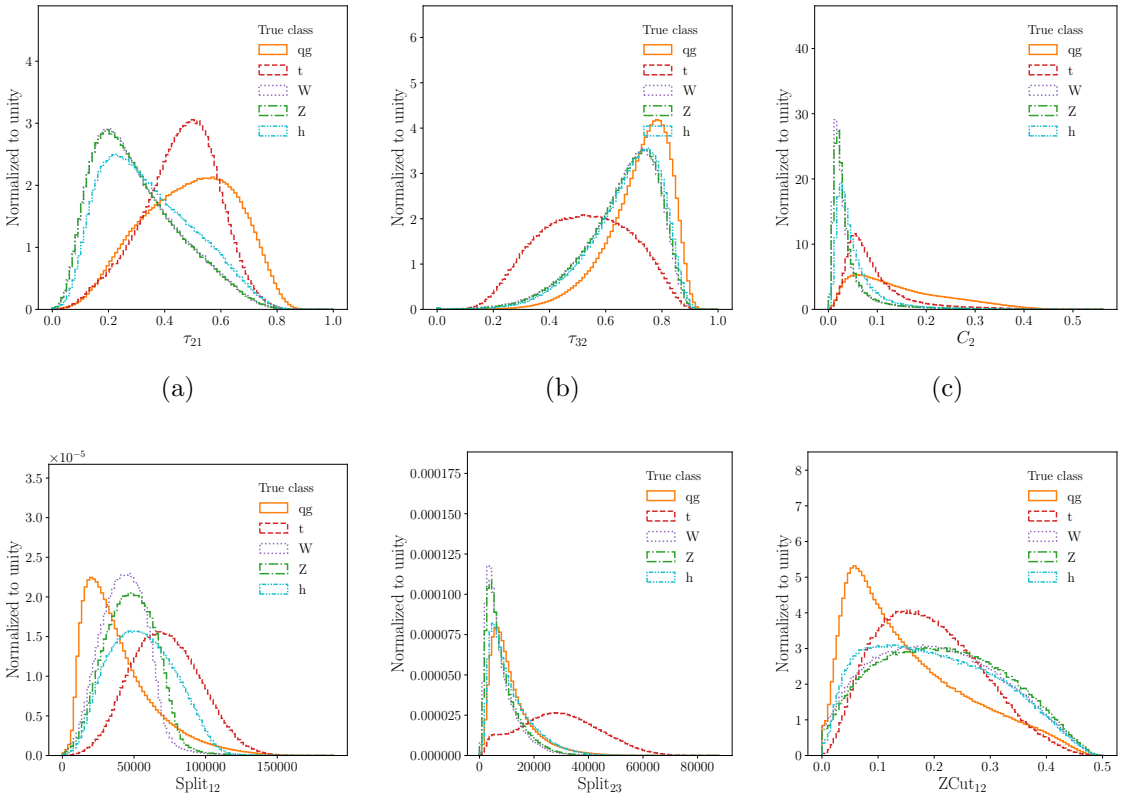
Pre-processing

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 variable x the mean μ and standard



(a) (b) (c)

Figure 7.8: Subset of input variables to the MergedMCT describing the large-R jet.



(a) (b) (c)

(d) (e) (f)

Figure 7.9: Examples of input variables to the MergedMCT describing the large-R jet substructure.

deviation σ were obtained only from the samples in the training dataset. Each event was

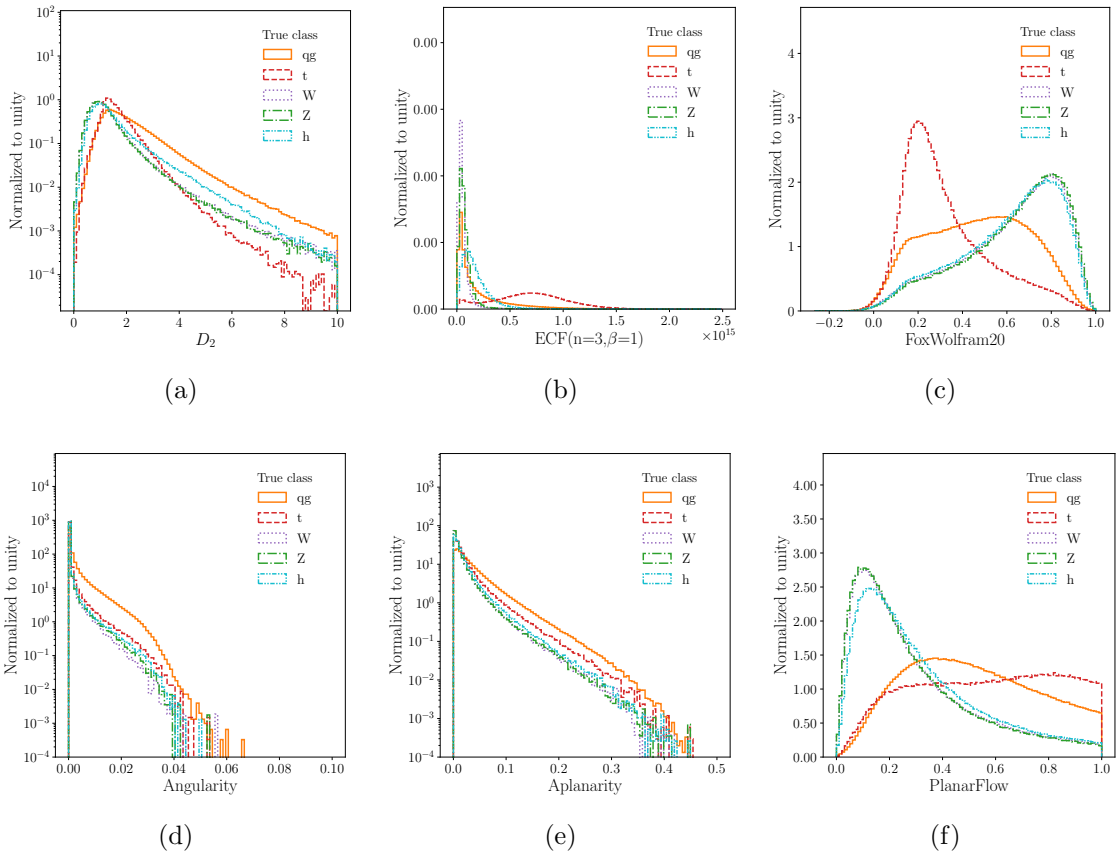


Figure 7.10: Examples of input variables to the MergedMCT describing the large-R jet substructure.

then standardized as

$$x' = (x - \mu)/\sigma \tag{7.9}$$

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

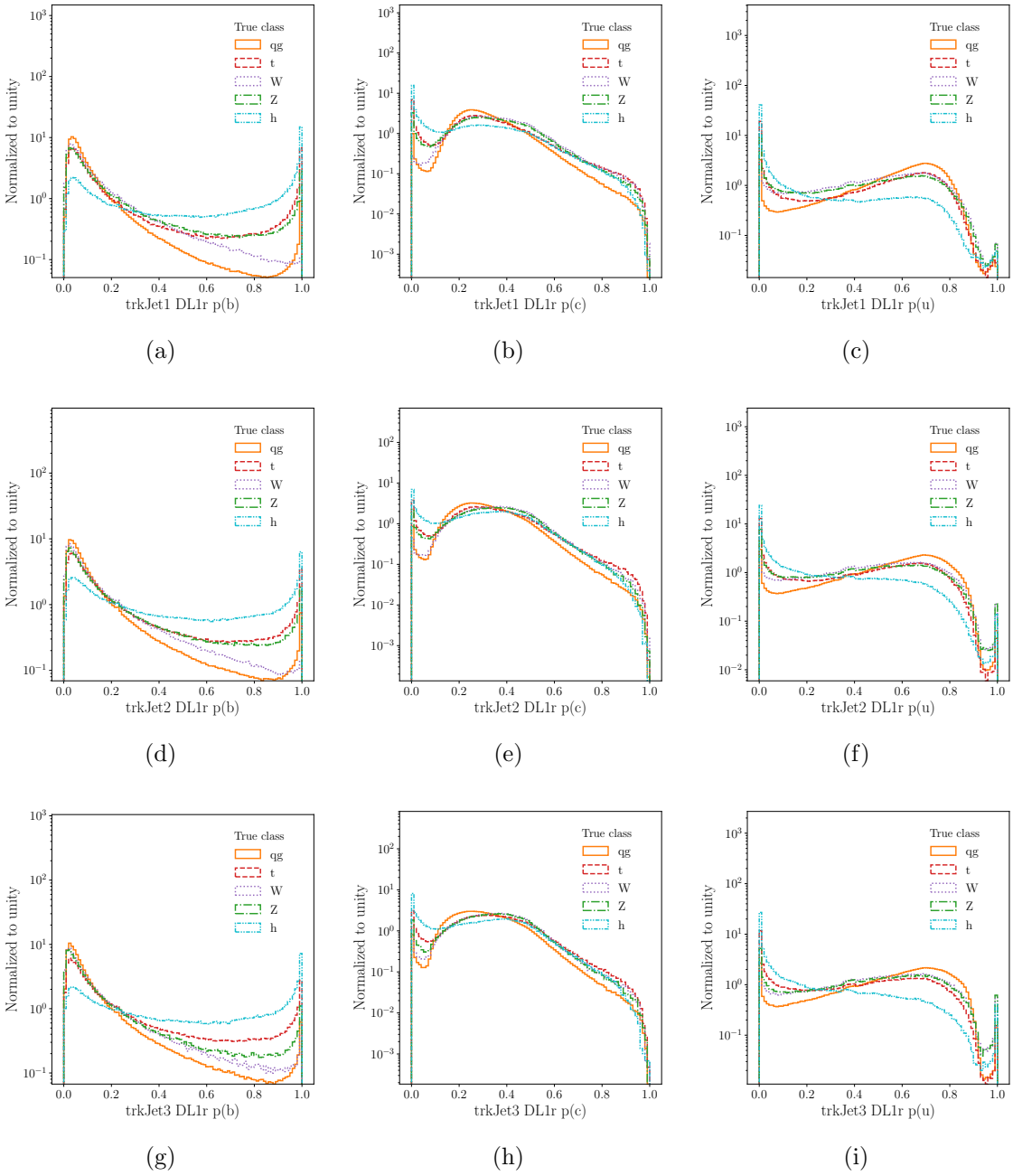


Figure 7.11: Examples of input variables to the MergedMCT describing the b-tagging DL1r scores of the three leading track jets inside the large-R jet.

samples were re-weighted to obtain a flat large-R jet p_T distribution. The re-weighting is performed for each class separately. In order to have the most accurate re-weighting, the $p_T$distribution is re-binned using the finest binning that allows to retain a statistical error

below 5% in each bi. The $p_T$density is 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 re-weighted 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 re-weighting procedure, each jet-event i true labeled, for example, as a Higgs boson is assigned a weight $w_i = w_i^{\text{Flat}p_T} \cdot w^{Higgs}$.

Hyper-parameters and training

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 the training when no further reduction in the loss was observed for more than 40 epochs. The hyper-parameters 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.

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 MCTis performing as expected. Similarly, the plots in Fig. 7.13 show the log-likelihood ratios for different true class combinations, showing significant discrimination power.

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. 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 W and Z scores, will be considered, resulting in an average accuracy in the network predictions above 75% for all classes.

The ROC (Receiver Operating Characteristic) curves in Fig. 7.15 show the signal efficiency vs. background rejection for all signal-background combinations. The ROC is built using the output score distributions for the signal class from the given true signal and true background samples. For each signal-background pair, the ROC is built using events in different p_T bins. Similarly to the confusion matrices, only a minor p_T dependence is observed, with degradation only in the lowest p_T bin of $[200, 250]$ GeV. 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. 7.16.

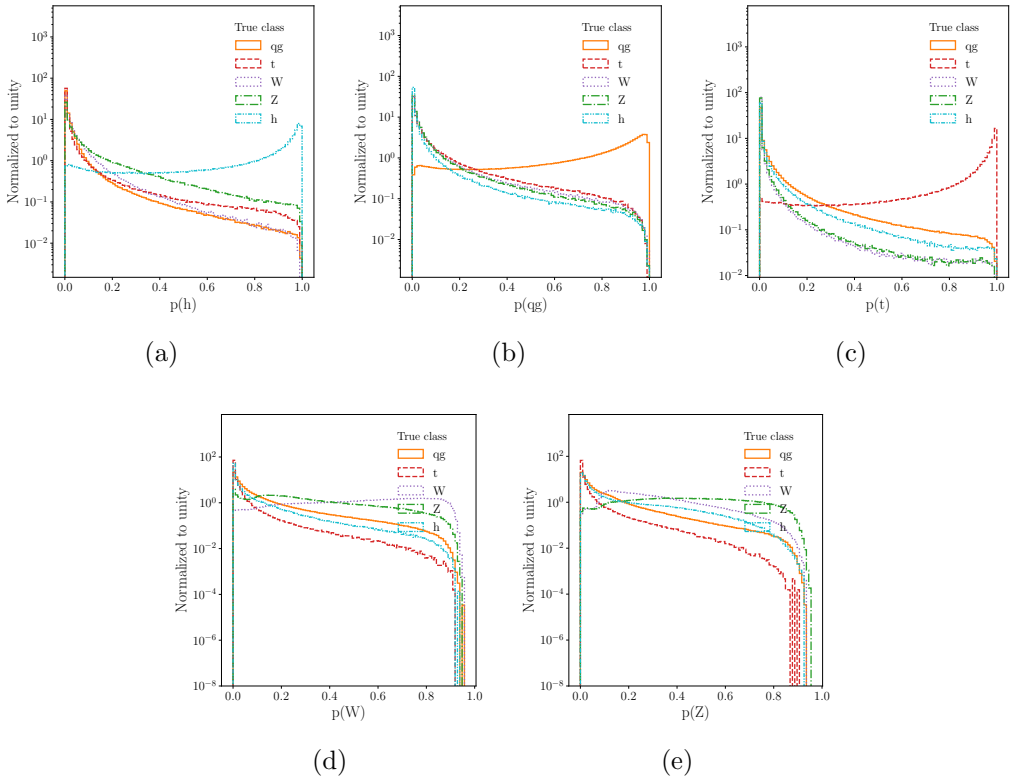


Figure 7.12: Merged MCT output probabilities tagging on true class label.

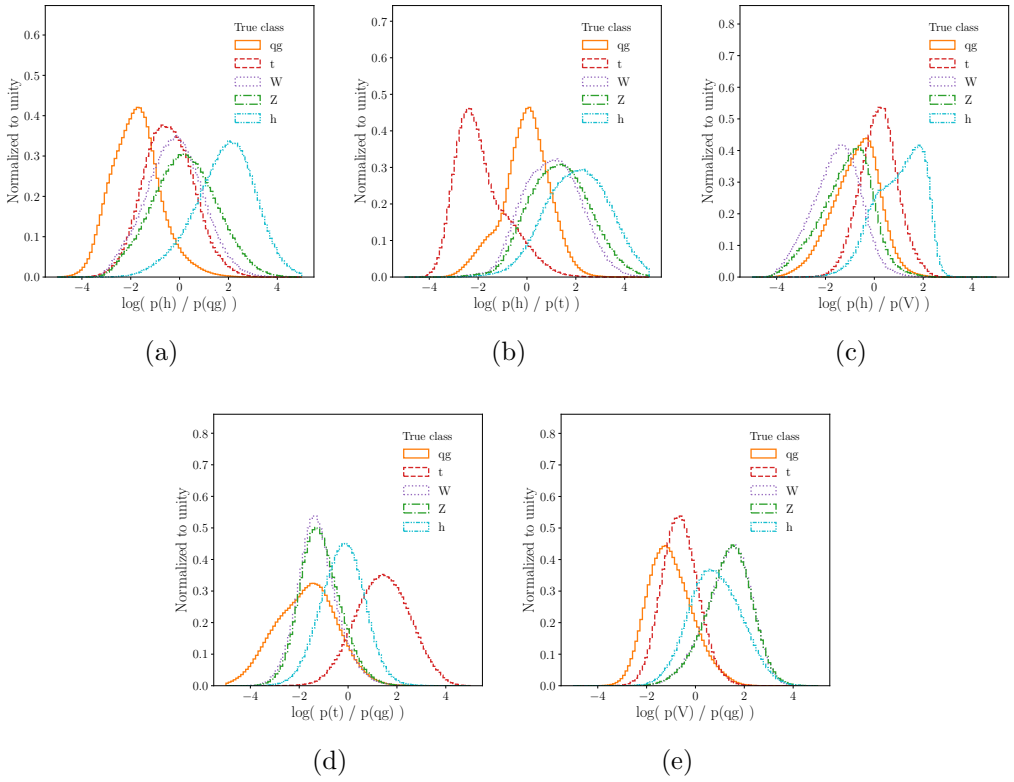


Figure 7.13: Merged MCT output log-likelihood ratios tagging on true class label. The spikes at +4 and -4 are due to a requirement on the scores to be greater than 1.e-4.

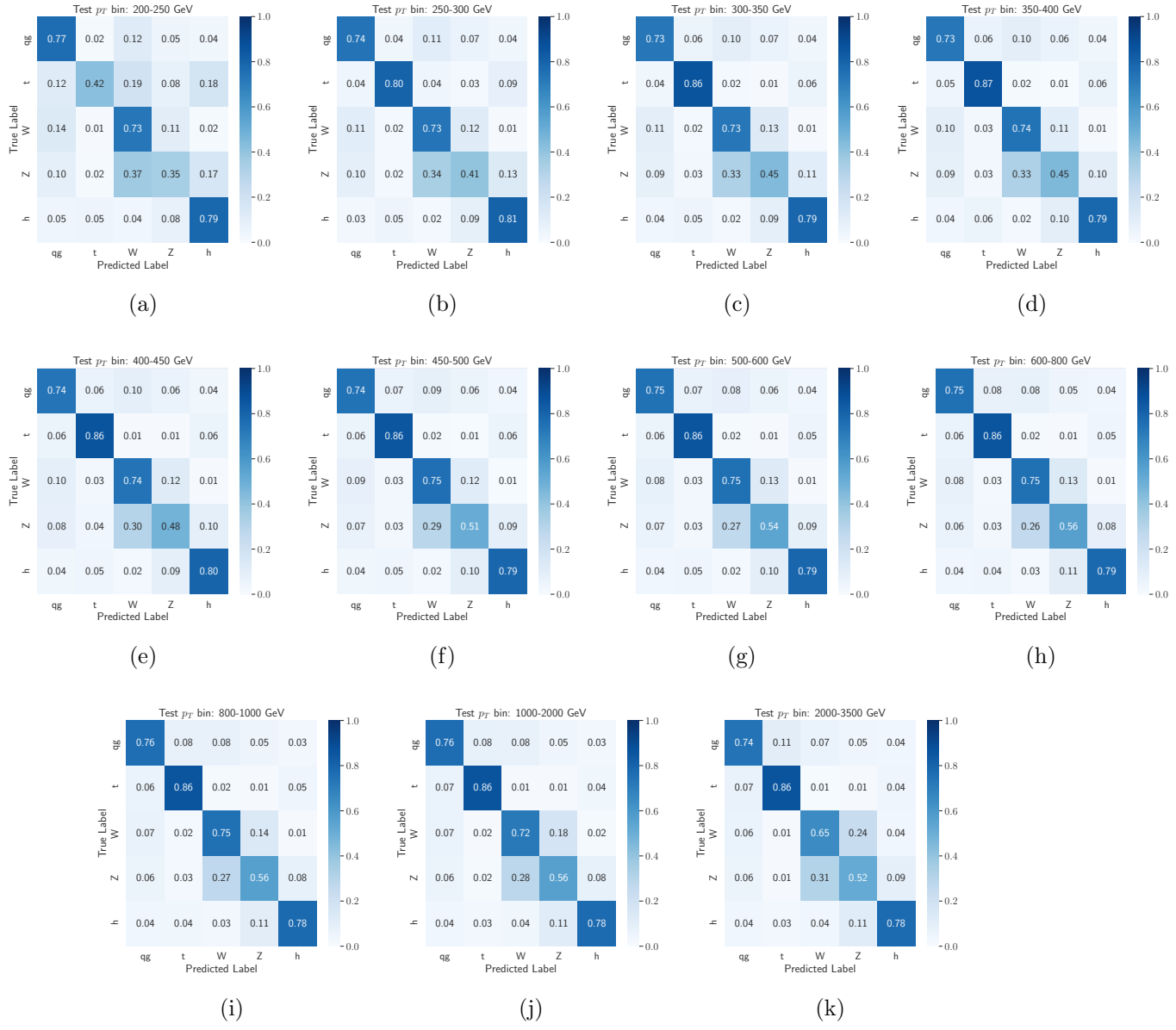


Figure 7.14: Merged MCT confusion matrices as a function of p_T of the large-R jet.

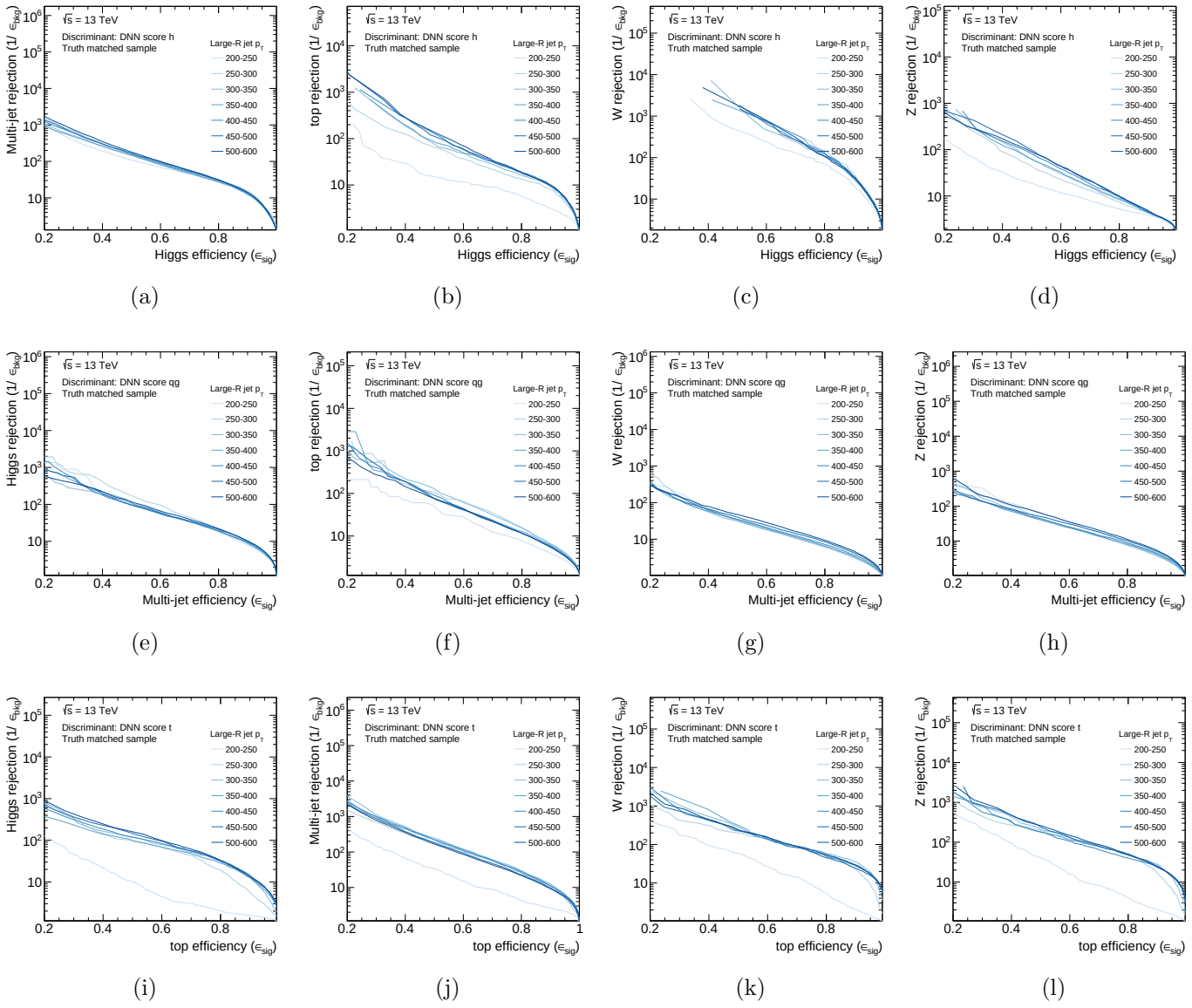
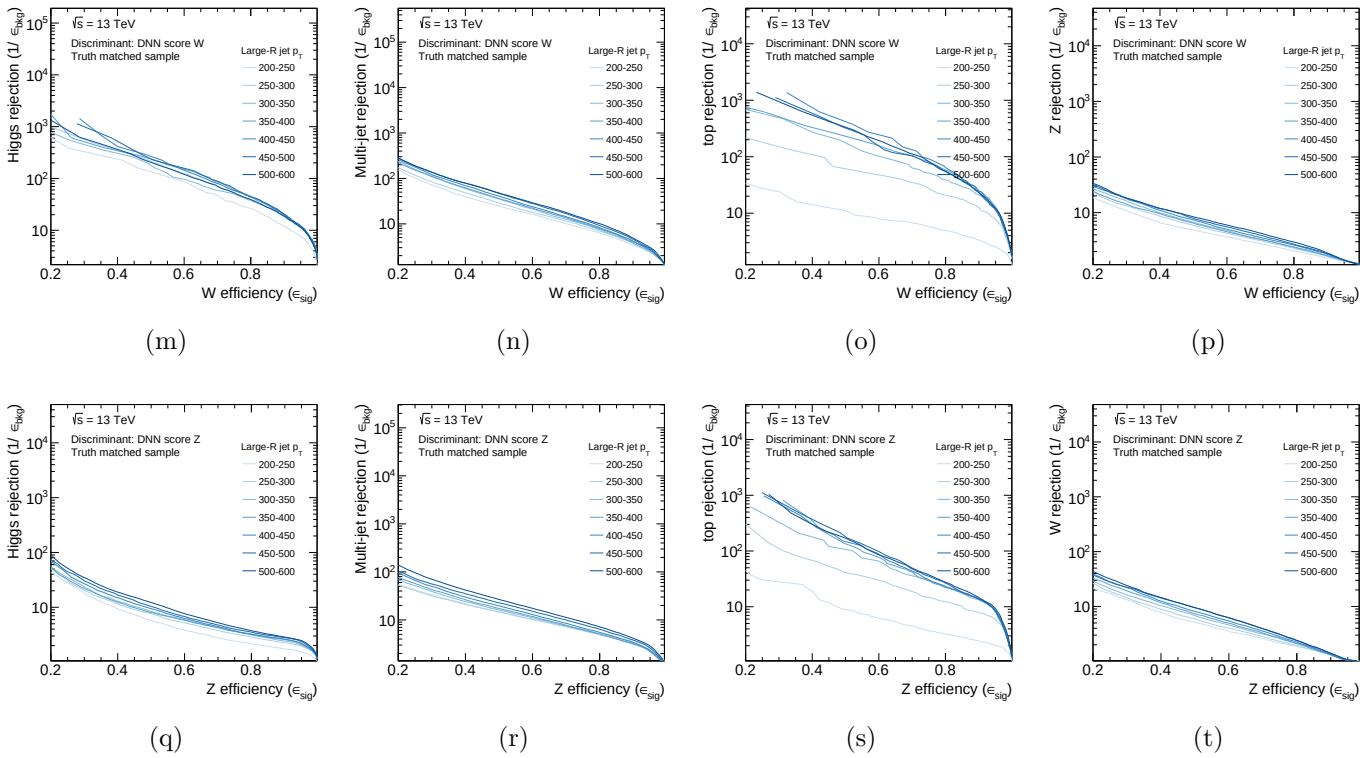

√s = 13 TeV
Discriminant: DNN score h
Truth matched sample
Large-R jet p_T
200-250
250-300
300-350
350-400
400-450
450-500
500-600
Multi-jet rejection (1/ ε_bkg)
top rejection (1/ ε_bkg)
W rejection (1/ ε_bkg)
Z rejection (1/ ε_bkg)
Higgs efficiency (ε_sig)
(a)
(b)
(c)
(d)
Discriminant: DNN score qg
Higgs rejection (1/ ε_bkg)
Multi-jet efficiency (ε_sig)
(e)
(f)
(g)
(h)
Discriminant: DNN score t
top efficiency (ε_sig)
(i)
(j)
(k)
(l)

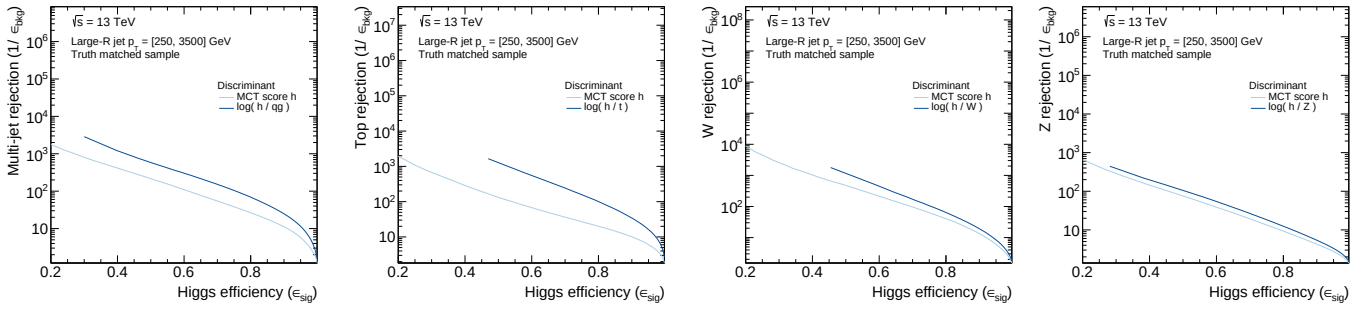
Figure 7.15



(m) (n) (o) (p)

(q) (r) (s) (t)

Figure 7.15: Merged MCT ROC curves built using the output score of the signal class for each signal vs. background combination. 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 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 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. These are $R = 0.4$ 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 $p_T > 20\,\text{GeV}$, with the leading jet required to have $p_T > 45\,\text{GeV}$.

MC samples and truth labeling

The resolved event topology is more dependent on the event type. For this reason, the samples used for training are a subset of the MC 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 RSG as sources for the W and Z boson classes. The ttbar samples were used as sources of top-quark, and the V+jets samples as sources of light quarks and gluons (QCD).

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 of the η acceptance region, as well as cases of "super merged" boson

Class	Process
Light quarks and gluons	V+jets
Top quark	$t\bar{t}$
W boson	HVT V'→ WW/WZ (W →qq), RSG-WW, Radion-WW
Z boson	HVT V' → WZ (Z→qq), RSG-ZZ, Radion-ZZ
Higgs boson	HVT V' → VH, gg-A

Table 7.7: Resolved MCT MC samples used to select truth matched jets for the training samples

decays, where the truth quarks overlap. For top quark samples, 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 $\eta < 2.5$. Moreover, 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 the default values of 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. This was an approximation to select events with resonance masses below 1 TeV.

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 be information describing 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 Fig. **??**.

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.

Pre-processing

Because of the much larger number of V+jets and ttbar MC 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 MergedMCT, 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.

Hyper-parameters and training

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 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 1000, with early stopping implemented to stop the training when no further reduction in the loss was observed for more than 40 epochs. The hyper-parameters 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

Fig. 7.19 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. 7.20 shows the log-likelihood ratios for different class combinations, where the $p(V)$ 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

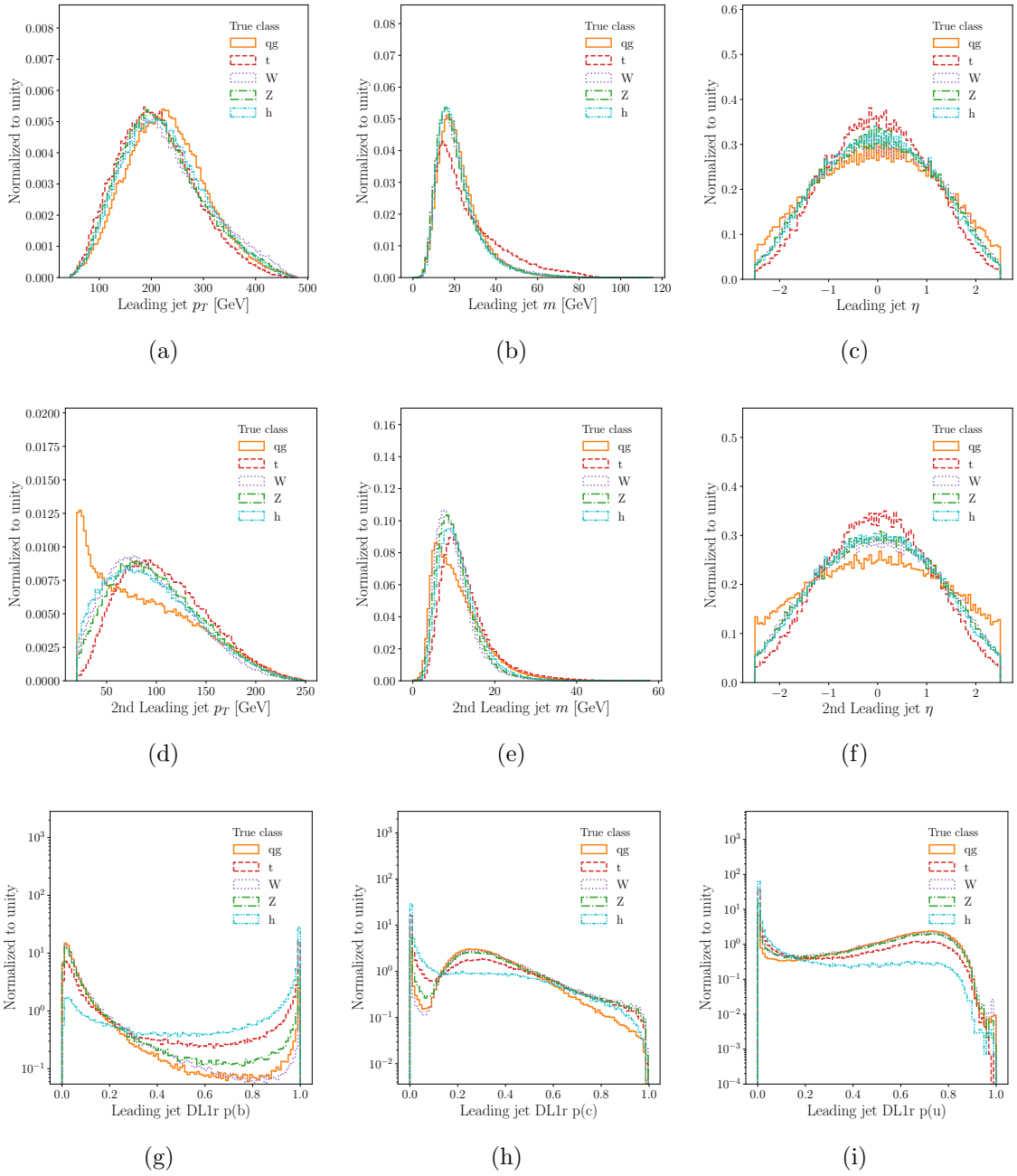


Figure 7.17: Set of input variables for the Resolved MCT tagged according to the true class representing the kinematic variables of the leading (top) and 2nd leading (middle) jets, and the b-tagging scores of the first leading je (bottom).

looking at the likelihood ratios, when uncertainty about the other classes is taken into account. A confusion remains between the Higgs and V classification, which will be discussed next.

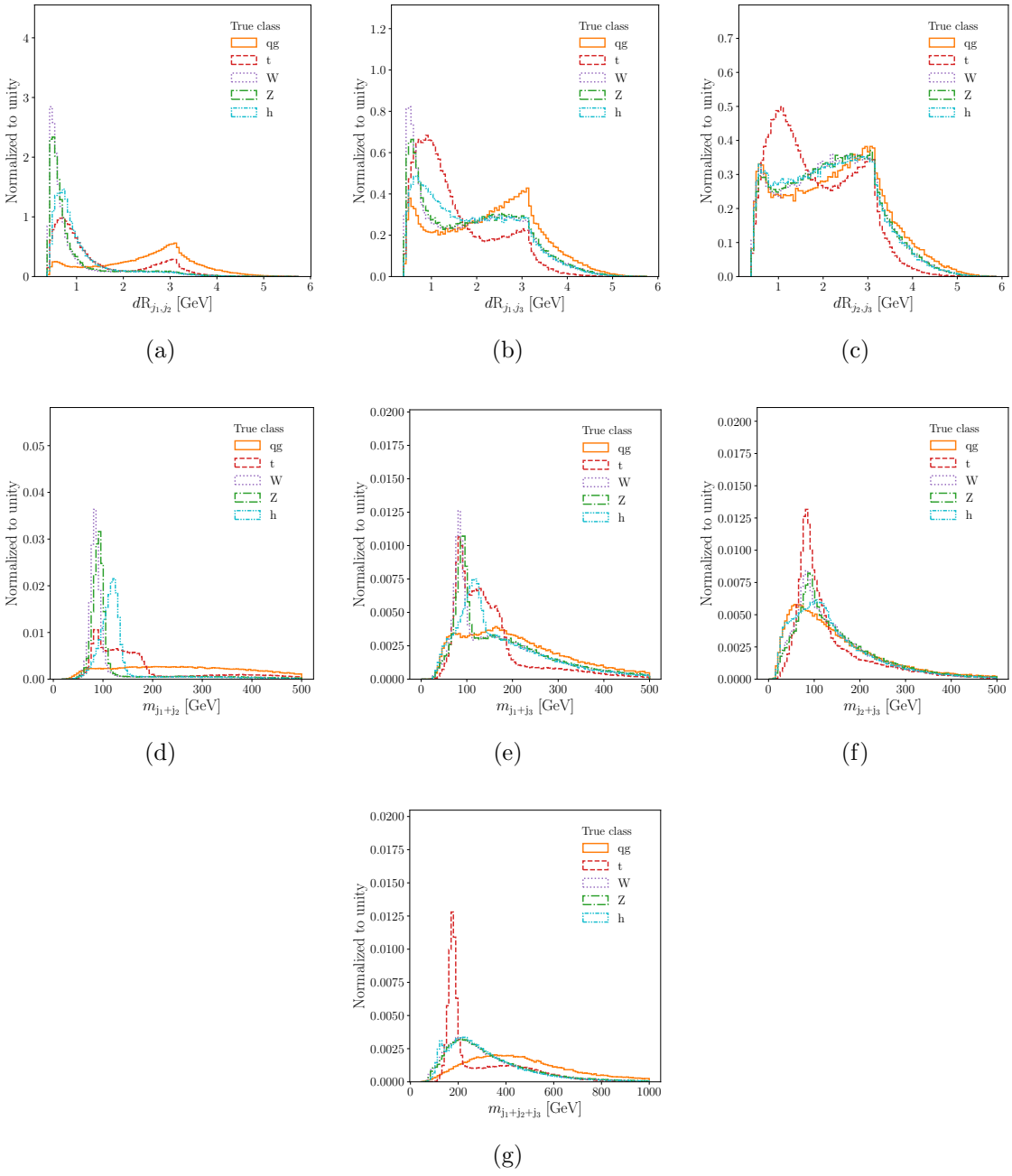


Figure 7.18: Set of input variables for the Resolved MCT tagged according to the true class representing the mass of reconstructed objects.

The confusion matrices are shown in Fig. 7.21. 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 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 average accuracy in the network predictions is above 75% for all classes. The confusion matrix for true Higgs, W, and Z events separated by true sample of origin is shown in Fig. 7.22. The same matrix is then shown in Fig. 7.23 separating the events in p_T bin and number of b-tagged three leading small-R jets. 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 only 15% of the times, and a Higgs boson hadronically decays to two b-quarks 95% of the times.

The ROC curves in Fig. 7.24 show the signal efficiency vs. background rejections for all signal-background combinations. The ROC is built using the output score distributions for the signal class for the given true signal and background samples. 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 QCD class and the greatest degradation on performance occurs for the lowest p_T bin.

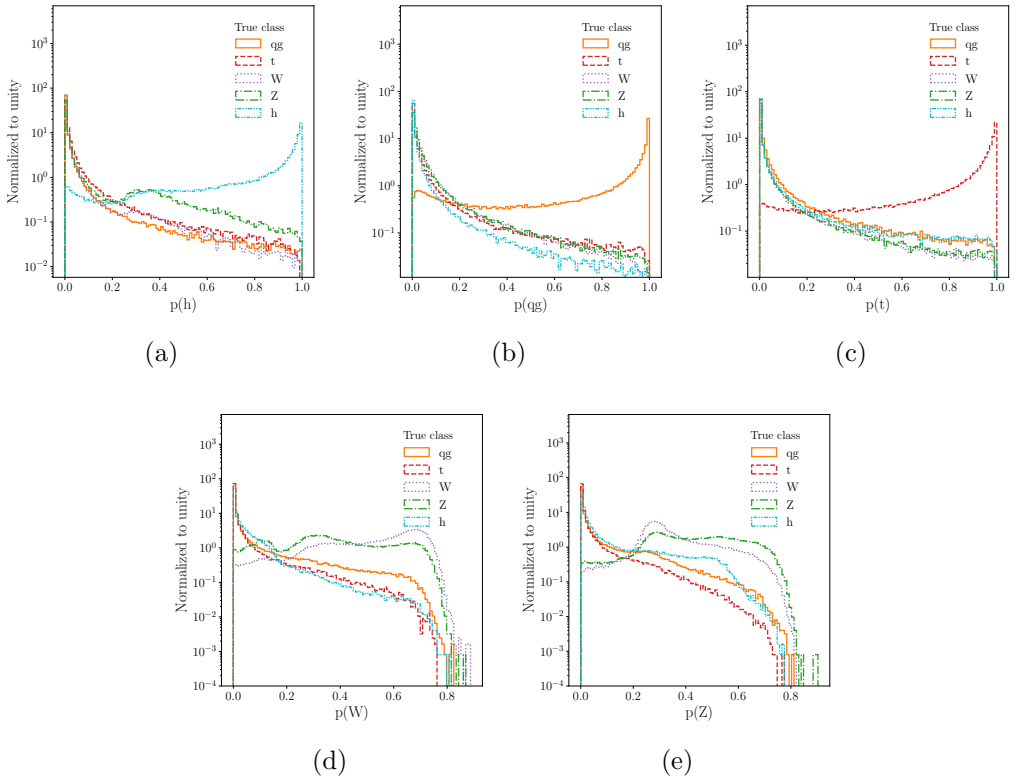


Figure 7.19: Resolved MCT output probabilities tagging on true class label.

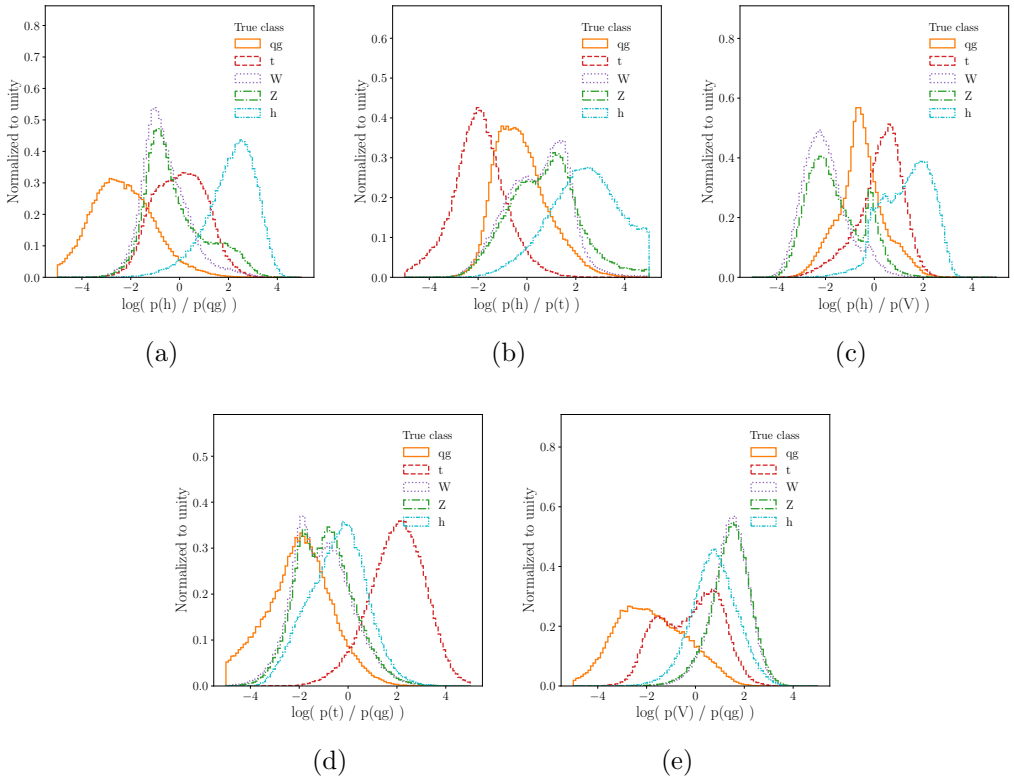
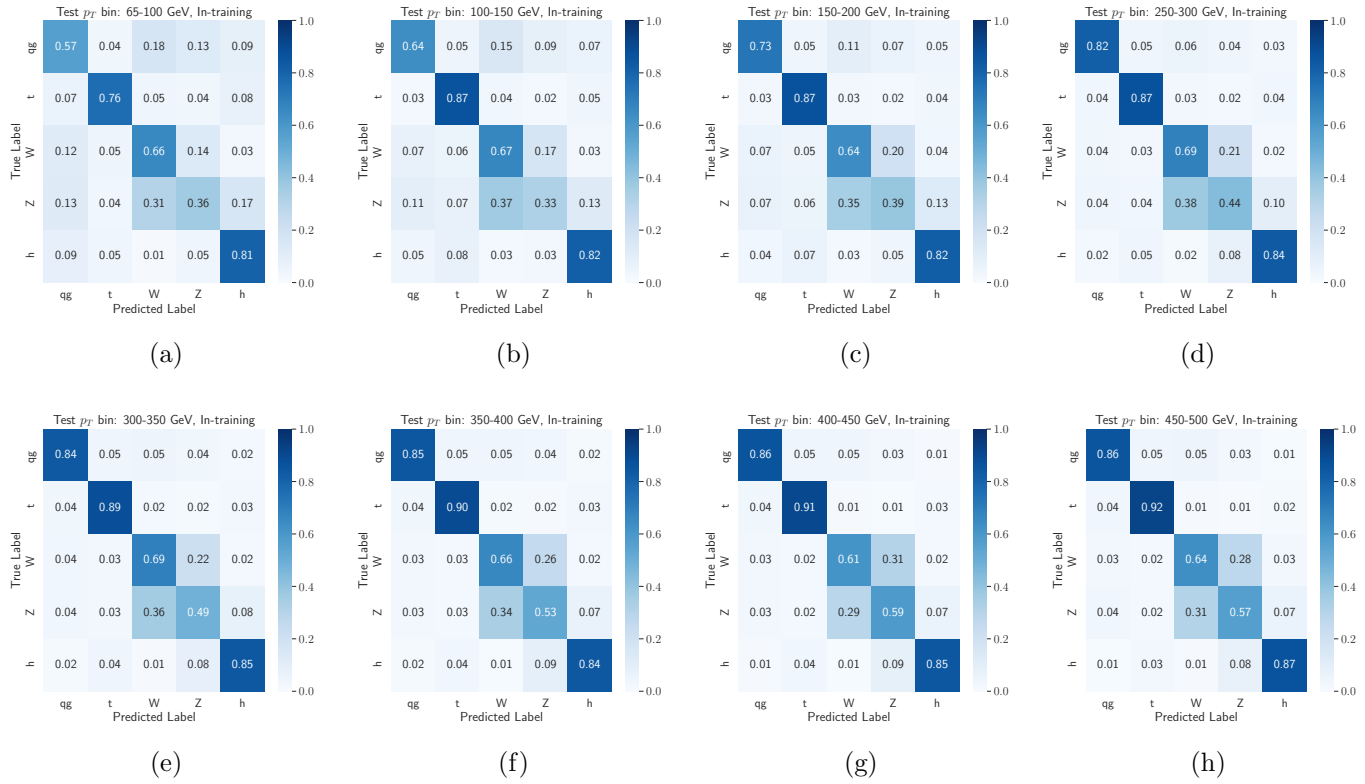


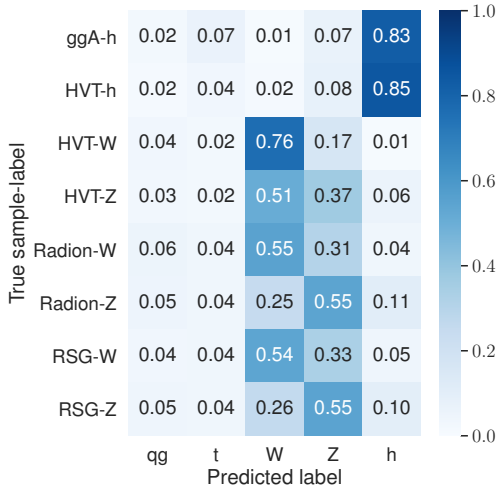
Figure 7.20: Resolved MCT output log-likelihood ratios tagging on true class label.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 7.21: Resolved MCT confusion matrices as a function of the sum of the p_T of the two leading small-R jets.



(a)

Figure 7.22: Resolved MCT confusion matrix separating the true signal classes by the true MC sample of origin.

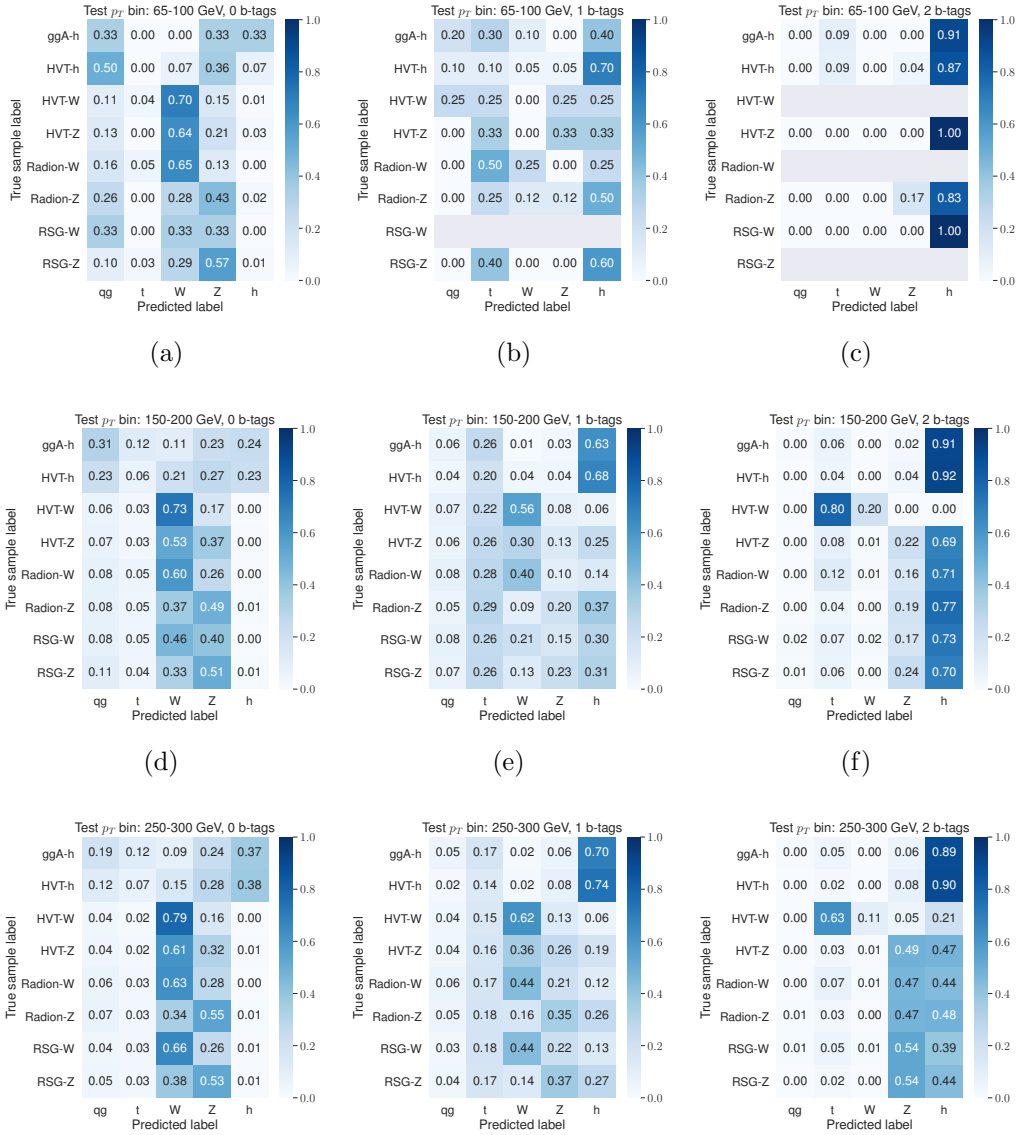

Test p_T bin: 65-100 GeV, 0 b-tags
Test p_T bin: 65-100 GeV, 1 b-tags
Test p_T bin: 65-100 GeV, 2 b-tags
Test p_T bin: 150-200 GeV, 0 b-tags
Test p_T bin: 150-200 GeV, 1 b-tags
Test p_T bin: 150-200 GeV, 2 b-tags
Test p_T bin: 250-300 GeV, 0 b-tags
Test p_T bin: 250-300 GeV, 1 b-tags
Test p_T bin: 250-300 GeV, 2 b-tags
True sample label
Predicted label
ggA-h
HVT-h
HVT-W
HVT-Z
Radion-W
Radion-Z
RSG-W
RSG-Z
qg
t
W
Z
h
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)

Figure 7.23

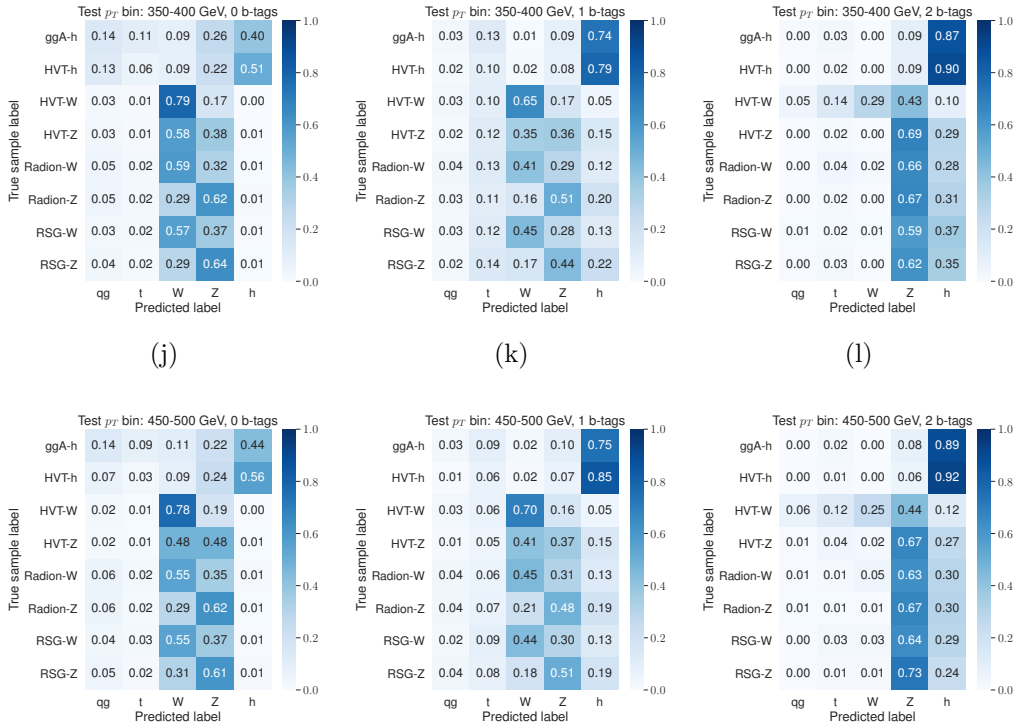


Figure 7.23: Resolved MCT confusion matrices separated by true sample of origin and number of b-tagged three leading small-R jets, as a function of the sum of the p_T of the two leading small-R jets.

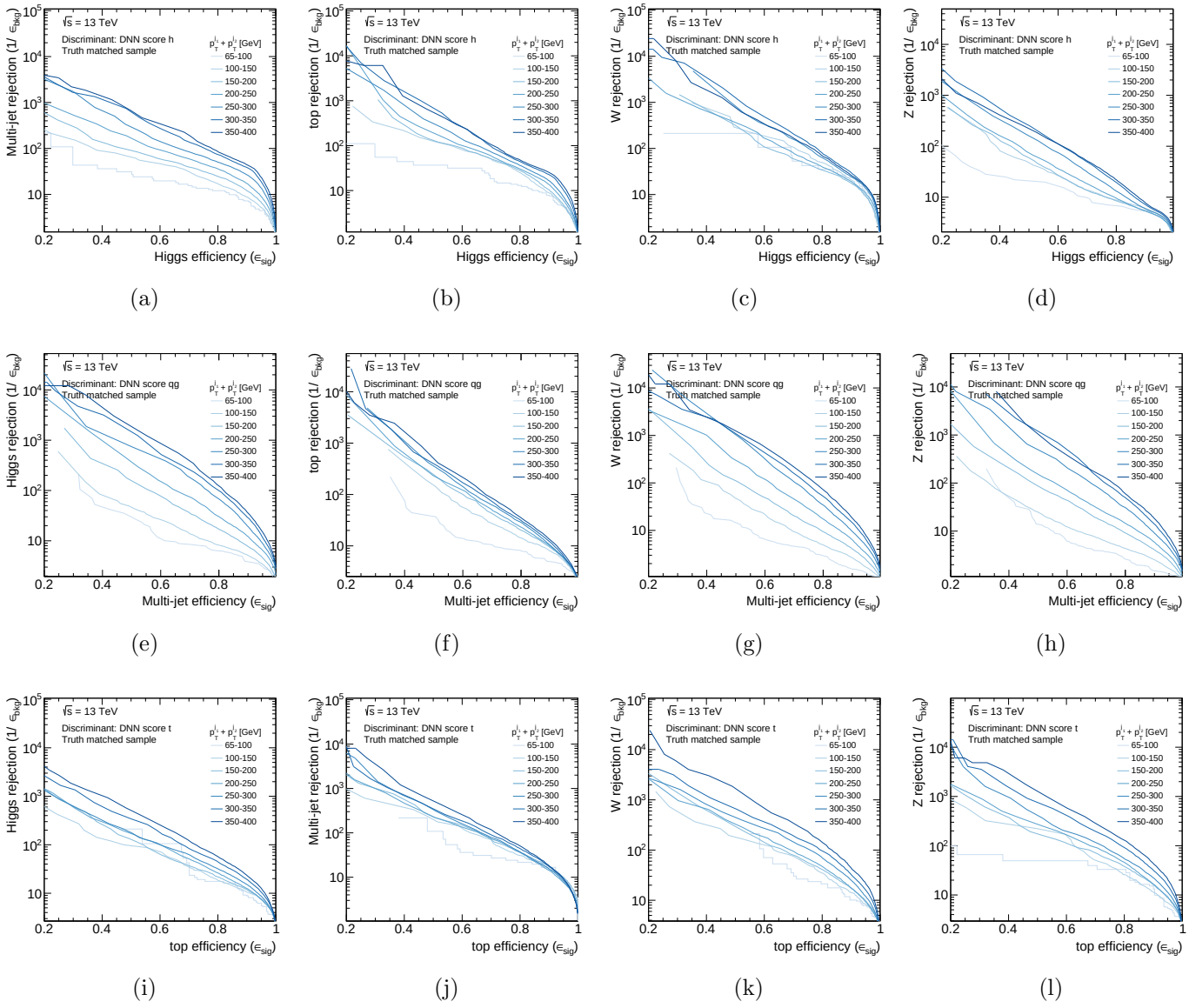
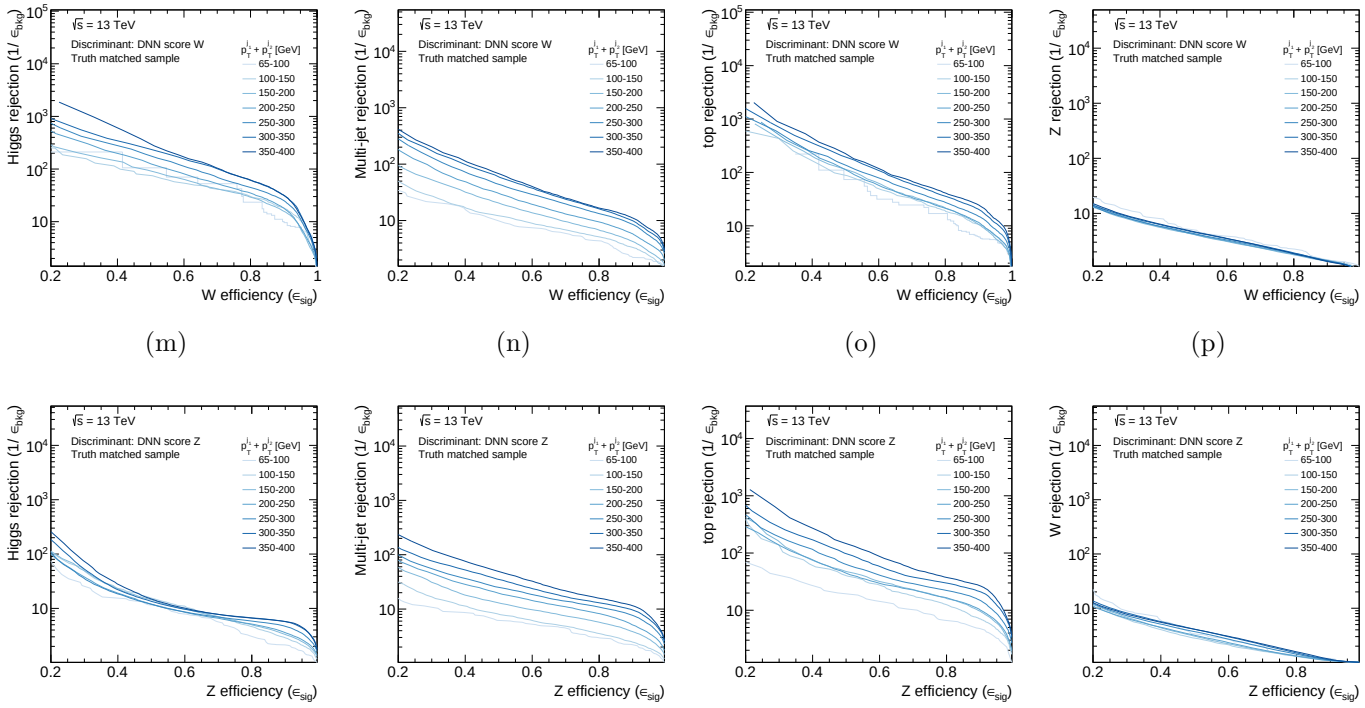

√s = 13 TeV
Truth matched sample
65-100
100-150
150-200
200-250
250-300
300-350
350-400
Multi-jet rejection (1/ ϵbkg)
top rejection (1/ ϵbkg)
W rejection (1/ ϵbkg)
Z rejection (1/ ϵbkg)
Higgs rejection (1/ ϵbkg)
Multi-jet rejection (1/ ϵbkg)
Higgs efficiency (ϵsig)
Multi-jet efficiency (ϵsig)
top efficiency (ϵsig)
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)

Figure 7.24



(m) (n) (o) (p)

(q) (r) (s) (t)

Figure 7.24: Resolved MCT ROC curves built using the score of the signal class for each signal vs. background combination. The ROCs are shown as a function of the sum of the p_T of the two leading small-R jets.

7.10 Orthogonalization of VV and VH signal regions in the search for a new HVT boson

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. 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.*These are `PriorityResolved` and `PriorityMerged` for the VH and VV analysis, respectively. After prioritization is enforced, the final regions within a given analysis are fully orthogonal.

However, 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 in Fig. 7.25. The x-axis shows the cuts on the small-R jet mass $m(jj)$, while the y-axis shows the cuts 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.

The VV and VH semi-leptonic analyses were part of previous combination efforts [158, 159, 178] 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

Analysis	Channel	Resolved	Merged
VH	0-lep 1-lep 2-lep	$75 \leq m(jj) \leq 145$ $100 \leq m(jj) \leq 145$	$75 \leq m(J) \leq 145$
VV	W Z	$68 \leq m(jj) \leq 98$ $78 \leq m(jj) \leq 106$	`Pass_WZTagger_WMassCut` `Pass_WZTagger_ZMassCut`

Table 7.10: Mass window definitions in VV and VH signal regions. All numbers are in units of GeV. In VH 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.

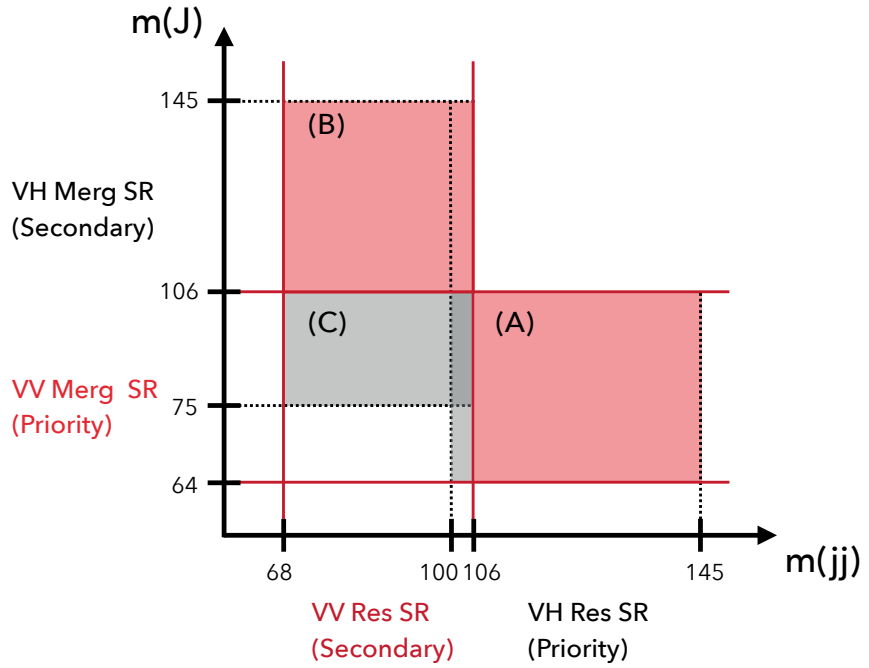


Figure 7.25: Schematic visualization of VV and VH resolved (x-axis) and merged (y-axis) mass windows overlap. See text for explanation.

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 in 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. 7.25, 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` production is shown in Tab. 7.11 for the 2-lepton channel.

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 that end up in both VV and VH signal regions, compared to the total number of events available in data and MC and to the total number of events entering any signal region. The number of events in the mixed-overlap regions that does not get orthogonalized by the mass cut strategy is also shown.

In this publication for the first time the VV and VH processes are 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, this opens up also the possibility for a more nuanced and 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.

7.10.2 Studies overview

The studies presented here use the `mc16a` dataset with $36.1\,\text{fb}^{-1}$ 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 will be used, as shown in Tab. 7.12, and the corresponding signal regions, shown in Tab. 7.13.

	0-lep	1-lep	2-lep
VH	HVT Z'→ ZH	HVT W' → WH	HVT Z'→ ZH
VV	HVT W'→ WZ	HVT W' → WZ	HVT W'→ 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). In these studies, an event is in a VV or VH SR, if any SR in the corresponding column is active.

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 orthogonalization strategy proposed in the analysis, 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 $p_{\text{max}} = \max(p(h), p(V))$.

  - If $p_{\text{max}} = p(h)$, 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 results will be compared:

- *Baseline*: no orthogonality.
- *MCT*: use MCT $p(h)$ and $p(V)$ scores.
- *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, signal significance as a function of heavy resonance mass, and expected limit sensitivity.

0 Lepton channel

		data	HVT-WZ	HVT-WW	HVT-ZH	HVT-WH	ttbar	Wjets	Zjets
MCT Sorting	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 Sorting	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 Sorting	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 Sorting	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 Sorting	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 Sorting	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

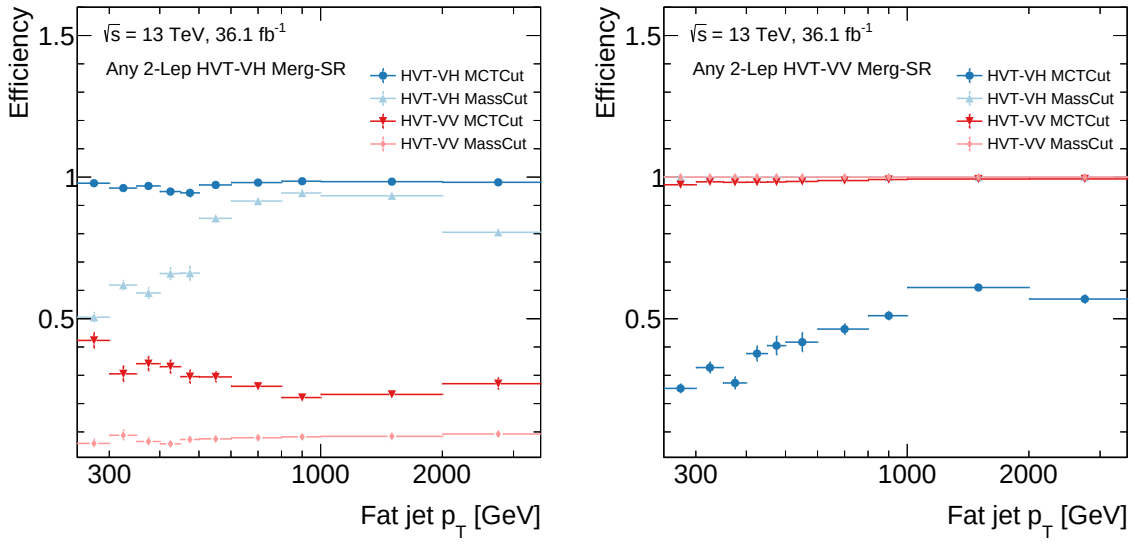
The effect of the orthogonality cut on the signal efficiency is compared for the *MCT* and *MassCut* strategies. This is shown for the 2-lepton channel in Fig. **??**, and for the 0- and 1-lepton channels in the Appendix **??**. The events are grouped into one of four categories – resolved or merged, and VV or VH – depending on which signal region they enter, according to Tab. 7.13.

Denoting the set of events entering a given signal region as $\mathcal{R}$, the signal efficiency ϵ_S of signal S in the region is given by the ratio of the number of events $\in \mathcal{R}$ that pass the orthogonality cut to the total number of 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 7.26 the p_T corresponds to the p_T of the large-R jet, while in the resolved regions 7.27 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. In dark hue the orthogonality cut is applied following the *MCT* strategy, in light following 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. As shown in Fig. 7.27, 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 in Fig. 7.26 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.

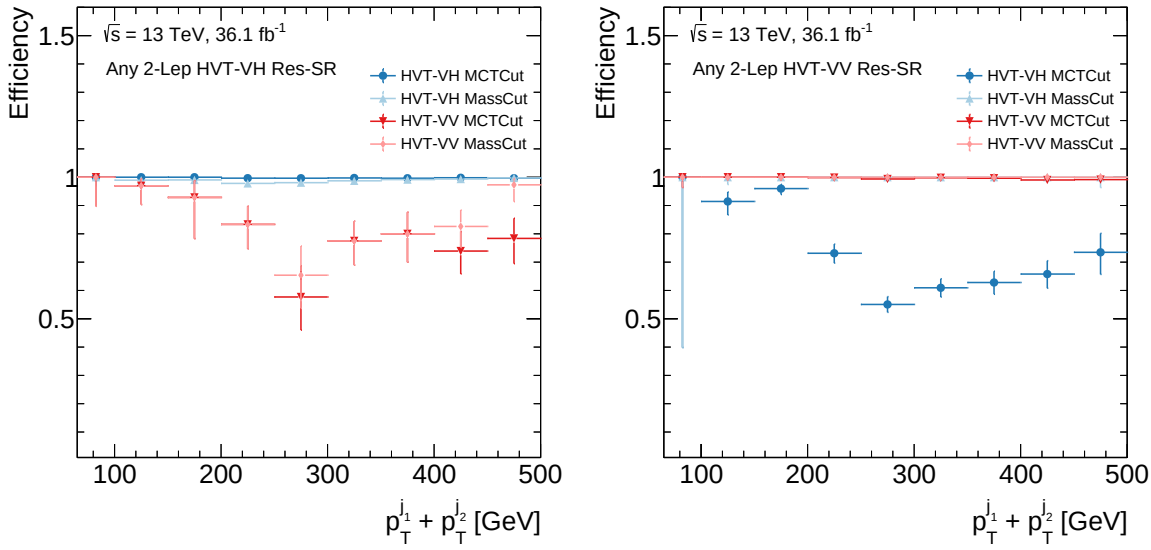
7.10.5 Signal significance

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 $\sqrt{n}$.



(a) (b)

Figure 7.26: 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 (a) and HVT-VV Merg SR(b).



(a) (b)

Figure 7.27: 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 (a) and HVT-VV Res SR (b) .

The significance S is then calculated as,

$$\mathcal{S}_i = \frac{n_s}{\sqrt{n_s + n_b + \sigma_b^2}}. \tag{7.10}$$

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,

$$\mathcal{S} = \sqrt{\sum_i \mathcal{S}_i^2}. \tag{7.11}$$

In the following the histogram used to calculated to total binned significance is that of the final discriminant for the analysis, i.e. the invariant mass distribution.

The significance scans are built for each signal region as follows. For each region, the histogram of the invariant mass distribution of the reconstructed VV or VH candidate is built using the SM background MC samples. An example of this is shown for one VH signal region in Fig. 7.28, where the background histograms are stacked on top of each other. Here, $n_b = \sum_j n_b^j$, where j represents each background sample. Then, the same histogram is built for each signal mass point and the total significance is calculated using Eq. (7.11). The total significance is then shown as a function of signal mass point.

Fig.**??** shows the expected HVT signal significance as a function of V' resonance mass in VV and VH signal regions. This is shown for the 2- lepton channels, with similar performance having been observed for the 0- and 1-lepton channel. The *MassCut* strategy causes a reduction in VH signal significance in the Merged VH signal regions, as shown in Fig. 7.29, while the *MCT* strategy does not affect the significance. In the VV signal regions in Figs. **??**, 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.

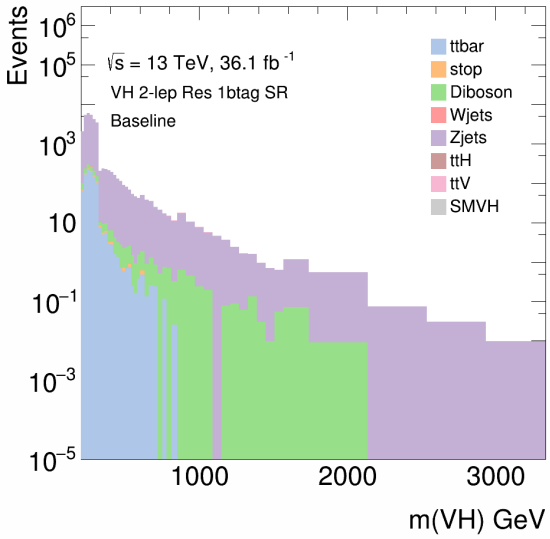
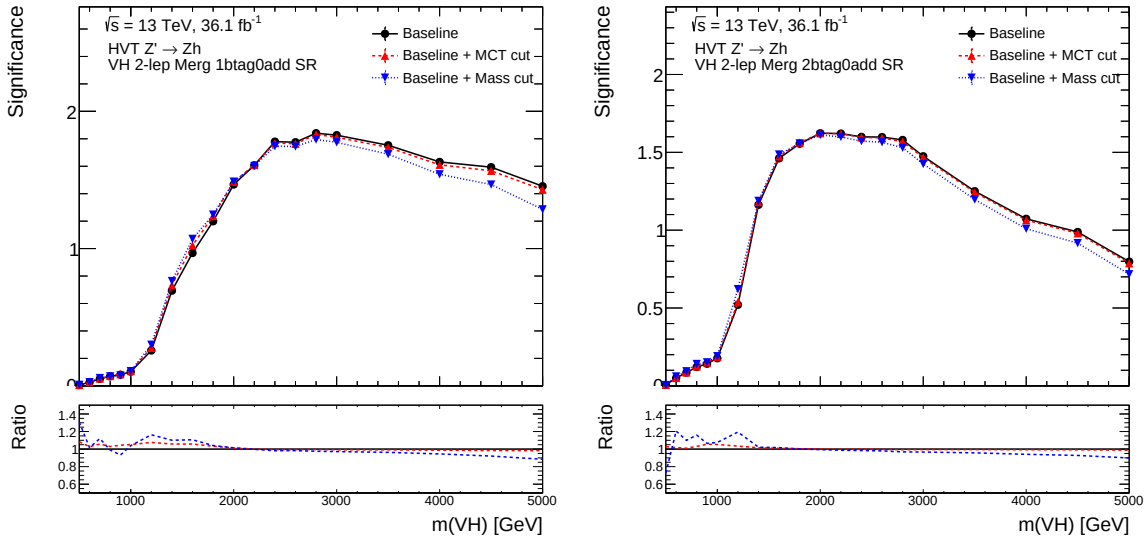
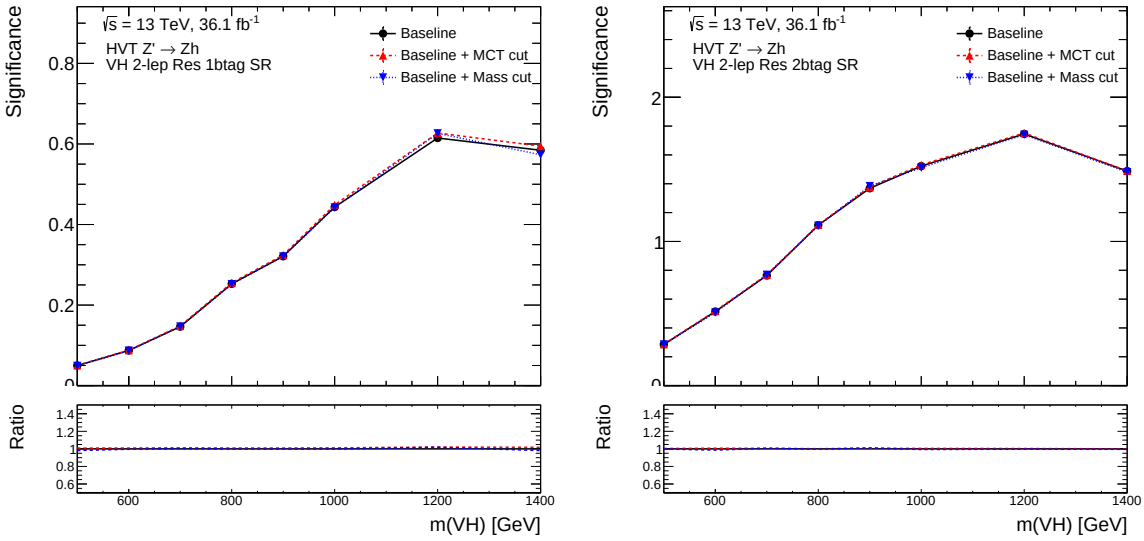


Figure 7.28: Example of invariant mass distribution histogram in one VH signal region for all SM backgrounds. The different background histograms are shown stacked on top of each other.



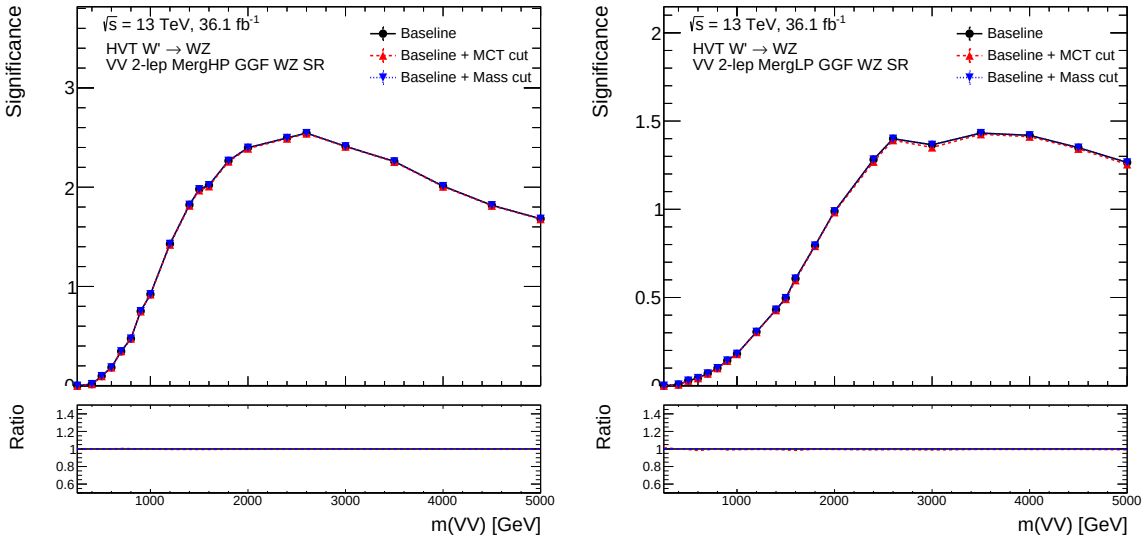
(a) (b)

Figure 7.29: Significance scans as a function of Z' resonance mass in VH merged signal regions. The baseline analysis cutflow is compared to the MCT and MassCut strategies for orthogonalization.



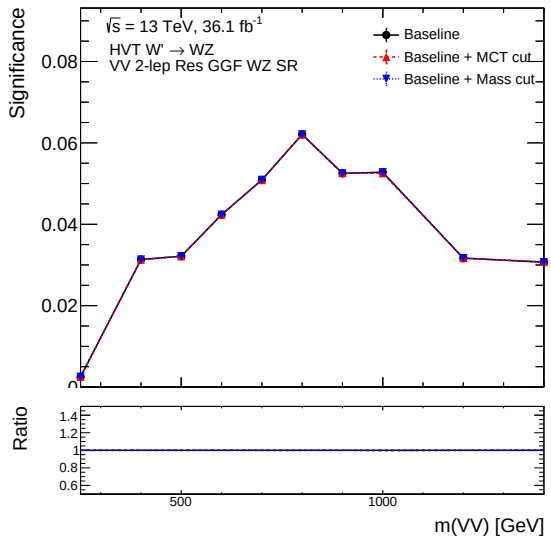
(a) (b)

Figure 7.30: Significance scans as a function of Z' resonance mass in VH signal resolvedregions. The baseline analysis cutflow is compared to the MCT and MassCut strategies for orthogonalization.



(a) (b)

Figure 7.31: Significance scans as a function of Z' resonance mass in VV merged signal regions. The baseline analysis cutflow is compared to the MCT and MassCut strategies for orthogonalization.



(a)

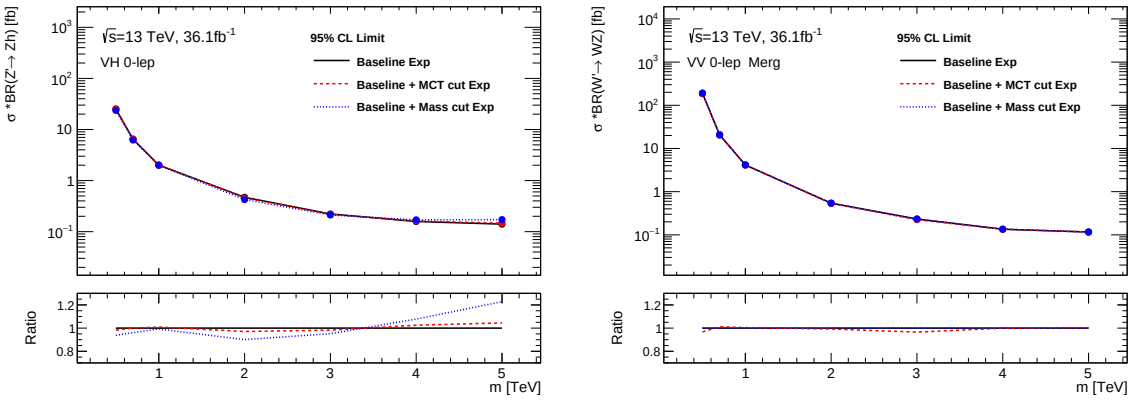
Figure 7.32: Significance scans as a function of Z' resonance mass in VV resolved signal regions. The baseline analysis cutflow is compared to the MCT and MassCut strategies for orthogonalization.

7.10.6 Expected limit sensitivity

The expected limit sensitivities when using the two orthogonalization strategies *MCT* and *MassCut* 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. **??** 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, respecitively. As expected from the previous studies, the *MCT* strategy does not cause any loss in performance with respect to the *Baseline* analyses. On the other hand, the *Mass cut* strategy causes a loss of sensitivity at high mass in the 0- and 2-lepton channel. 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

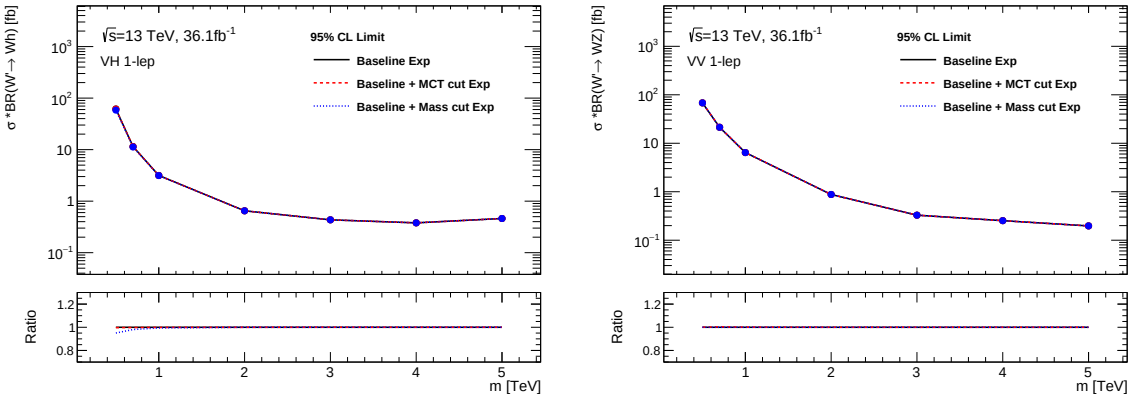
significance.

In conclusion, the *MCT* strategy does not cause any loss in sensitivity in the limits, allows to recover up to 20% loss in sensitivity with respect to *Mass cut* strategy, and allows to simplify the combined search for new heavy resonances.



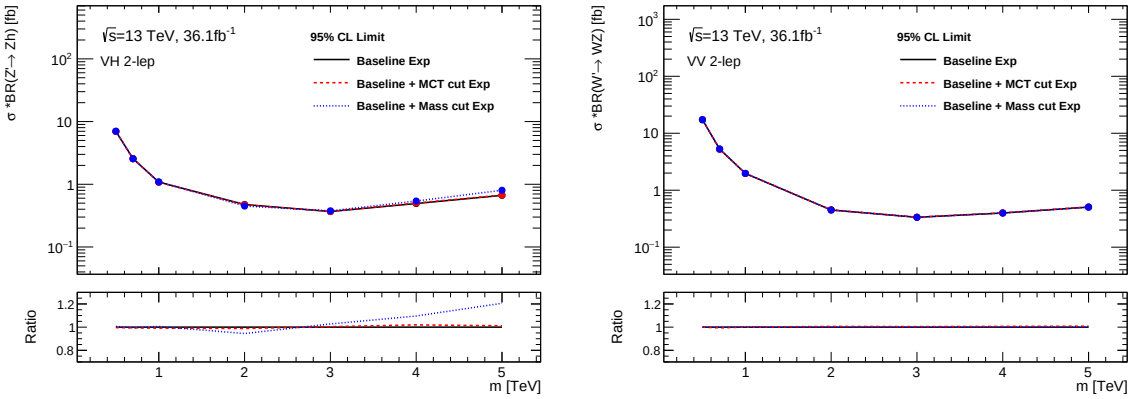
(a) (b)

Figure 7.33: Comparison of expected limits in VH and VV 0-lepton channel, shown in the inclusive regions (note that VV does not have a resolved region in the 0-lepton channel). The baseline analysis cutflow is compared to the MCT and MassCut strategies for orthogonalization.



(a) (b)

Figure 7.34: Comparison of expected limits in VH and VV 1-lepton channel, shown in the inclusive regions. The baseline analysis cutflow is compared to the MCT and MassCut strategies for orthogonalization.



(a) (b)

Figure 7.35: Comparison of expected limits in VH and VV 2-lepton channel, shown in the inclusive regions. The baseline analysis cutflow 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 pre-selection and control regions of the VV+VH semi-leptonic 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 pre-selection (Sec. 7.11.2) and control regions (Sec. 7.11.3). In the following, representative studies are shown in the 2-lepton pre-selection regions, with the 0- and 1-lepton channels provided in the Appendix A. Modeling in a top-enriched control region is also 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 on the MCT on the final analysis sensitivity and no effect is observed.

7.11.1 Derivation of background normalization scale factors

The normalization scale factors of the background processes are obtained by a binned maximum likelihood (Eq. (6.7)) fit of the predicted observable of interest to the observed data.

The number of background events 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 determines a scale factor $\tau = N^{\text{data}}/N^{MC}$. Normalization nuisance parameters can be included in the fit model to allow to vary the normalization of a given background process and obtain a maximum likelihood estimator of the scale factor. 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 $\hat{\tau}$. This type of nuisance parameters introduce significant freedom in the optimization procedure, so they should only be used when necessary to avoid overfitting. In the case of constrained nuisance parameter, 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 $\epsilon^b(\alpha) = \epsilon_0^b \times (1 + \delta \cdot \alpha)$, where α is sampled from 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 an 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 ± 2.4) 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 mis-modelling. Nonetheless, the residual mismodellings 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 larger 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 for the 2-lepton channel are shown as a representative example. The 0- and 1- lepton channels can be found in the Appendix A.

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 different flavor contributions, this was necessary to disentangle mis-modelling originating from incorrect background normalizations from mis-modelling induced by the MCT. The

following V+jets sub-samples were defined according to the truth flavour 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 contribution[4].

When considered inclusively, the different sub-samples have a similar shape. In order to provide the fit with a handle on the different flavour 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 **??** and in the zero, one, and two b-tagged regions **??**. The latter distributions show the different contributions from the V+jets components, with $Z + l$ dominating in the 0 b-tagged region, $Z + bl + Zcl$ in the 1 b-tagged region, and $Z + bb$, $Z + bc$, $Z + cc$, $Z + l$, and $t\bar{t}$ in the 2 b-tagged

[4]The 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.

region. The corresponding normalization uncertainties were left unconstrained in the fit. The jet mass distribution used for the fit is also shown to gauge the presence of residual shape mis-modeling. The uncertainties in the plots are given by the combined statistical uncertainties and the fit uncertainties for the normalization coefficients. The only noticeable mis-modelling appears in the resolved $p(h)$ score in Fig. 7.37h. However, part of the mis-modelling is likely due to the residual shape mis-modelling in the jet mass distribution 7.37g. 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.

Merged MCT	Resolved MCT
$m^J \in [50, 200]$ GeV	$p_T^{j1} > 45$ GeV and $\lvert\eta^{j1}\rvert < 2.5$
$\lvert\eta^J\rvert < 2$	$p_T^{j2} > 20$ GeV and $\lvert\eta^{j2}\rvert < 2.5$
$p_T^J \in [200, 3500]$ GeV	$p_T^{j1} + p_T^{j2} < 500$ GeV

Table 7.15: MCT training cuts applied to the pre-selection region definition.

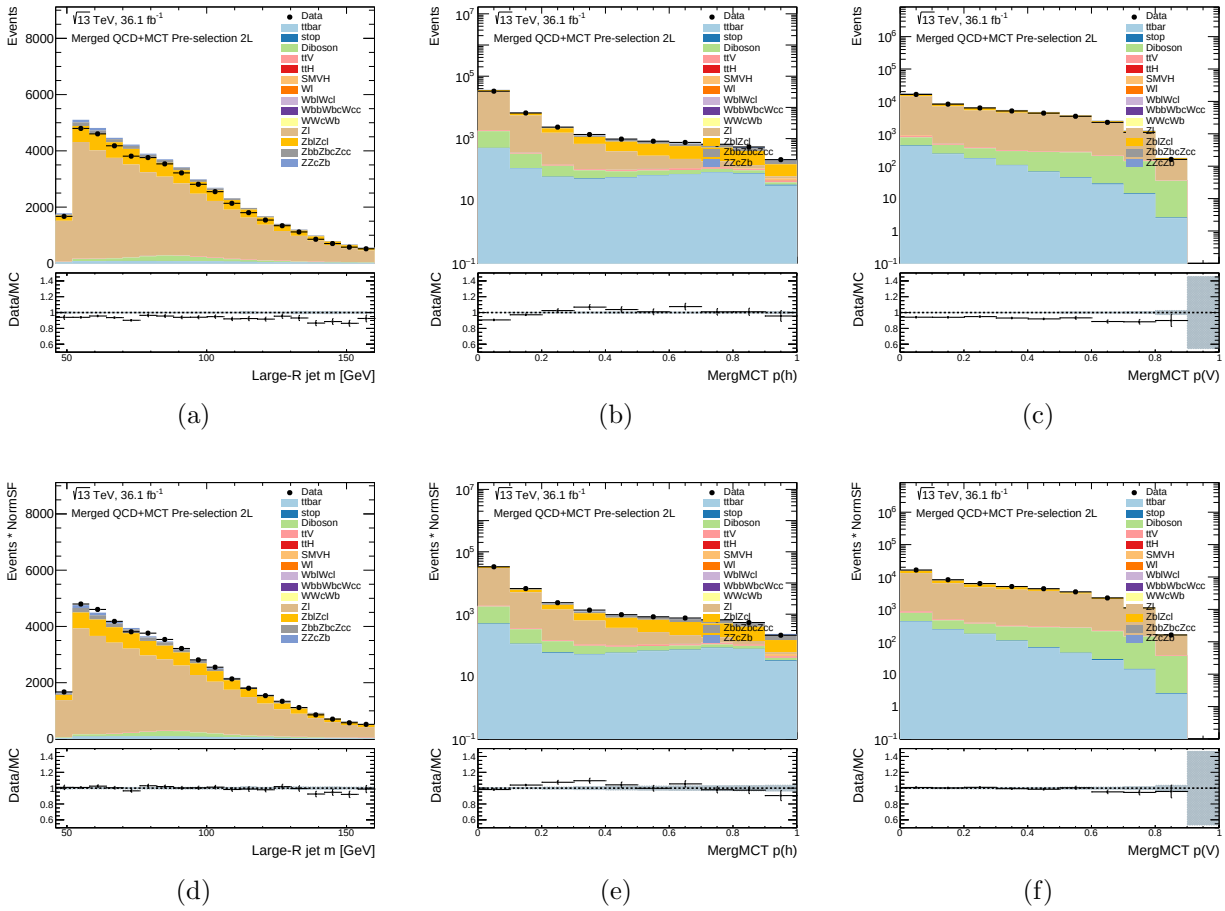
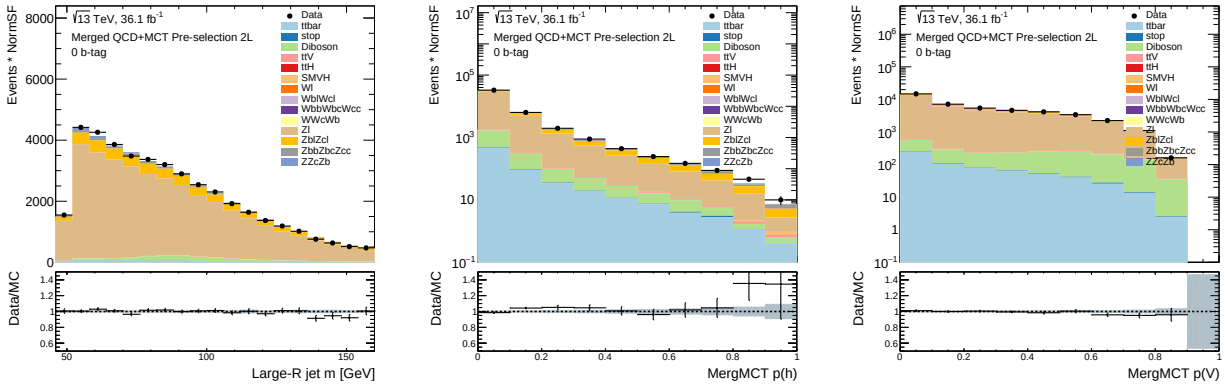
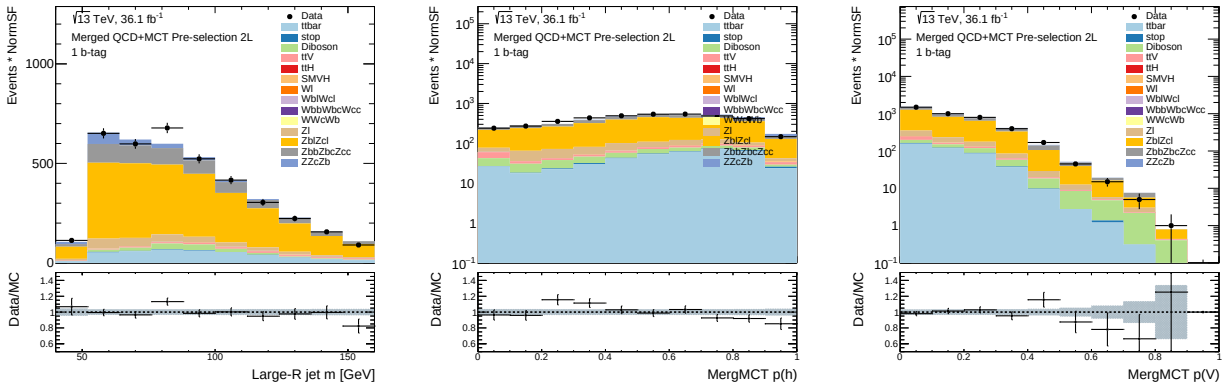


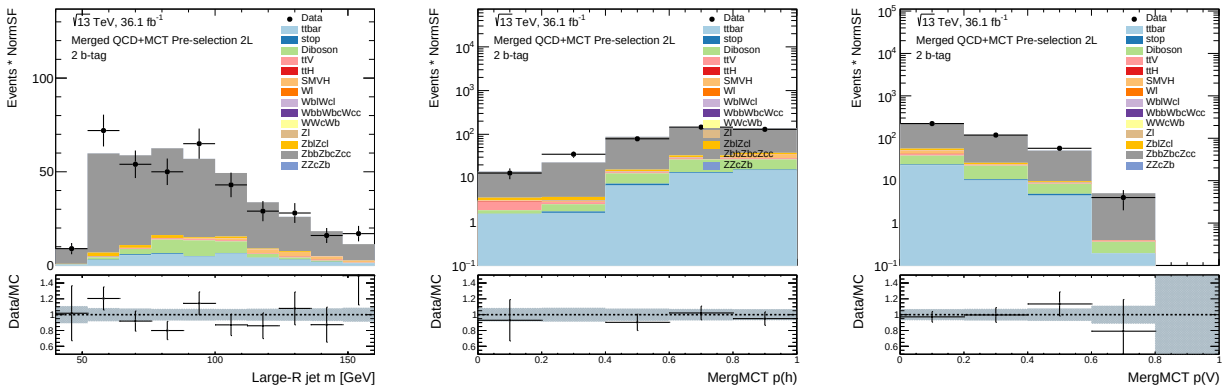
Figure 7.36: Data and MC comparison in inclusive merged pre-selection region in the 2-lepton channel of the large-R jet mass and the raw merged MCT scores p(h) and p(V), shown before (top) and after (bottom) applying the normalization SFs.



(g) (h) (i)



(j) (k) (l)



(m) (n) (o)

Figure 7.36: Data and MC comparison in merged pre-selection n b-tagged region in the 2-lepton channel of the large-R jet mass and the raw merged MCT scores p(h) and p(V), after applying the normalization SFs in the 0 b-tag (top), 1 b-tag (middle), and 2 b-tag (bottom) regions.

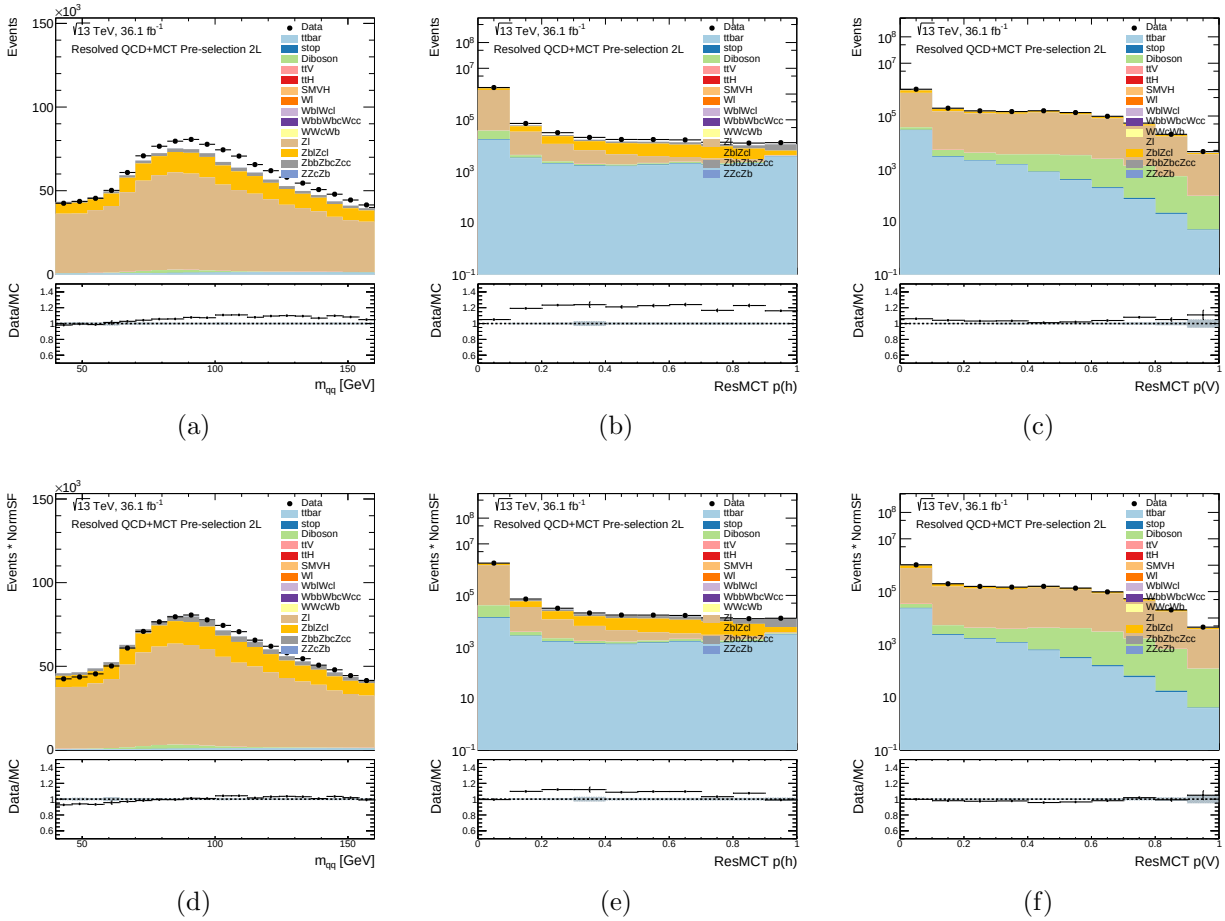


Figure 7.37: Data and MC comparison in inclusive merged pre-selection region in the 2-lepton channel of the large-R jet mass and the raw merged MCT scores p(h) and p(V), shown before (top) and after (bottom) applying the normalization SFs.

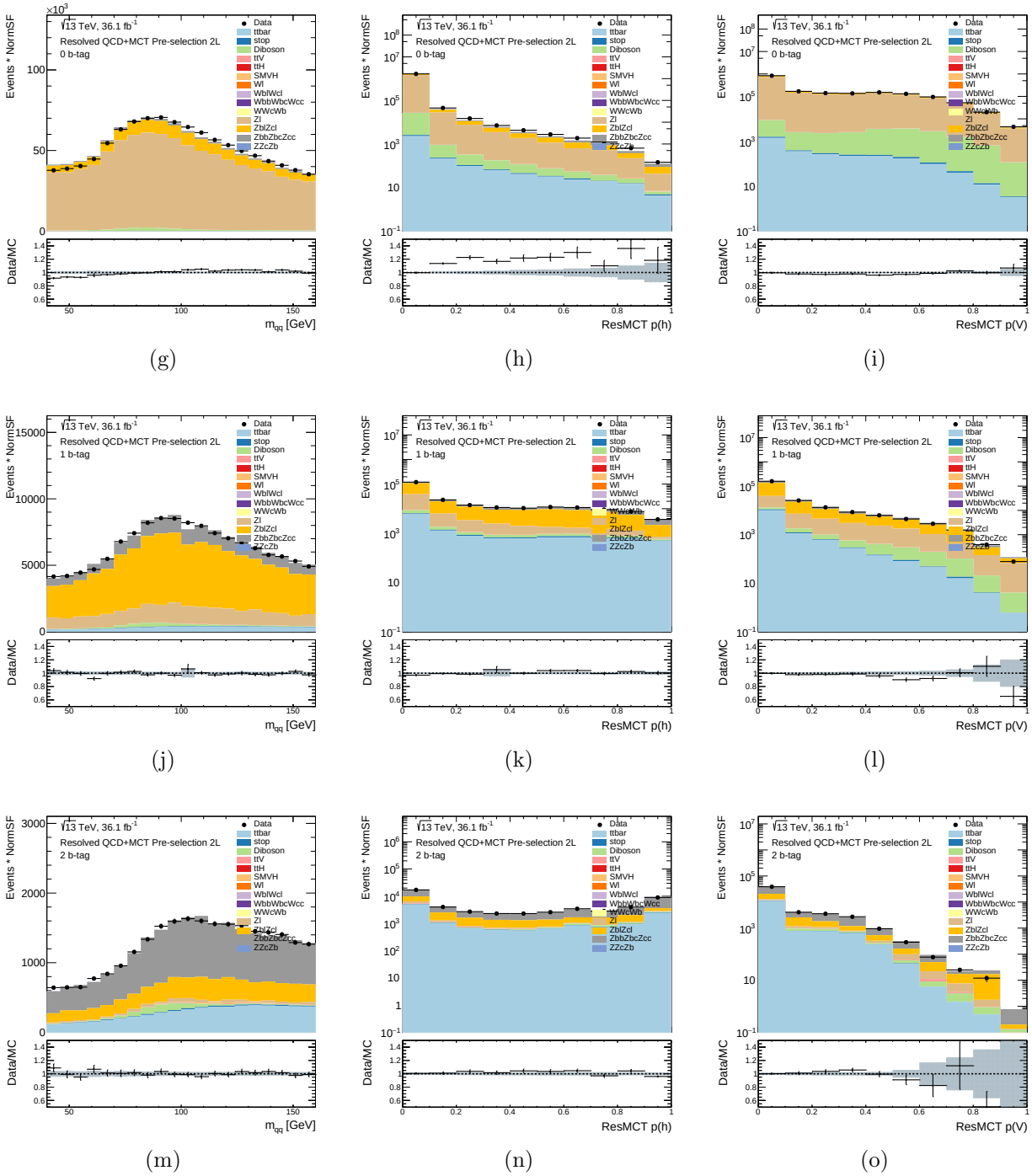
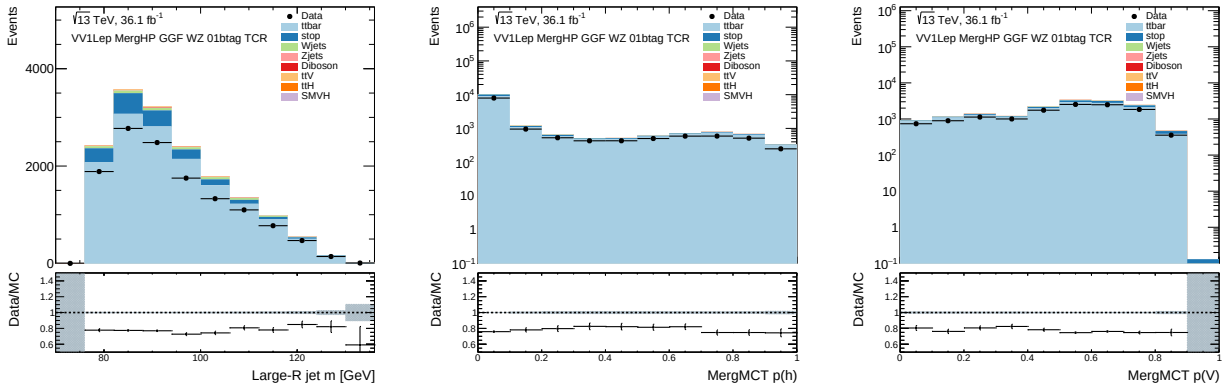


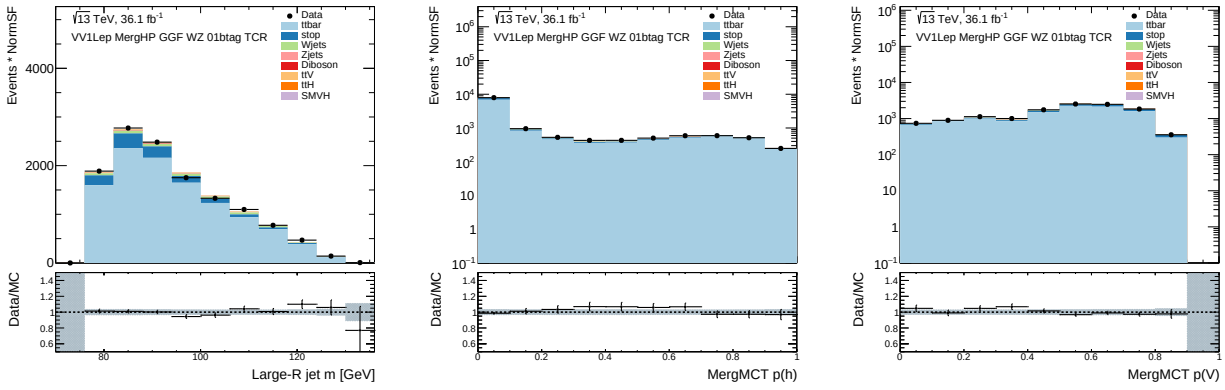
Figure 7.37: Data and MC comparison in resolved pre-selection n b-tagged region in the 2-lepton channel of the di-jet jet mass and the raw resolved MCT scores p(h) and p(V), after applying the normalization SFs in the 0 b-tag (top), 1 b-tag (middle), and 2 b-tag (bottom) regions.

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_01btag_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 flavour components, as the 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. 7.38. As observed in the pre-selection regions, after applying the normalization SFs no differences are observed between data and MC.



(a) (b) (c)



(d) (e) (f)

Figure 7.38: Data and MC comparison in top-enriched CR `VV1Lep_MergHP_GGF_WZ_01btag_TCR`. The large-R jet mass and the merged MCTscores 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 studied 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 mis-modelling was simulated as $\pm 10\%$ 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 the scores are then 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 between the other scores.

Tab. 7.16 shows the migration of events after the MCT sorting in the nominal, up, and down variation scenarios. Fig. **??** shows the inclusive signal efficiencies as a function of p_T, similarly to the studies presented in Sec. 7.10.4, for the Nominal, up and down variation. 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.

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.

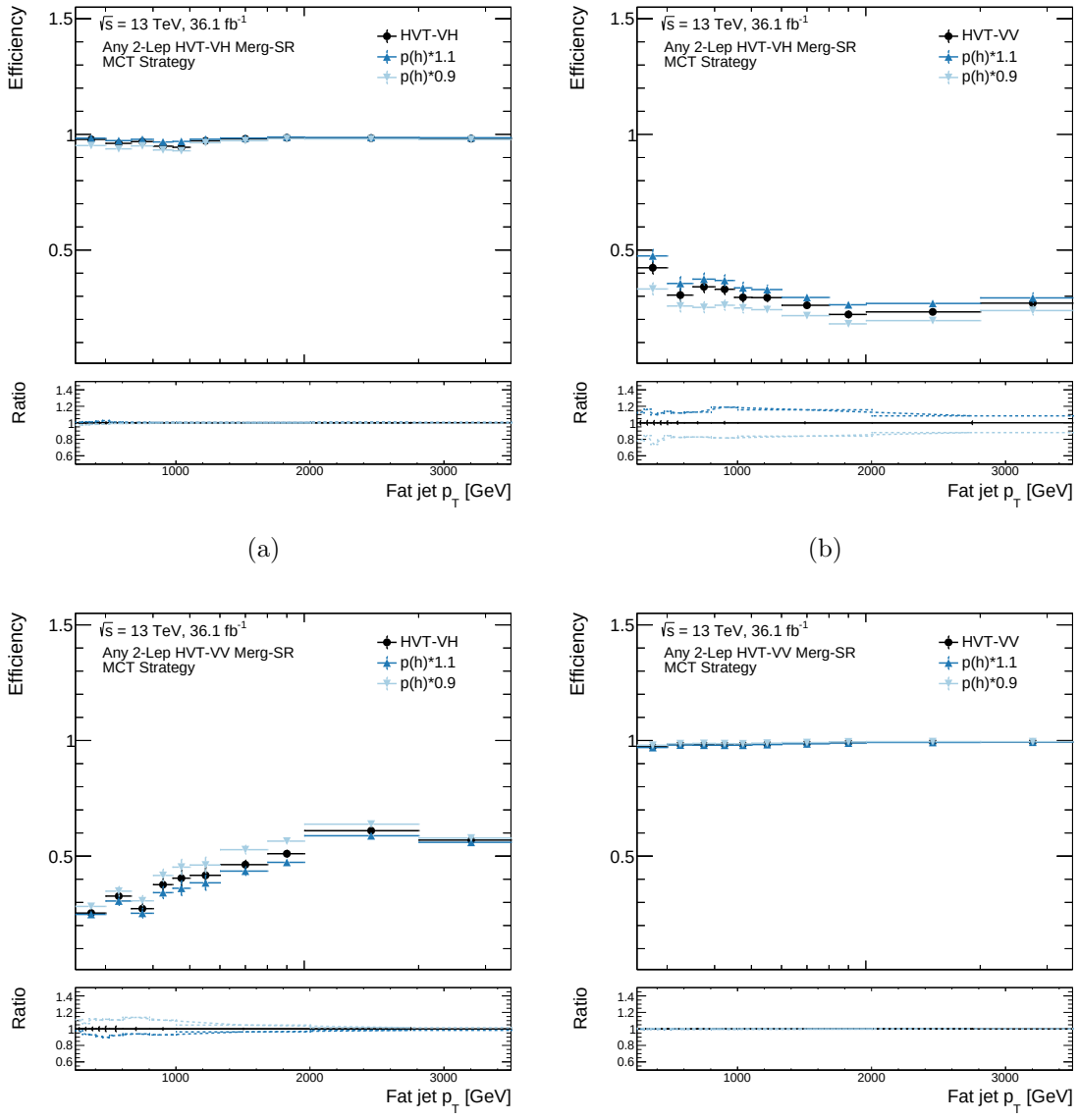


Figure 7.39: Signal efficiency as a function of p_T in the 2-lepton channel for VV and VH merged signal regions, after orthogonalization with the MCT strategy. The nominal result is compared with the effect of a ±10% variation on the resolved and merged p(h) scores propagated to p(V) as p(V') = p(V) + (p(h) - p(h')).

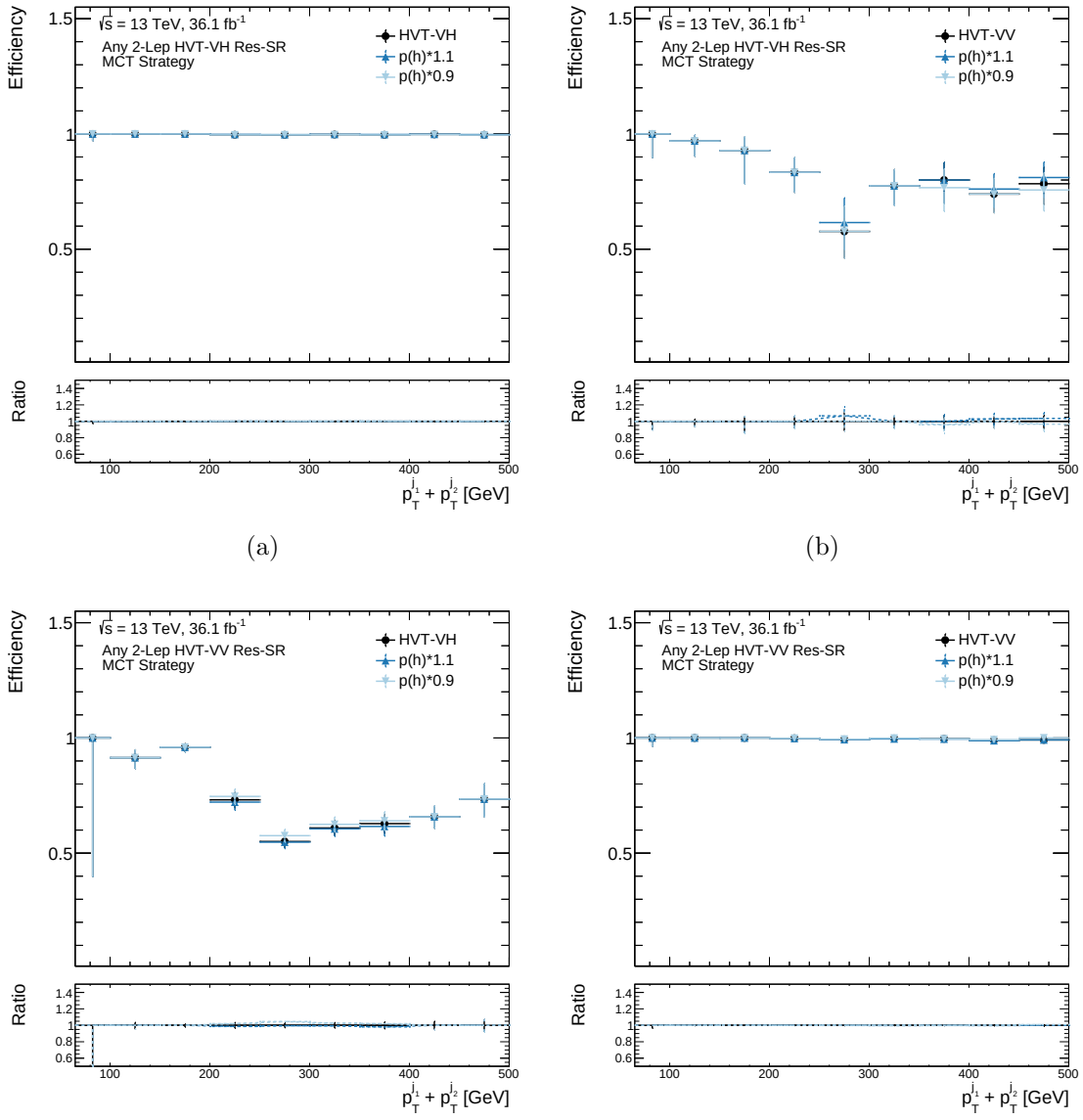


Figure 7.40: Signal efficiency as a function of p_T in the 2-lepton channel VV and VH resolved signal regions, after orthogonalization with the MCT strategy. The nominal result is compared with the effect of a ±10% variation on the resolved and merged p(h) scores. The variation is propagated to p(V) as p(V') = p(V) + (p(h) - p(h')).

Chapter 8

The trigger upgrade for the HL-LHC

The Global Trigger (GT) will be a major addition to the ATLAS trigger system to be installed during the Phase II upgrades in preparation for the High-Luminosity LHC (HL-LHC)[1]. The GT upgrade is necessary in order to benefit from the increase in luminosity and keep ATLAS physics goals. 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 will provide 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 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.

[1] (See Sec. 4.3.5) for an overview of the Phase II Trigger & Data acquisition (TDAQ) upgrade

Most of the work within the GT upgrade project goes into the software and firmware co-development of the new trigger algorithms. A large contribution of this thesis was the development of the software simulation framework for firmware algorithm development for the GT. In addition, one of the main objectives of this work was the development of a new jet reconstruction and jet triggering strategy for the GT.

Every trigger is designed to target a specific physics signature. 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 online small-R jet has a p_T above a given threshold. Some signatures are characterized by a higher jet multiplicity, which usually decreases the average p_T of the final state jets. A multi-jet trigger requires therefore the presence of three or four jets above a p_T threshold that is lower than the $p_T$requirement of a one-jet trigger. From the point of view of the analysis, the best case scenario would be if all events of the desired topology were recorded. In practice, of the forty million events that the L0 trigger has to analyze per second, only 100 thousands of them can be accepted. This 100 kHz event rate has to be shared between the different L0 triggers according to ATLAS physics goals and priorities. The number of events that a jet trigger can accept is thus determined by the fraction of the 100 kHz event rate allocated to it. The online pt threshold is found such that the number of events that pass the p_T cut is equal to the allowed rate threshold.

Recall that the majority of the events seen by the detector are minimum bias events, while interesting collisions are orders of magnitude rarer. An additional difficulty then lies in not cutting away the rare signal events at the same time as one removes background. The development of a trigger algorithm consists precisely in optimizing this trade-off between signal efficiency and background rejection. The more events can be discarded as being

background at the moment of the trigger selection, the lower the p_T threshold can be, the higher the signal efficiency of the offline analysis. Note that this is one of the fundamental motivations behind the Global Trigger upgrade. With the current trigger algorithms, the larger rates of the HL-LHC would have pushed the trigger thresholds up, decreasing the signal efficiency and the physics reach of analyses in the lower $p_T$regimes. The development of more complex algorithms able to perform a more sophisticated signal to background discrimination at L0 is necessary to keep the rates and the online pt thresholds low enough for physics analyses to have access to important events.

A brief introduction to the GT is given in Sec. 8.1. The remaining sections of this chapter will describe the contributions to the GT project performed in this thesis, including the development of the software simulation framework for trigger algorithms (Sec. **??**), the design and validation of a new jet reconstruction strategy in Sec. 8.3, and studies for a novel pileup-jet suppression method using deep neural networks to improve multi-jet triggers in Sec. 8.4.

8.1 The Global Trigger

The Global Trigger (GT) will be a new addition to the L0 hardware-based trigger for the HL-LHC. The task of the GT is to reduce the event rate from 40 MHz down to 1 MHz. In order to achieve this, the GT will bring Event-Filter like capabilities at L0 by deploying complex offline-like algorithms on fast FPGA hardware.

The GT 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 FEXs, 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 fully independent and decoupled from the BC rate and can be processed in parallel in the 48 GEP nodes, allowing the implementation of asynchronous complex algorithms. The event data from each BC is distributed to the GEP nodes in a round-dobin fashion, with each node receiving new data every 48 BCs, increasing the latency between the arrival of two events from 25 ns to 1.2μs. This process is shown in Fig. 8.1.

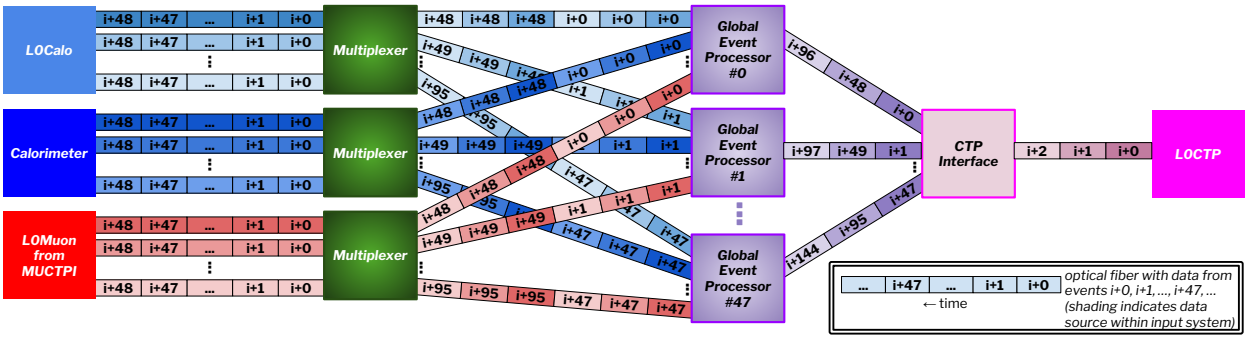


Figure 8.1: Multiplexing of incoming serial data at 40 MHz by the MUX nodes. Each MUX receives data for every bunch crossing. The data is processed, organized, and dispatched to the GEP nodes. Each GEP node receives the complete event data and analyzes the data asynchronously. Results from the GEP nodes 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 shown in Fig 8.2. As the data is being received, regional algorithms can already start processing before iterative algorithms that require the full event data can start, making full use of the data transmission time. Ordering data geometrically also allows the

pipelining of the steps of non-iterative algorithms on the FPGA board. For example, if the data arrives ordered in η, any regional algorithm can start processing the detector plane in full slices over ϕ. Once all the event data is available, the next step of the algorithm can start right away. This pipelined processing design of individual algorithms, together with the parallel execution of different algorithms, allows to drastically reduce the FPGA resource utilization and extend the L0 latency up to $\sim 6\mu$s.

The output of a GEP node is the Trigger Input (TIP), containing flag bits of which trigger requirements have been satisfied and possibly multiplicities of identified objects. A Global-to-CTP Interface (gCTPi) demultiplexes the data, re-builds the event with the correct ID 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 Global Trigger and MUCTPI, as well as from the forward detectors and other detector calibration sub-systems, and makes the final L0 trigger decision. It also applies deadtime and prescales. The Level-0 accept (L0A) rate is of 1 MHz.

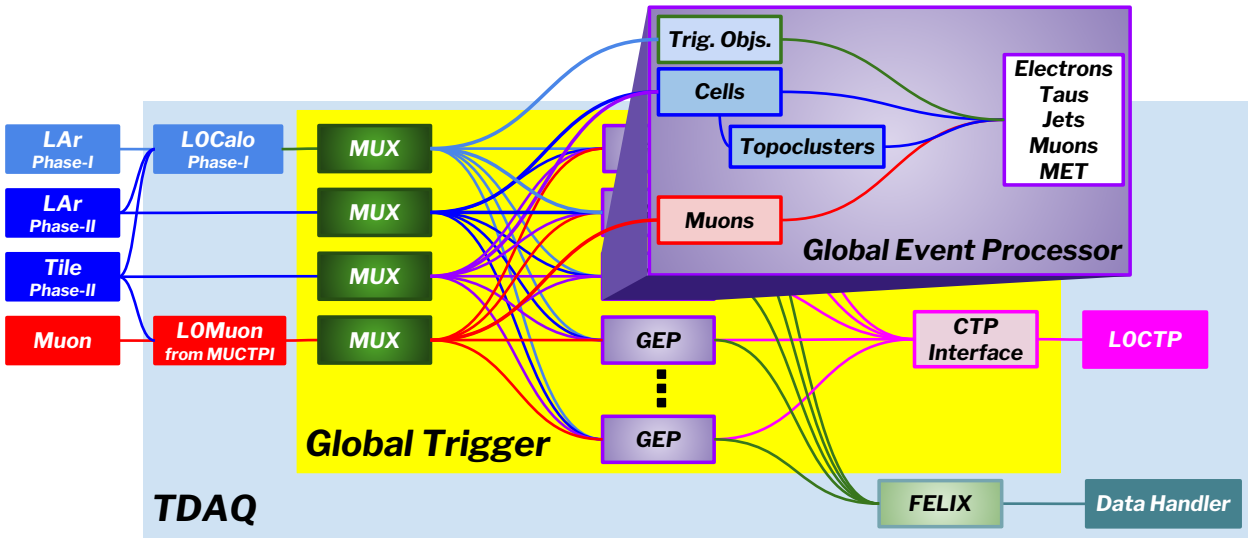


Figure 8.2: Schematic view of the Global Trigger processing. [10].

Inputs and processing

Every LHC bunch crossing the GT receives the full granularity noise-suppressed ($|E_T| > 2\sigma/$) calorimeter data from the LAr and Tile calorimeter front-end electronics, where the noise threshold on the cells is determined by the bandwidth limitations of the hardware. Access to the cells information will allow to run for the first time topological clustering at L0. Global will also receive TOBs from the L0Calo eFEX, jFEX, and gFEX processors, which can be used as seeds for GEP algorithms. Lastly, Global will receive TOBs from the MUCTPI. These are muon-track candidates defined by p_T, η, and ϕ information.

The LAr and Tile inputs are expected to arrive first in $1.4-1.7\mu$s. The data from L0Calo is expected to arrive in $2-2.6\mu$s, followed last by the L0Muon inputs at $4.5-5.3\mu$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. A significant contributor to the GT latency is the data transfer time from a MUX to a GEP node, of 1.25 μs.

Hardware

The hardware is based on a common design called the Global Common Module (GCM) to minimize the complexity of the firmware. A GCM with dedicated firmware will be used for each of the three main components of the Global Trigger: a multiplexing module, an Global event processing module, and a demultiplexing module. Each node consists of an FPGA with substantial input and output capabilities to handle the large bandwidths and buffering. The GCM design is an ATCA Front Board that 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 a Versal Premium VP1802.

Firmware

The various algorithms that compute the trigger objects are executed on the common GEP hardware modules, which will represent the majority of the resource consumption. 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, 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 FPGA GEP module is shown in Fig. 8.3. The FPGA is divided in four Super Logic Regions (SLR) pipelined in latency intervals of 1.2μs. The first SLR0 receives LAr data, so that regional algorithms, such as topological clustering, can start processing. After 1.2 μs the processing 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 in time with algorithm development. From this picture it is clear that the GT is primarily a firmware project.

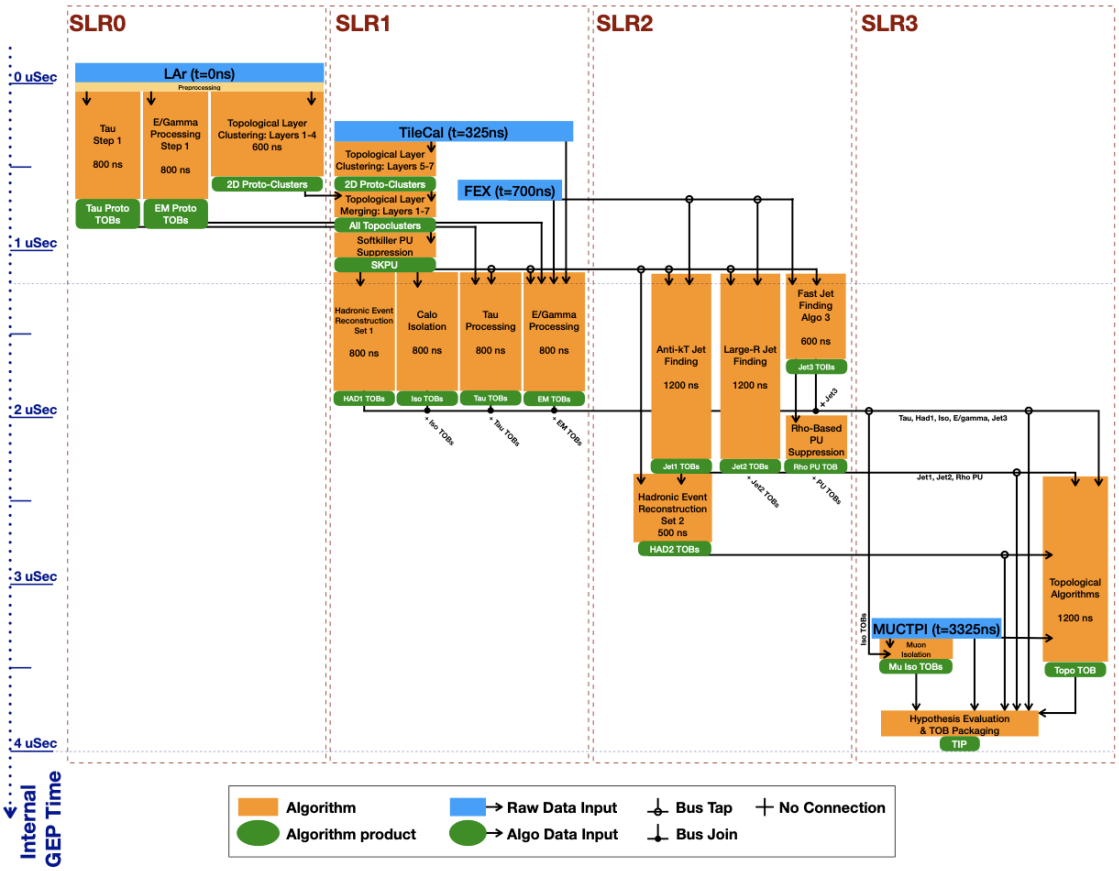


Figure 8.3: Preliminary plan of Global Trigger algorithm dataflow on one FPGA corresponding to a GEP module. The FPGA is divided in four Super Logic Regions (SLRs) for pipelined data processing. This picture might evolve with time.

8.2 Trigger algorithm development

Global is primarily a firmware project, where the trigger algorithms that analyze the event data and produce the trigger objects are implemented in reprogrammable firmware. From the hardware side, they are constrained by the I/O ports, latency, and bandwidth requirements set by the Global Common Module. From the physics side, each algorithm is generally 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 and stay 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 high-level packages, such as High-Level-Synthesis (HLS). Often this requires several steps of optimization in order to stay within resource allocations. In particular, if the resource usage significantly exceeds the allocated lot, it is possible that significant changes to the algorithm itself are needed, so that the software and firmware co-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 implementation is finalized. In the context of trigger performance studies there are two levels of reconstruction, which will be referred to as *online* and *offline*. 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 produces 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 have very few limitations in terms of CPU time and are therefore maximally optimized. The objects output by these two stages will be referred to correspondingly as *online* and *offline* objects.

8.2.1 The Global Trigger software simulation framework

Part of this thesis work was to develop the software simulation framework for the Global Trigger. The package was written in $C++$ and Python as part of the ATLAS offline software

Athena [90]. A sketch of the functionalities of the framework is shown in Fig. 8.4.

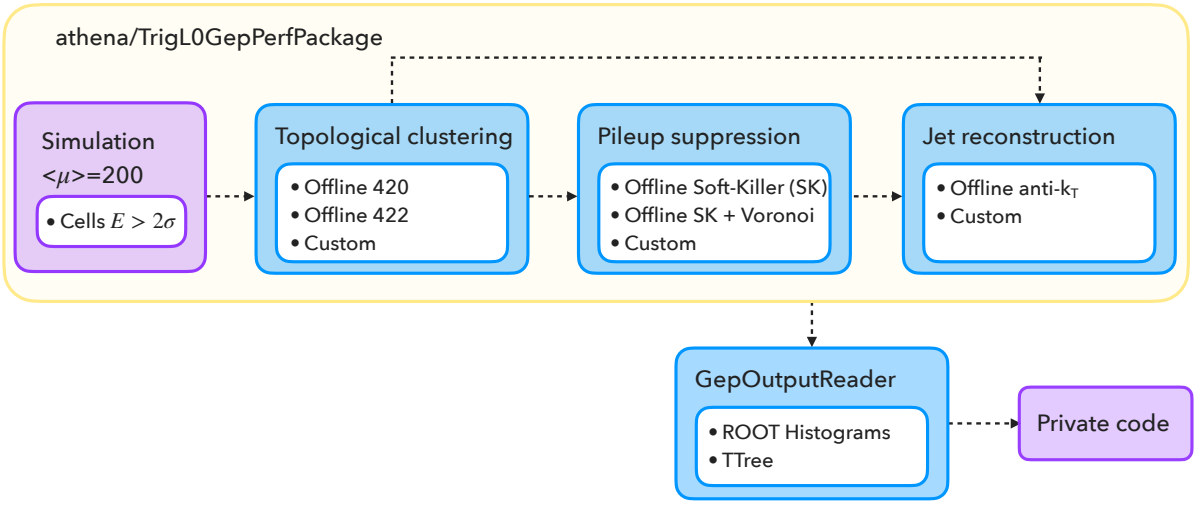


Figure 8.4: Diagram of the software simulation framework for the Global Trigger.

The framework was designed to provide an efficient way to develop and study new candidate algorithms for the GT, as well as to provide L0 objects as input to physics performance studies of downstream trigger algorithms. The first point was particularly important at the time of this work, as all algorithms were under development. The framework was therefore developed to be a skeleton where algorithms could be easily added for testing. The label "custom" in the diagram represents a possible plug-in for developers to interface their C++ code with the Athena workflow. The ATLAS offline versions of the algorithms are also always provided for a solid baseline comparison. As most algorithms require topocluster or jet information, the focus was put on the 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. The MC samples read by the framework are produced with the standard ATLAS MC production path, but with HL-LHC settings. The samples are simulated at $\sqrt{s} = 14\,\text{TeV}$, including the simulation of the new detector

components, and with the number of simultaneous pp interactions per bunch crossing set to $< \mu >= 200$. This represents the extreme pileup scenario that the trigger is expected to handle. Di-jet events are generated using PYTHIA8 [**?**], with the NNPDF23LO set and A14 ATLAS parameter tuning, to simulate the multi-jet background. These are 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 is used as representative of a minimum bias sample. For single-jet trigger studies, a simulation of fully-hadronic $t\bar{t}$ events is used. The sample was generated using Powheg [**? ? ?**] interfaced with Pythia with the NNPDF23LO set and the A14 tune for the parton shower. For multi-jet trigger studies, samples with two Higgs bosons, each decaying to a pair of b-quarks were produced. The di-Higgs pair is produced via gluon-gluon fusion (ggF) assuming the SM trilinear coupling. The matrix element is calculated using Poweheg at NLO including finite top mass loop calculations. The showering is performed using PYTHIA8. The primary input objects are calorimeter cells. The cells are fed as input to the topological clustering algorithm. The standard ATLAS offline topoclustering algorithm (see Sec. 4.4.5) provides topocluster collections at the EM scale built with both the 4-2-0 (Calo420) and the 4-2-2 (Calo422) setting. 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, Global will only receive cells with $E > 2\sigma$. Therefore, the Calo422 offline collection represents the "best case scenario" for topoclustering at L0. As the online topoclustering algorithm for the GT is still under development, only the Calo422 collection was used as the GT representative topocluster collection.

The topoclusters can then 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 (VorSK). The modified topocluster collection can then be fed to downstream algorithms, such as jet reconstruction. In some ATLAS internal studies it was shown that running SK on the Voronoi-suppressed topoclusters provided the best performance, while running Voronoi alone should be avoided. For this reason, in the following studies we will mainly discuss the SK and VorSK options. As no online version of these algorithms was available at the time of these studies, only the offline versions will be used in the following.

Lastly, the chosen jet reconstruction algorithm is run on the input topocluster collection. The offline anti-k$_t$algorithm is always provided as a reference, 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 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 allows to lower the $p_T$threshold. In the following, we refer to the cut on the online jet p_T as p_T^{cut}

The performance of a jet trigger algorithm is studied with MC simulations in terms of the offline signal efficiency, as shown in Fig. 8.5. The efficiency is generally displayed as a function of the offline-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

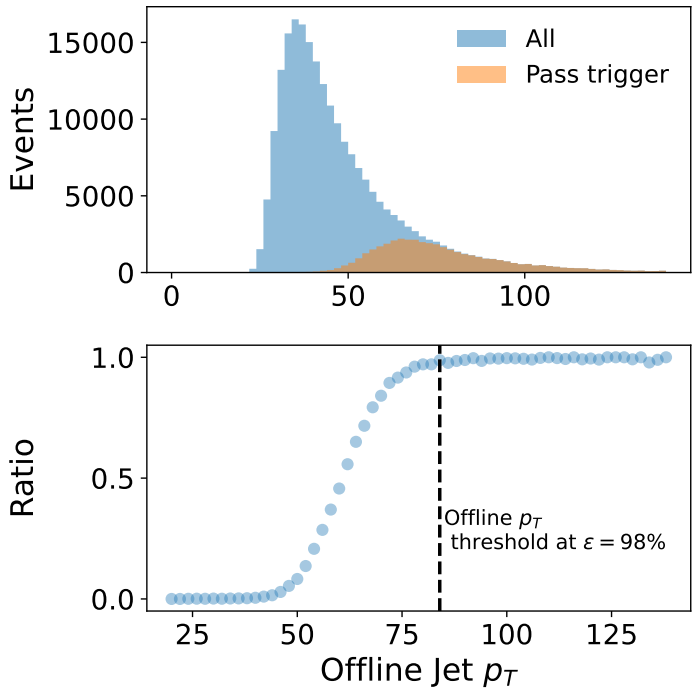


Figure 8.5: Example of trigger efficiency curve (see text for explanation).

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 100% efficiency point (the dashed line in the figure). 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 when the turn-on curve reaches 100% efficiency as p_T^{thresh}. In practice, the lower the p_T^{thresh}, the broader the $p_T$range the analysis will be sensitive to. 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 relationship between online and offline reconstruction.

Recall that offline reconstruction represents the best one can do when virtually no latency or resources limitations are present. Then one can 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 Θ 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% for values above. Mistakes in online reconstruction result in a deviation from the scenario $p_T^{\text{thresh}} = p_T^{\text{cut}}$. One can then identify two ways in which online jet reconstruction can go wrong:

1. One can reconstruct the right jet, but with the wrong energy

2. One can reconstruct the wrong jet

Sometimes the online and offline jets represent the same energy deposition in the detector (they have the same (η, ϕ) coordinates), but they have different transverse momenta. This can happen because offline and online reconstruction use different inputs and generally different algorithms. For instance, while offline small-R jets are produced using the Calo420 collection, online jets in the GT are produced using the Calo422 topoclusters, which are lower in energy, as they are built from fewer cells. Therefore, one can expect online jets to have lower transverse momenta than offline jets.

If all online jets transverse momenta differ by a constant factor, this is just a matter of normalization, that does not affect the value of p_T^{thresh}. However, issues can arise when the factor is not constant, but dependent on the phase space. As an example, consider the case of two nearby offline 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 low energy jets might see their energy increase. Without an energy overlap removal strategy, the online jets would see a fictitious increase only in high energy jets induced by these cases of nearby cones. This would cause an increase only in the high energy tail of the cumulative distribution that would raise the online p_T^{cut} *and* the offline p_T^{thresh}. It might also produce an over-efficiency, as online jets that should have not passed the trigger will pass.

Other times the reconstructed online and offline n^{th} leading jets correspond to actually different objects. This can happen quite often 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 built 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 pt cut is really being applied to the leading jet. Consider an event where the offline third leading jet has $p_T < p_T^{\text{thresh}}$. The expectation is that the event be discarded in order to save space for events that are more likely to contain interesting physics. However, in this scenario the event will pass the third jet trigger, producing a non-zero efficiency at a p_T value below p_T^{thresh}.

Assume instead that seeds are required to not be closer than $dR = 0.5$. Consider then a signal topology where two pairs of nearby jets are created and where, for some events, at least one pair is closer than $dR = 0.5$. While the offline jet reconstruction will find four jets, the online algorithm will find only two hard jets and the 3rd and 4th online jets will either be soft or missing. In either case, the event will not pass the three- and four-jet trigger. For signal samples characterized by this topology, this will result in a plateau efficiency below 100% across the $p_T$spectrum.

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, whether this association between online and offline objects is correct or not does not affect the performance of the trigger, as long as the event passes the selection anyway. This is why no truth-matching is performed when looking at

the signal efficiency curves. However, these mis-matches can cause inefficiencies, ultimately resulting 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.

Clearly, the performance of a trigger algorithm relies on the jet reconstruction strategy, including the jet reconstruction algorithm and possible forms of pileup suppression. The development and optimization of a trigger algorithm has therefore two goals:

1. Improve the jet energy resolution of the online vs. offline jet reconstruction, i.e. make the turn on as steep as possible.

2. Reduce the rate of high energy background jets, while keeping the signal efficiency high, i.e. reduce p_T^{thresh} as much as possible.

The first can be studied by comparing trigger efficiency curves at a fixed trigger rate. A maximum minimum bias trigger rate is assumed to be allocated, the online p_T cut is found that would remain below the given rate threshold, and the trigger efficiency is built. This method can be used to gauge the effect of pileup removal strategies, including Et cuts on jet constituents. The second effect can be studied by overlaying the turn-on curves from the different algorithms adjusting the online p_T threshold so that the efficiency plateaus are aligned at a given offline p_T threshold. This decouples the problem from the choice of online p_T cut, and allows to study only the jet energy resolution. The best algorithm 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.

8.3 A cone jet reconstruction algorithm

While offline anti-k$_t$(see Sec. 5.2.3)is the standard for jet reconstruction, it is also a computationally intensive algorithm, as for each iteration one has to calculate dR distances and take $1/p_T^2$ divisions. Additionally, anti-k$_t$is a highly iterative algorithm, with a non deterministic number of operations. This makes it non-scalable on parallelizable firmware, removing one of the main advantages of fast FPGA hardware. This makes the anti-k$_t$algorithm prohibitively expensive in terms of the resource share on the Global Event Processor. The goal of this study was therefore to find an alternative to anti-k$_t$that would allow to keep the performance high enough for the goals of the L0 trigger, 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 L0 trigger, IRC safety is not a requirement, as the goal of the L0 trigger is to accept the right events, not to reconstruct physics-analysis-ready objects. The objects themselves are reconstructed with the actual offline algorithms at the event filter (EF) level and again offline for use in analyses. The ability of identifying 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. 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 development of this jet reconstruction strategy will be discussed in the following. The new strategy aimed at using a simpler jet reconstruction algorithm that would produce coarser jets in exchange for fewer resources, and planned to subsequently refine these jets using deep learning. Hence,

the goal was to trade the benefits of a highly iterative algorithm, with those of a highly parallelizable neural network. However, as it will be shown, a thorough optimization of the cone jet algorithm allowed to attain an equivalent trigger performance to that of offline anti-k_t. The study using deep learning to potentially improve the multi-jet trigger performance will be discussed in the next Sec. 8.4.

These studies 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 could be envisioned, including using the leading small-R cones as seeds or as input to a reclustering algorithm. However, these studies go 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 4-2-0 topoclusters at the EM scale. These will be referred to as `AntiKt420`. The inputs to the jet reconstruction algorithm in Global will be 4-2-2 topoclusters (Calo422). 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 be 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 $p_T > 10\,\text{GeV}$.

The performance of the algorithm was benchmarked against target signal simulations. The $Z' \to t\bar{t}$ 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, with the 3rd and 4th leading jet close to the 1st or 2nd leading jet. The $hh \to b\bar{b}b\bar{b}$ sample with $\kappa_\lambda = 1$ was used as a representative sample for a type of analysis 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 [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 candidate algorithm will be discussed, while in Sec. 8.3.2 the performance of the final optimized cone algorithm will be benchmarked against online `AntiKt422` jets. Offline pileup suppression schemes will also be tested. Lastly, Sec. **??** will discuss the results of a preliminary firmware simulation.

8.3.1 Development

Recall that a jet definition is defined by the jet algorithm, the jet inputs, and the recombination strategy.

Inputs

As previously stated, the inputs to jet reconstruction in Global will be 4-2-2 topoclusters, which for these studies will be the Calo422 collection. Pileup suppression is expected to be applied before the topoclusters are fed to the jet reconstruction. However, the jet algorithm development was performed without pileup suppression applied, as no online version of the algorithms were available. This also allowed to reduce the number of competing factors in the optimization process. Robustness against offline pileup suppression schemes will be shown for the optimized algorithm.

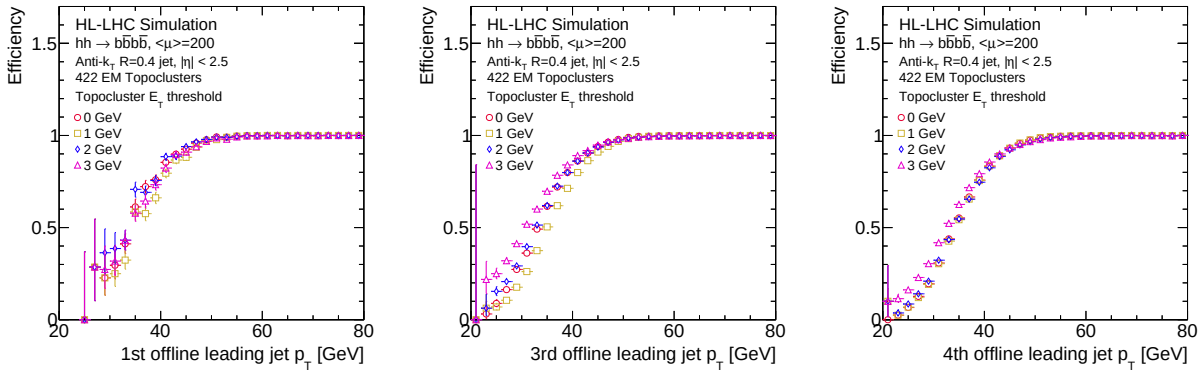
The maximum allowed number of inputs to be processed has to be fixed in firmware, and the smallest this is, the fewer the footprint of the algorithm on the FPGA. In particular,

without any E_T cut on the topoclusters, the number of topoclusters reconstructed per event peaks around 700, a prohibitively large number. This is shown in Fig. 8.8 for the minimum bias and the signal samples. It was therefore necessary to threshold the input topoclusters. The E_T cuts at 1, 2, and 3 GeV were tried, motivated by the range of the Soft-Killer p_T cut simulated for $< \mu >= 200$ simulations in Ref.**??**. The effect of the different cuts was studied using the `AntiKt422` online jet collection in different energy regimes. Fig. 8.6 and Fig. 8.7 show the trigger efficiencies for di-Higgs signal samples 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. This indicates 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 of using the `Calo422` instead of the `Calo420` collection.

The conclusion from these studies was to exclude the 3 GeV 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.

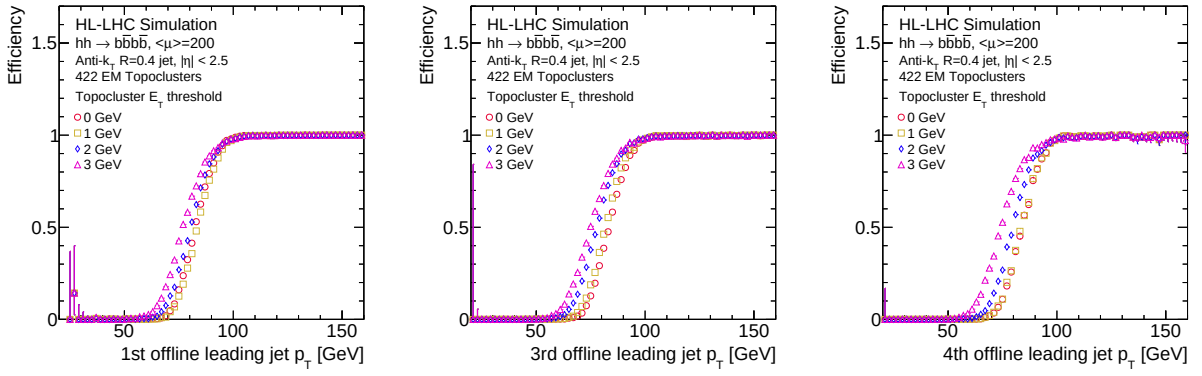
Recombination scheme

The standard 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



(a) (b) (c)

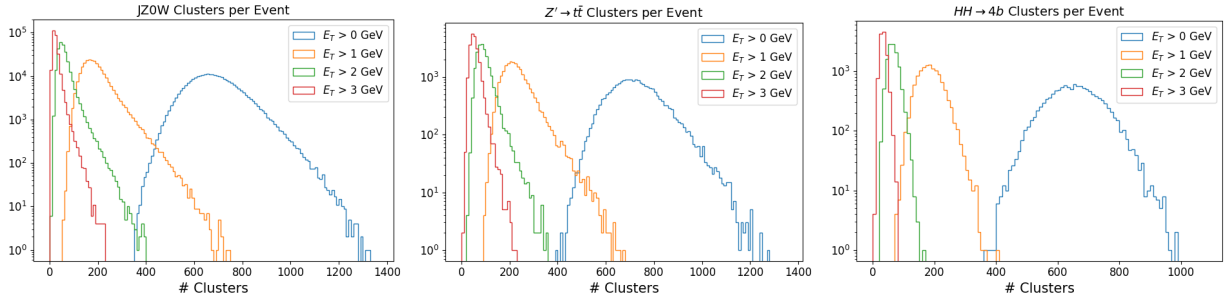
Figure 8.6: Trigger efficiencies using 1, 2, and 3 GeV input topoclusters E_T thresholds and no threshold. Turn-on curves are built with a fixed offline p_T threshold of 50 GeV using di-Higgs signal sample and the `AntiKt422` online jet collection.



(a) (b) (c)

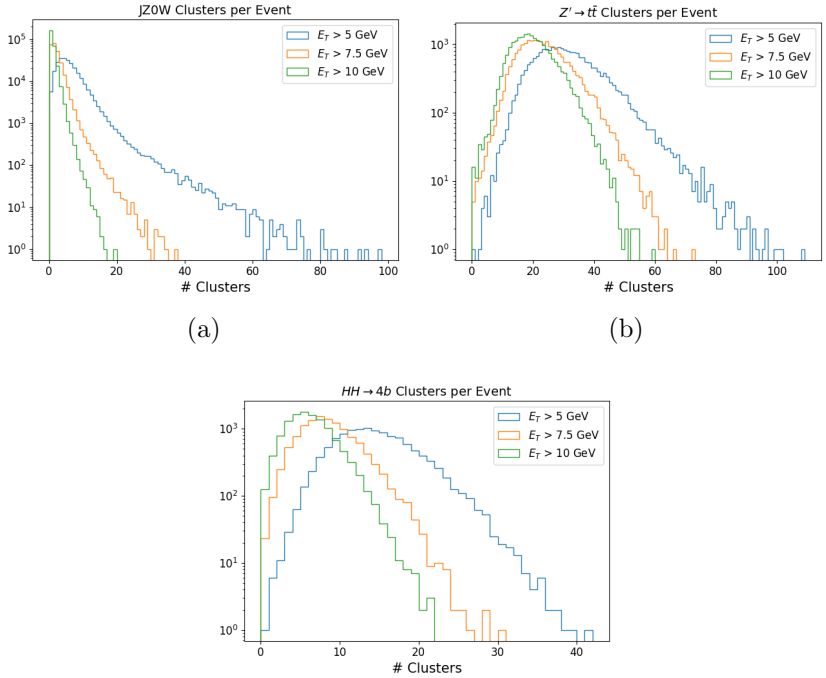
Figure 8.7: Trigger efficiencies using 1, 2, and 3 GeV input topoclusters E_T thresholds and no threshold. Turn-on curves are built with a fixed offline p_T threshold of 100 GeV using di-Higgs signal sample and the `AntiKt422` online jet collection.

was observed to significantly increase the FPGA resource utilization beyond acceptable limits. A new scheme was therefore created, 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 constituent, the (η, ϕ) coordinates are given by the coordinates of



(a) (b) (c)

Figure 8.8: Number of topoclusters per event with E_T thresholding shown, from left to right, for minimum bias, Z' and di-Higgs samples.



(a) (b)

(c)

Figure 8.9: Number of topoclusters per event after E_T thresholding to select seed collection. This is shown, from left to right, for the minimum bias, $Z' \to t\bar{t}$, and di-Higgs samples. (Plots from Garrit)

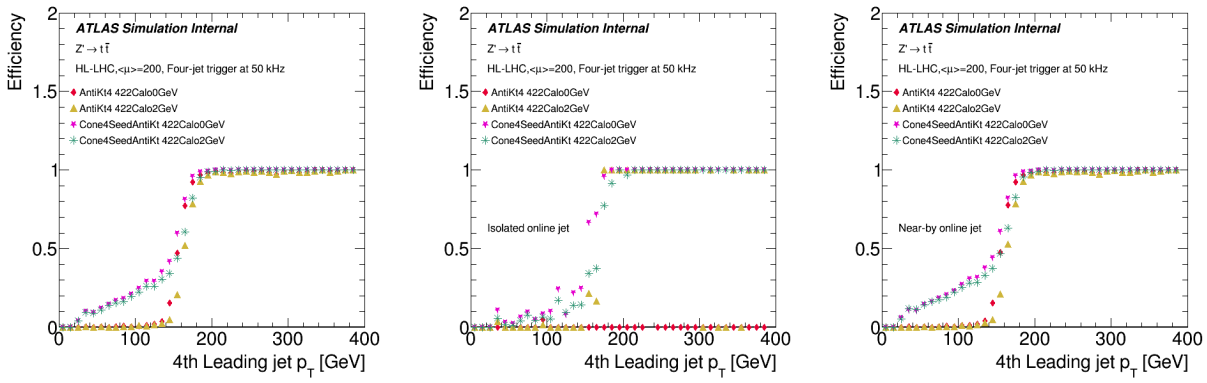
the seed, and the jet is assumed massless,

$$E_T^{\text{jet}} = \sum_{i \in Clusters} E_T^i, \quad \eta^{\text{jet}} = \eta^{\text{seed}}, \quad \phi^{\text{jet}} = \phi^{\text{seed}}, \quad m^{\text{jet}} = m^{\text{seed}}. \tag{8.1}$$

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 is more significant the lower the transverse energy of the seed. However, the resolution is still contained within an average area of 0.1×0.1. The computation of the jet E_T assumes the jet constituents to be collimated. Again, this assumption can be good for boosted jets, but 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, the type of seeds to use and how to select them were non-straightforward questions with, as we will see, several interrelated consequences. 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. Cone jet reconstruction can go wrong in two ways: it can reconstruct the wrong jet, or it can use the right seed, but reconstruct the wrong energy. Using offline anti-k$_t$jets as seeds made possible to decouple the two issues and focus on the latter. Fig. 8.10 shows the comparison of `AntiKt422` jets and `Cone422` seeded with offline anti-k$_t$jets for a four-jet trigger. These are shown with the $Z' \to t\bar{t}$ 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. A similar behavior is observed for the three-jet trigger. When separating the signal sample into isolated and non isolated jets, as shown in Fig. 8.10b and 8.10c respectively, this over-efficiency was shown to be coming from the subsets of events with non-isolated online jets. This study motivated the need for an energy overlap removal strategy.



(a) (b) (c)

Figure 8.10: Four-jet trigger efficiencies for $Z' \rightarrow t\bar{t}$ signal samples. The offline jet collection is `AntiKt420`. 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. An obvious candidate seed type were the trigger objects (TOBs) output by the jFEX algorithm (described in Sec. 4.3.4), which will be available in the GT. However, the minimum dR distance requirement between seeds of the the jFEX sliding window algorithm was observed to cause a constant inefficiency for multi-jet triggers and signal samples with nearby jets. This comparison will be shown in Sec. B.0.0.1. 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, with the expectation that this would exacerbate the energy overlap issue, as well cause a double counting of jets. Both issues will be solved with an energy overlap removal step, as discussed later.

An E_T thresholding was needed to select the topoclusters to be used as seeds. From now on when referring to seeds it will be assumed to be the list of topoclusters with E_T above

the selected seed E_T threshold. The number of potential seeds as a function of possible E_T thresholds is shown in Fig. 8.9.

Similarly to the maximum number of inputs, the maximum number of seeds that can be computed for any event has to be fixed for the firmware. In principle, once the seed list of thresholded topoclusters is formed, only the N highest E_T seeds should be used as seeds. The choice of the maximum number of seeds is therefore 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 downstream algorithms that might need a larger number of jets as input. For this reason, only the seed E_T threshold was considered in these studies and all the seeds above threshold were used.

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 a cut of 5 GeV and 7.5 GeV performed equally well without introducing any inefficiency. For the following studies, a seed threshold of 5 GeV was used.

Energy overlap removal

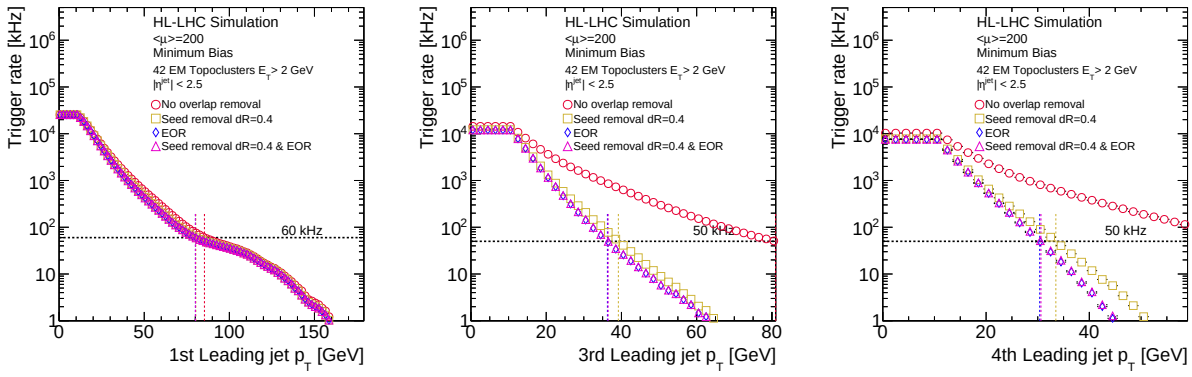
An energy overlap removal strategy was necessary to remove energy and jet double counting. Two strategies were developed. The first strategy, referred to as "seed removal" (SR), aims at removing the possibility of a seed to be a constituent of more than one jet. The list of sorted seeds is parsed in 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 within $0.4 < dR < 0.8$. 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 higher energy seed.

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 is assigned only to the highest energy seed. The jets transverse energy is built only afterwards, to avoid repeating the computations. This procedure is similar to the CMS Cone algorithm described in Sec. 5.2.2 (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.

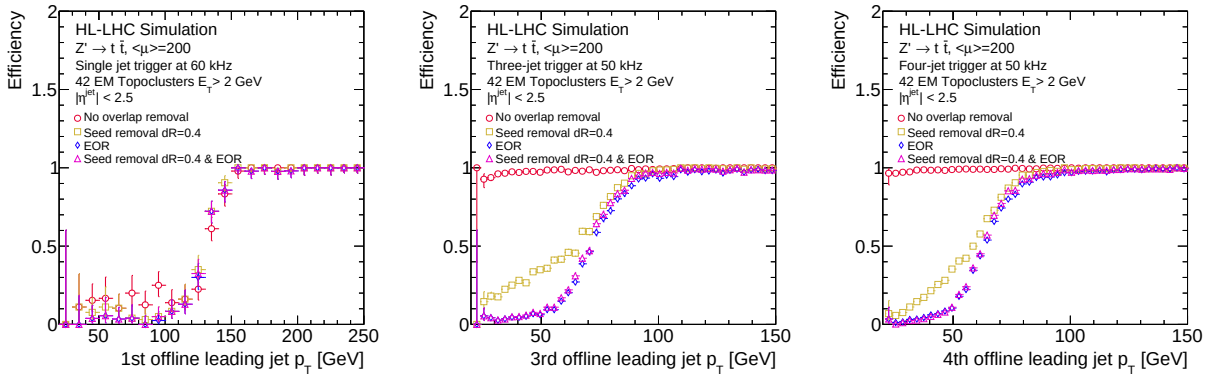
The effect of these strategies was studied in terms of trigger rates and trigger efficiencies. These are shown in Fig. 8.11 and **??**, respectively. First, note that the energy double counting issue does not affect the one jet trigger. Without any overlap removal strategy (red circles), the rate of high energy background jets is artificially increased, as shown in Figs. **??**. This causes the online p_T threshold to be significantly higher than it should be. For a signal sample like Z', 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 trigger 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 energy regime. Some jet double-counting occurs, producing an over-efficiency below the 100% efficiency threshold, but this is not sufficient to compensate for the loss of efficiency due to the increase in rates and the higher online p_T^{cut}. Clearly, an energy overlap removal is necessary. 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.



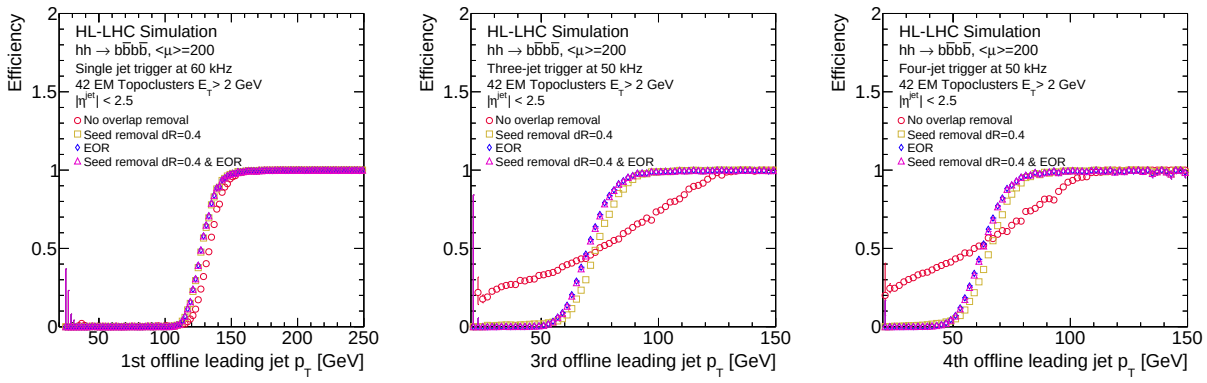
(a) (b) (c)

Figure 8.11: One-jet and multi-jet trigger rates using cone jets with different overlap removal strategies.



(a) (b) (c)

Figure 8.12: One-jet and multi-jet trigger efficiencies for $Z' \to t\bar{t}$ signal using cone jets with different overlap removal strategies.



(a) (b) (c)

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 Performance

This section presents the performance of the optimized cone algorithm. The final cone jet collection, referred to a `ConeTopo`, is reconstructed with the following algorithm:

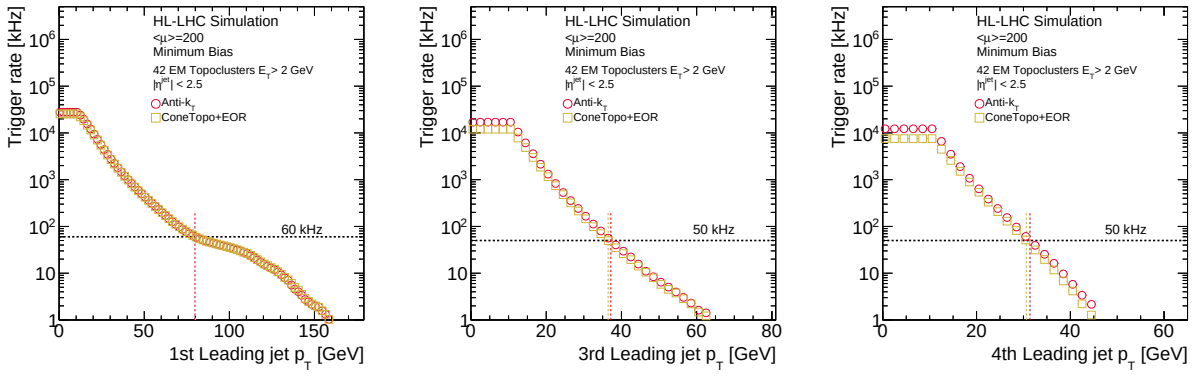
- **Inputs:** Topoclusters with $E_T > 2\,\text{GeV}$
- **Seeds:** Topoclusters with $E_T > 5\,\text{GeV}$
- **Recombination scheme:** Approximate-E_T
- **Overlap removal:** EOR

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 $E_T > 1\,\text{GeV}$ and $E_T > 3\,\text{GeV}$ thresholds.

Trigger efficiencies

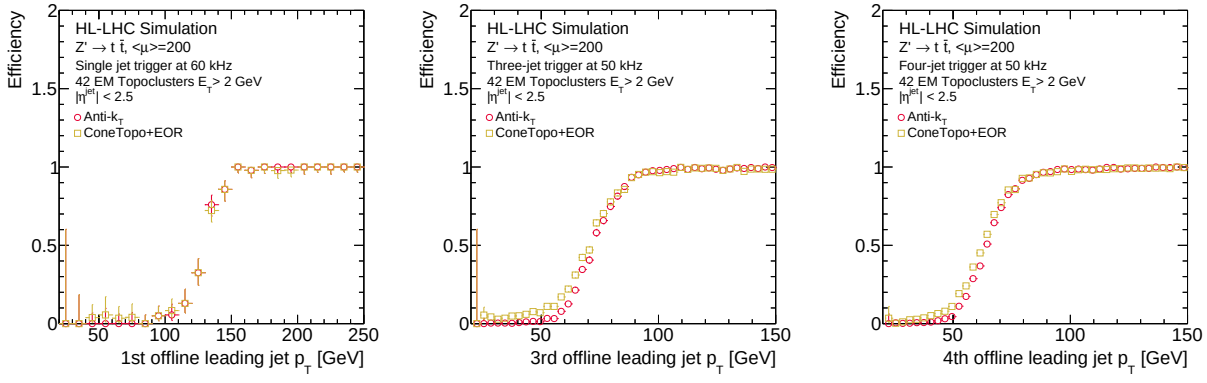
Fig. 8.14 shows the trigger rates with the minimum bias sample. The rates are identical

for ConeTopo and Antikt422 jets. Fig. ?? show the trigger efficiencies for $Z' \to t\bar{t}$ and $hh \to b\bar{b}b\bar{b}$ signals, respectively. Again, the performance is almost identical. The only observable difference is a slight over-efficiency for multi-jet triggers with the $Z' \to t\bar{t}$ 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.



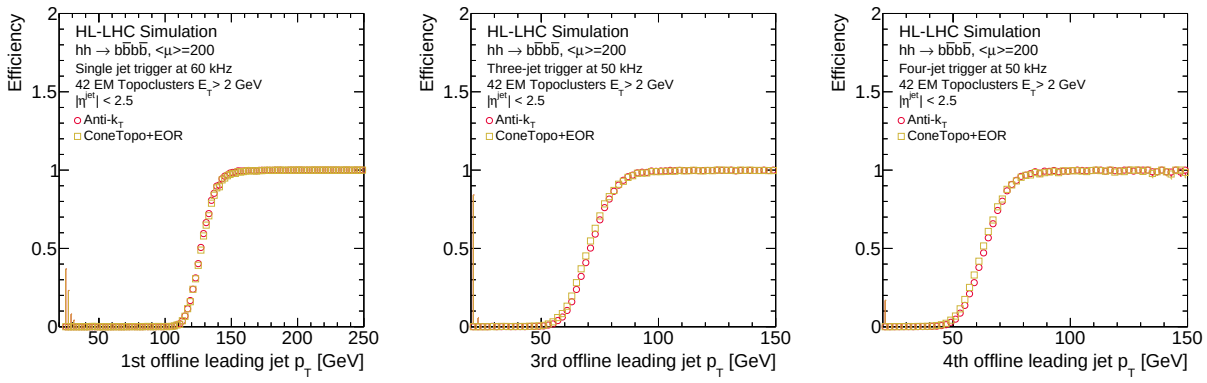
(a) (b) (c)

Figure 8.14: One-jet and multi-jet trigger rates comparing cone jets and AntiKt422 online jet collections.



(a) (b) (c)

Figure 8.15: One-jet and multi-jet trigger efficiencies for $Z' \to t\bar{t}$ signal comparing cone jets and AntiKt422 online jet collection.



(a) (b) (c)

Figure 8.16: One-jet and multi-jet trigger efficiencies for di-Higgs signal comparing cone jets and `AntiKt422` online jet collection.

Di-Higgs m_{HH} signal efficiency

After studying the trigger performance, the effect of the trigger selection was studied on a metric more closely related to the offline analysis. In particular, it was important to check that the cone algorithm was not introducing unexpected bias through, for instance, the energy overlap removal. 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.

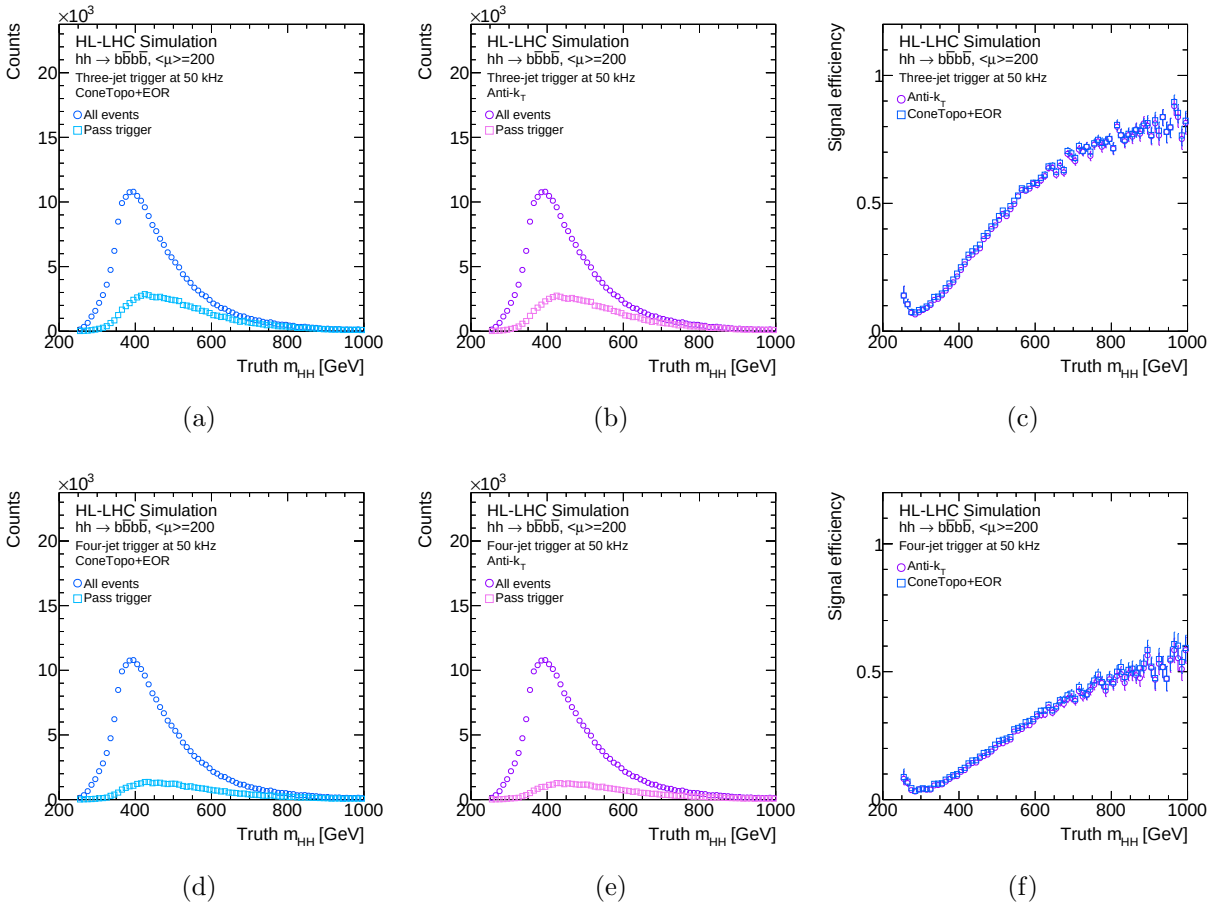


Figure 8.17: Efficiency as a function of reconstructed m_{hh} after passing the three-jet (top row) and four-jet (bottom row) trigger comparing the performance of ConeTopo and AntiKt422 online jet collections.

8.3.3 Constituent-level pileup suppression

In this section the robustness of the cone jet algorithm is studied against the offline pileup suppression algorithms, Soft-Killer (SK) and Voronoi+Soft-Killer (VorSK). The SK algorithm removes topoclusters from the event 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, what matters is whether this decrease affects signal jets in the same way. If this is the case, then this accounts as a simple normalization difference and will not change the offline p_T threshold.

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 most affected by the presence of pileup, but similar results were observed for the $Z' \to t\bar{t}$ signal. The the `ConeTopo` jet reconstruction is used, but identical conclusions were obtained with online anti-k$_t$. 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. **??**, respectively.

As expected, applying pileup suppression on the topoclusters reduces the rate of background jets across the entire $p_T$regime, 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 signal efficiencies show that SK can bring a significant improvement to the offline analysis sensitivity by reducing the offline p_T

threshold. However, this effect diminishes with increasing topocluster E_T threshold. The effect of applying Voronoi before running the SF algorithm is instead negligible. This might indicate that Voronoi is not sufficiently discriminating between signal and background jets, and the same reduction factor is applied to all.

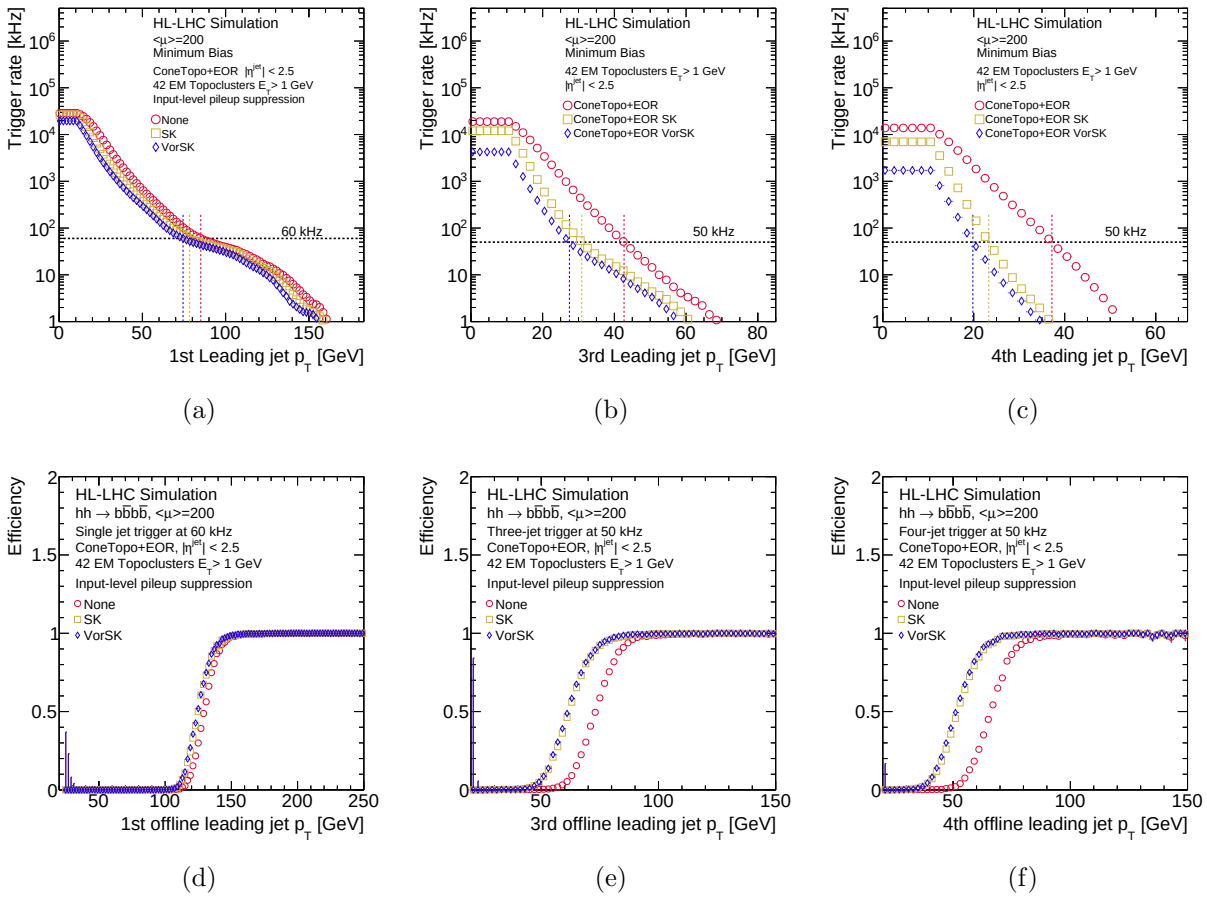


Figure 8.18: Compare trigger rates (top) and trigger efficiencies (bottom) using using different constituent-level pileup suppression methods, as well as no pileup suppression. Turn-on curves are built using online p_T cuts derives from the corresponding fixed trigger rate threshold. The efficiens are shown for the di-Higgs signal sample and the ConeTopo online jet collection built from input topoclusters with $E_T > 1\,\text{GeV}$.

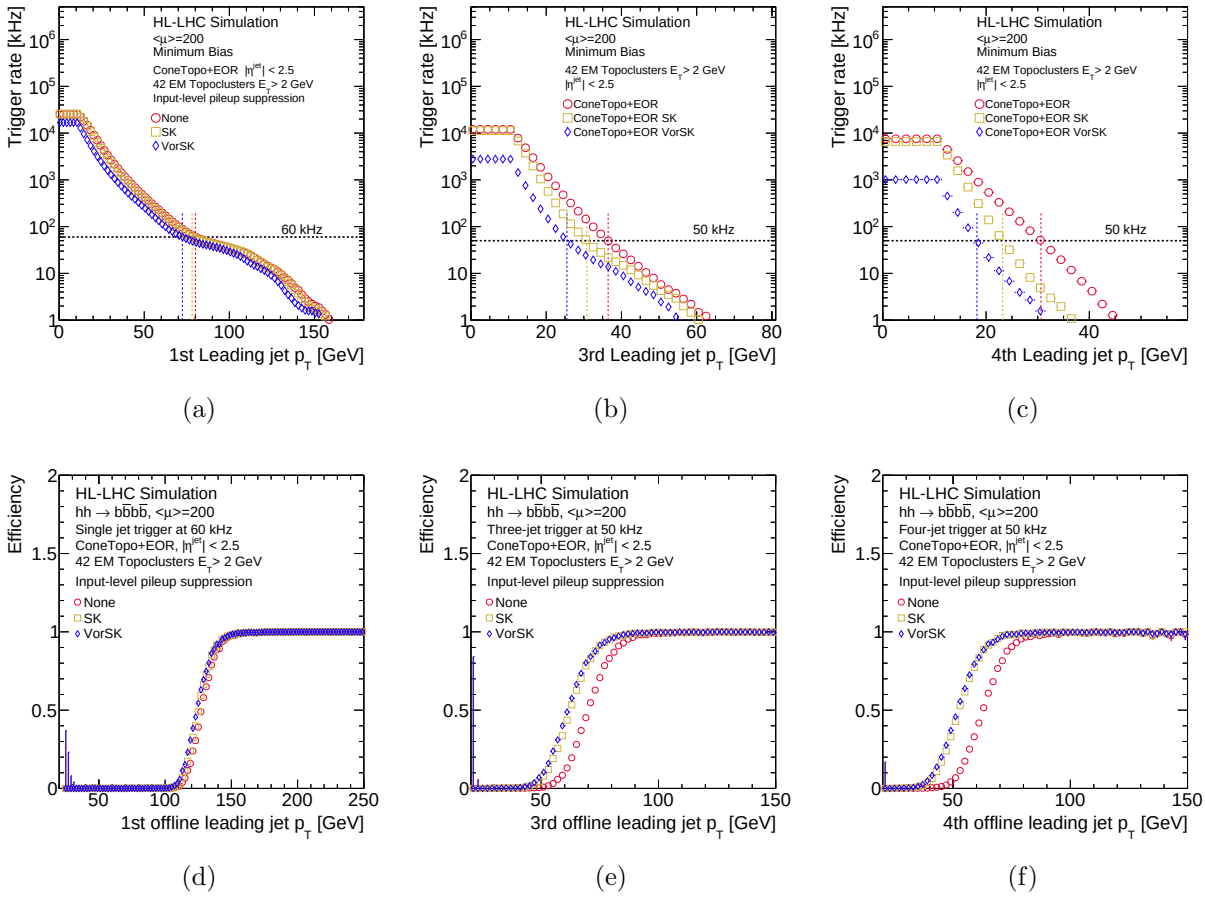


Figure 8.19: Trigger rates (top) and efficiencies (bottom) using different constituent-level pileup suppression methods, as well as no pileup suppression. Turn-on curves are built with online p_T cuts at fixed trigger rate threshold. The efficiencies are shown for the di-Higgs signal sample and the `ConeTopo` online jet collection built from input topoclusters with $E_T > 2\,\text{GeV}$.

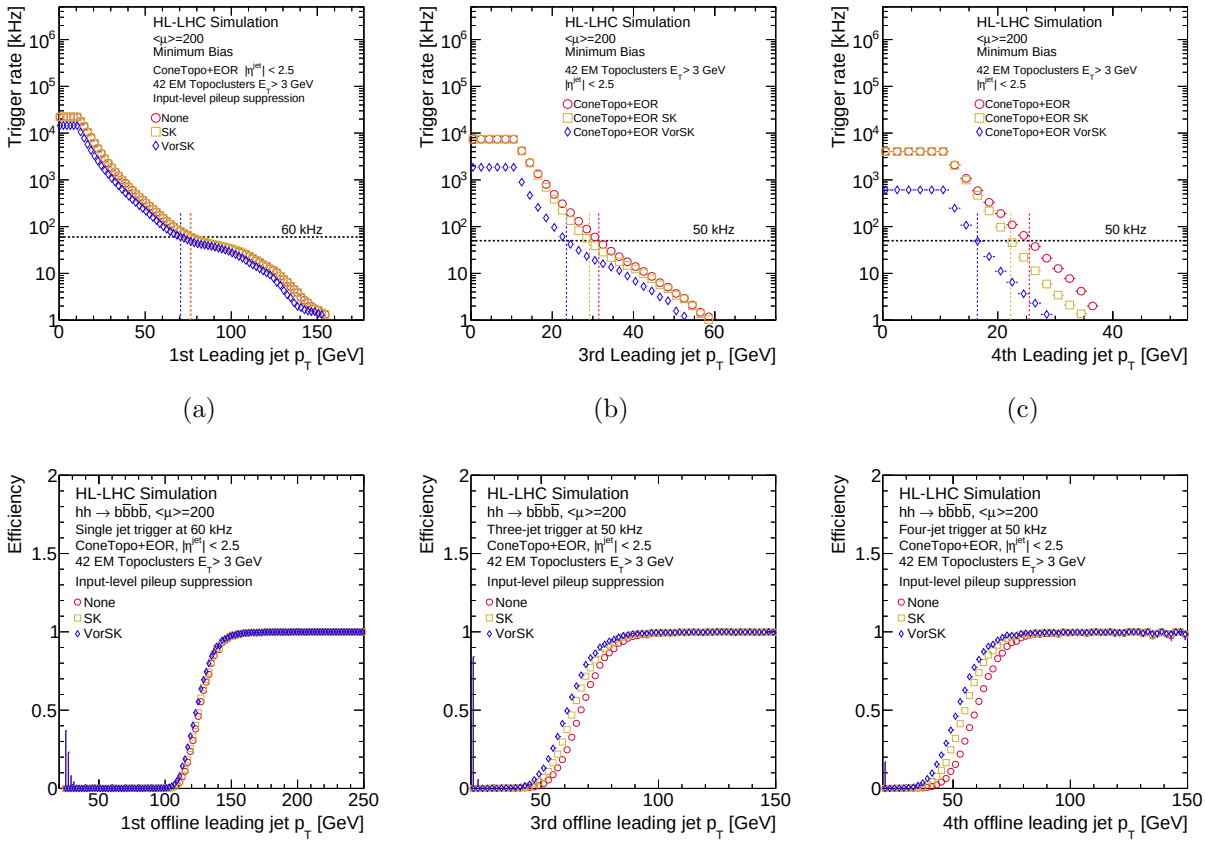


Figure 8.20: Trigger rates (top) and efficiencies (bottom) using different constituent-level pileup suppression methods, as well as no pileup suppression. Turn-on curves are built with online p_T cuts at fixed trigger rate threshold. The efficiencies are shown for the di-Higgs signal sample and the `ConeTopo` online jet collection built from input topoclusters with $E_T > 3\,\text{GeV}$.

8.3.4 Preliminary firmware simulation

A preliminary firmware simulation of the cone jet algorithm was performed in order to get an estimate of the FPGA resource utilization. As we saw in the previous sections, this helped guiding certain decisions during the development.

The simulation was performed using the Vitis High Level Synthesis (HLS) package provided by Xilinx, the same company manufacturing the FPGA boards that will be used in the GT. The software suite provides a software developer friendly interface to synthesize hardware code. It allows a faster turn-around to optimize code for throughput, power,and latency. The workflow was the following:

1. 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 memory cannot be implemented.

2. Provide a set of inputs and expected results to use for validation. Write a *testbench* that reads the input test files, runs the algorithm, and compares the output to the output test file. 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 same output as the provided test file.

4. Synthesis: Synthesize C algorithm in an RTL implementation. Vitis HLS will effectively compile the C code into hardware description language, VHDL or Verilog.

5. C-RTL Cosimulation: Use C testbench to validate RTL design and confirm it produces the same output as C-level code.

6. Use code directives to produce different RTL versions and analyze the designs by looking at the reports of the resource consumption.

7. Export RTL design as an IP block that can be integrated into the hardware.

A preliminary firmware simulation of the cone jet algorithm was performed. The energy overlap removal step is not included in the simulation. It is also assumed that the input list of topoclusters is 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 `N_MAX_SEED` $= 10$ and `N_MAX_CONSTITUENT` $= 40$, but these values are not optimized.

Before the simulation, certain constraints have to be set by the user. These are the clock period, the clock uncertainty, and the FPGA target. The target clock period was set to the L0 trigger rate of $240MHz$, or $4.17ns$. 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. The latency is the number of cycles it takes to produce the output. When synthesizing the RTL 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. The

final sequence of operations is extracted as an RTL design, which is analyzed to obtain the performance and resource estimates.

The estimate of the fastest achievable clock frequency was 3.027 ns, while the latency was 378 cycles. This gives a reasonable 1.144μs latency. The resources used to implement the design are shown in Tab. 8.1, in terms of absolute numbers and of fraction of resources of a single SLR on a VP1802, the target device for the GT. These are close to the expected FPGA budget for jet reconstruction, currently set at 25% of one SLR. Note that the simulation was performed using floating point data-types, so the resources utilization is expected to go down with fixed-point precision.

	DSP	FF	LUT
Absolute numbers	681	131719	81755
% VP1802 single SLR resources	19	7.8	9.7

Table 8.1: Resource estimates of cone jet algorithm from preliminary firmware simulation.

This preliminary simulation was used as a guidance to understand the roadblocks of the algorithm. For instance, this motivated the investigation of a different recombination scheme option. The firmware was not optimized, as the algorithm was still under development. The purpose of this study was to confirm that the algorithm could be a viable option for the GT and the results were indeed promising.

8.4 Pileup-jet suppression using machine learning

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 an important result, as `AntiKt422` represents the upper bound on jet reconstruction performance at the Global Trigger level. However, the question

remained whether this was enough to retain ATLAS physics goals and, if not, whether one could improve the performance even further.

One of the main difficulties forseen for the HL-LHC are the extreme pileup conditions, with up to 200 secondary interactions per bunch crossing. As described in Section 5.5, pileup can significantly impact object reconstruction. In the case of jets, the effect is usually that of introducing a positive bias in the reconstructed transverse momentum and an overall resolution degradation of the reconstructed quantities due to pileup fluctuations. While these effects were discussed in the context of offline reconstruction, they affect equally the online reconstructed jets. 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. Another important practical consequence of pileup is its impact on the TDAQ system. The busier environments result in a larger quantity of data to be processed per event. This stresses the DAQ system, increasing the latency and bandwidth constraints on the trigger, while at the same time requiring more sophisticated 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 $hh \to b\bar{b}b\bar{b}$.

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. 3.0.1). This is a very challenging mesurement, 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 is complicated further by a destructive interference between

the two contributing diagrams (Fig. 8.21), as shown in Fig. 8.22. Only the low m_{hh} mass region provides a handle to measure possible deviations from the SM Higgs self-coupling, an experimentally difficult region dominated by pileup. The latest results from ATLAS and CMS, which used Run 2 data, have only been able to set limits [7, 8]. However, more than 10^5 pairs of Higgs bosons are expected to be observed by each experiment by the end of the HL-LHC and the projected combined signal significance with (without) systematic uncertainties is of 4.0σ (4.5σ) [9]. The HL-LHC is therefore expected to reach the ultimate sensitivity, making di-Higgs one of the flagship signatures of the HL-LHC upgrade.

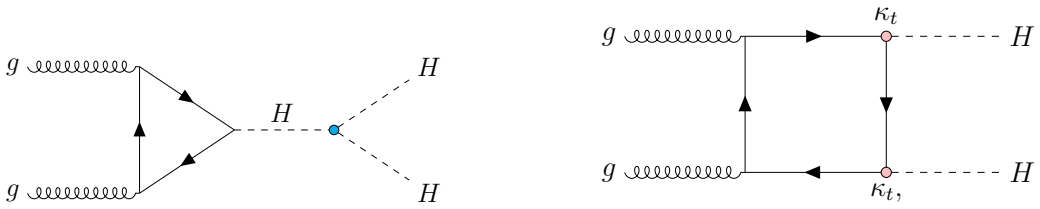


Figure 8.21: Leading order diagrams contributing to gluon-gluon fusion di-Higgs production cross section. Only the right diagram depends on the Higgs self-coupling λ.

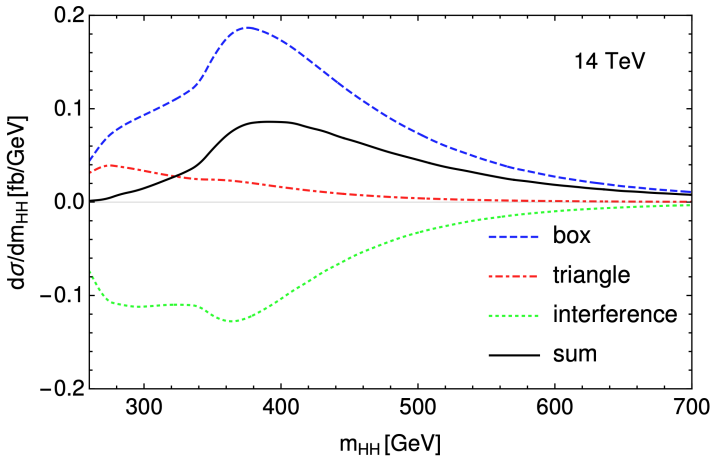


Figure 8.22: Invariant mass distribution of di-Higgs pair for the gluon-gluon fusion production mode at leading order for the different contributing diagrams and their intereference [9].

The decay channel $hh \to b\bar{b}b\bar{b}$ ($hh4b$) has one of the largest sensitivities thanks to the largest branching ratio. In the low mass non-resonant region the $hh4b$ final state is charac-

terized by four low energy jets, a region of phase space dominated by multi-jet background and pileup. Masurement 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 Global Trigger. In this section a new method for mitigating the impact of pileup on the Level-0 (L0) multi-jet trigger performance targeting $hh \to b\bar{b}b\bar{b}$ signal is investigated.

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 running 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. A jet whose transverse momentum is mostly due to pileup particles will be referred to as a *pileup 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. Of particular importance is the contribution of pileup jets to three and four-jet trigger rates, which rapidly grows larger with increasing number of pileup vertices. In this study we investigate the use of deep learning to identify and reject pileup jets in the Global Trigger. In this context, we define a signal jet as a jet truly originating from the hard scatter. The novel feature of the approach is that it applies pileup mitigation on a jet-level basis. Three pieces of information are useful for pileup identification: the local energy and multiplicity distributions of pileup particles with respect to particles originating from the hard vertex, the number of charged tracks originating from pileup vertices, and the event pileup density. While tracking information is a very useful pileup discriminant,

it will not be available in the ATLAS L0 trigger. A metric of the hadronic activity in the event would also be desirable, but it is unclear at the moment whether such a computation would be feasible for the Global Trigger. The goal of this study was therefore to understand whether the first piece of information, the local distribution of the jet's constituents, could be sufficient to identify a jet as being pileup or not.

This question was motivated by the fact that the parton shower-like radiation found in a jet induced by a hard quark or gluon tends to be collimated and with a more uneven energy sharing among the constituents, while the energy pattern of a pileup jet will be more diffuse and with most constituents around the same energy scale. This is shown in Fig. 8.23. These studies were produced in the following way. The jets[2] were built with all the topoclusters above a given E_T threshold, without pileup suppression. After the jets were formed, the jet constituents were compared with the Soft-Killer pileup suppressed topocluster collection to determine if the 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 and pileup jets. Note that each bin content is given by the sum of the E_T of the jet constituents at a given dR distance from the jet center. Similar profiles were obtained when weighting by the constituent E_T normalized to the E_T of the leading constituent or to the jet p_T. The histograms are shown in bins of jet pt: [15,35], [35,50], [50,70], [70,100] GeV. Radiation that is deemed "pileup-like" by SK is uniform and low in energy, resulting in an energy profile that is a straight line increasing with increasing distance from the jet center. This feature is identical for signal and pileup jets, subject to the same uniform pileup contamination. The

[2]The jet collection used for these studies was cone jets built with input topoclusters with $E_T > 1\,\text{GeV}$ and with seed removal with $dR = 0.3$. This is the cause of the upward shift in some histograms at $dR = 0.3$, when jets overlap starts being allowed.

energy profile of signal jets after pileup-suppressed topoclusters are removed peaks close to the center of the jet and falls off rapidly at large radii, as expected. 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 above 50 GeV. However, discrimination power is present in the lower energy bins which fall around the region of interest for a multi-jet trigger efficiencies, motivating this study.

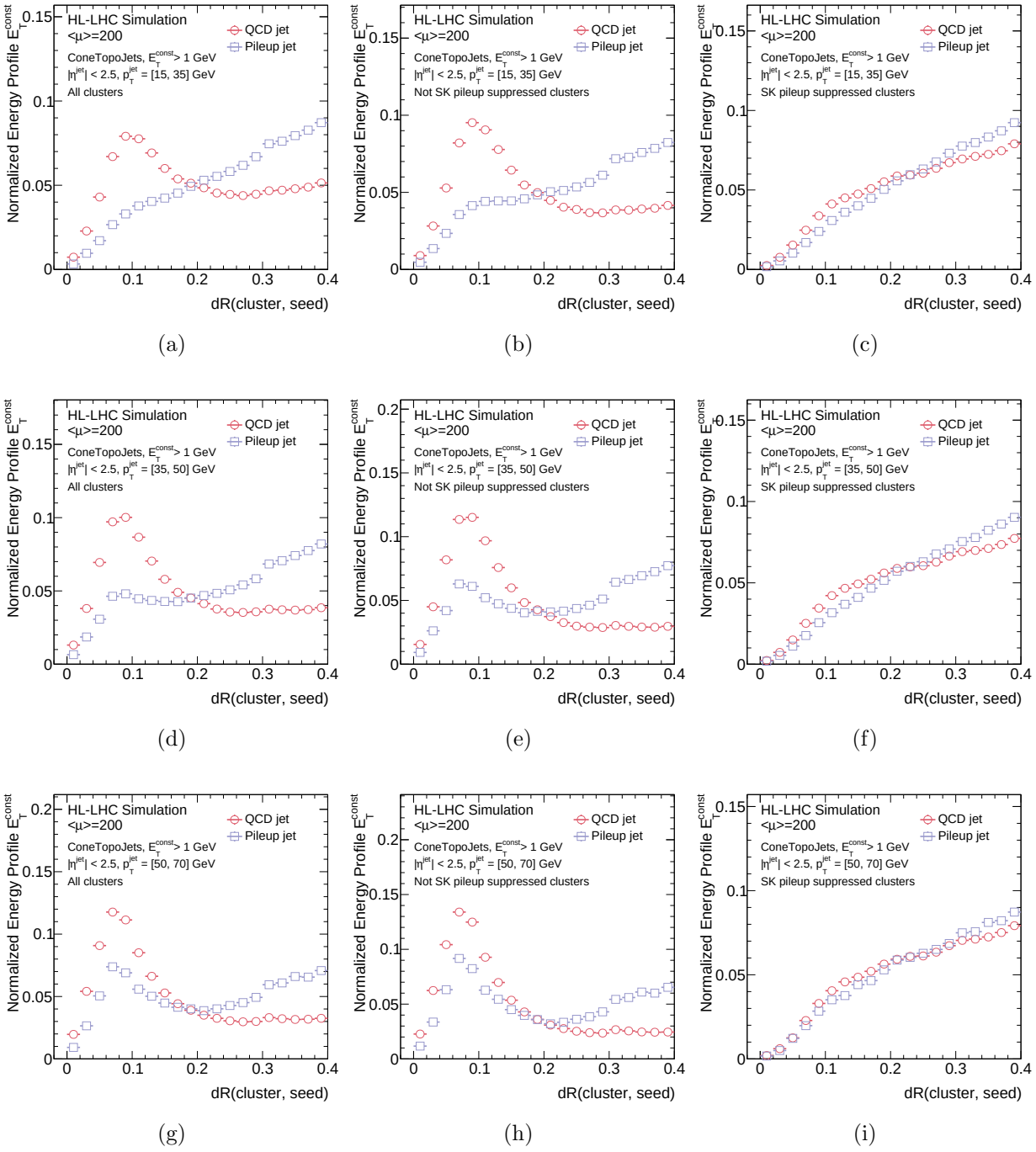


Figure 8.23: Jet energy profile for `ConeTopo` jets built with seed removal using $dR = 0.3$. Comparing all constituents (left), only constituents that would have not been pileup suppressed (center), and constituents that would have been pileup suppressed (right). The studies are performed separating the jets into p_T bins. From top to bottom: [15,35], [35,50], [50,70] GeV.

8.4.2 Neural network for pileup jet rejection

A neural network (NN) was developed as a jet-by-jet classifier to output the probability of a jet to be signal ($p = 1$) or pileup ($p = 0$). Two different NNs were trained using different input variables.

Training samples

The samples used for training were the multi-jet MC samples described in Sec. 8.2.1. Pileup (background) samples were selected by identifying jets not truth matched to any truth quark from the minimum bias sample generated with truth di-jet transverse momenta in the range $[0, 20]$ GeV. The sample of QCD jets (signal) was obtained by selecting only jets truth matched to a true quark across a wider p_T range. The jet collection used to select training jets were `ConeTopo` jets built with topoclusters with $E_T > 1$ GeV and the seed removal $dR = 0.4$ (SR4) applied. Energy overlap removal was not applied to preserve jets with a constant round shape for training. A further requirement on the leading E_T constituent to be the seed of the jet is imposed to avoid a higher energy seed removed by SR to be associated to a lower energy seed.

As will be shown later, the selected signal and background samples are characterized by very different p_T distributions, with signal jets covering a wide p_T range and background jets peaking at small values. In order to avoid the network to classify only based on the jets p_T, different steps were taken. First, the training was performed only in the region around the turn-on curve, for jets between 25 and 50 GeV. Values below this range are not useful, since such low p_T jets do not contribute to the rates that drive the online trigger p_T cut. Above this range, the discrimination power in the local distribution of the jet constituents was observed to degrade, as shown in Fig. 8.23h. The samples were then reweighted to have

a flat p_T distribution and an equal class normalization. The data set was then split into training, testing, and validation samples, with $\approx 500k$ training events.

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 the topoclusters forming 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 pileup suppressed. Different input variables were considered for each case. Here the decision process is described, while the actual distributions are discussed 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. **??**) and the discrimination power between signal and background for the n^{th} leading topocluster decreases for increasing n, as will be shown in Fig.s **??**. 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 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, as shown later in Fig. 8.28. 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. 8.23. This will be referred to as DNN-B. The training was performed using cone jets built with non-pileup suppressed topoclusters. If DNN-A is sufficient to improve the trigger performance, the conclusion is that SK is not needed in the GT. If DNN-A is not sufficient, then DNN-B allows to test whether including information from the SK algorithm can improve the trigger performance more than running jet reconstruction on pileup suppressed topoclusters.

8.4.3 Training performance

The NNs were trained using Keras with Tensorflow 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 hyper-parameters. 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 crossentropy

loss with the Adam optimizer, using a learning rate of $1.e-4$, 80 epochs, and a batch size of 500. The final set of input variables were taken to be the $\Delta R(\text{cl}, \text{lead cl})$ and $E_T^{\text{cl}}/E_T^{\text{lead cl}}$ of the 10 leading topoclusters. Information on the leading topocluster was removed, as by construction the leading topocluster has $\Delta R = 0$ and $E_T^{\text{cl}}/E_T^{\text{lead cl}} = 1$. The additional boolean pileup-suppression flag was included for DNN-B. This accounts for 18 input variables for DNN-A and 27 input variables for DNN-B. The training sample 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.

The final metrics computed on the entire validation data set for DNN-A and DNN-B is shown in Tab. 8.2.

The accuracy is the percentage of correct predictions, or the sum of the true positive and true negative rates. 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 (true positive vs. true negative rate) represents the trade-off between signal efficiency and background rejection. 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 high precision (accurate results) and high recall, and so 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.

	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 full validation sample for DNN-A and DNN-B. The metrics include the accuracy, precision, recall, area under the ROC curve, area under the precision-recall curve.

8.4.4 Testing performance

An unbiased evaluation of the model performance was obtained with the unseen testing sample. Fig. 8.24 shows the comparison between DNN-A and DNN-B in terms of output scores, confusion matrices, and ROC curves. In particular, specific working points (WPs) at fixed signal efficiencies are provided for comparison and for later use. As expected from the previous look at the training metrics, DNN-A has a slightly lower performance, especially in terms of true positive rate. While the differences in metrics and confusion matrix are in the few percentage points, the WPs are more telling, as DNN-A has a significant larger background efficiency at a fixed signal efficiency, which will have a significant impact in trigger performance. In Fig.s **??** the input variables are shown for selected n^{th} leading topoclusters. In order to visualize what the network is learning, the distributions are shown for events tagged according to the true label and to the 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 for the distribution of the dR between the higher energy topoclusters and the leading one, as the network learns that signal jets are more collimated, while background jets are more diffuse. Fig. 8.27 shows similar distributions for out-of-training variables. These show that

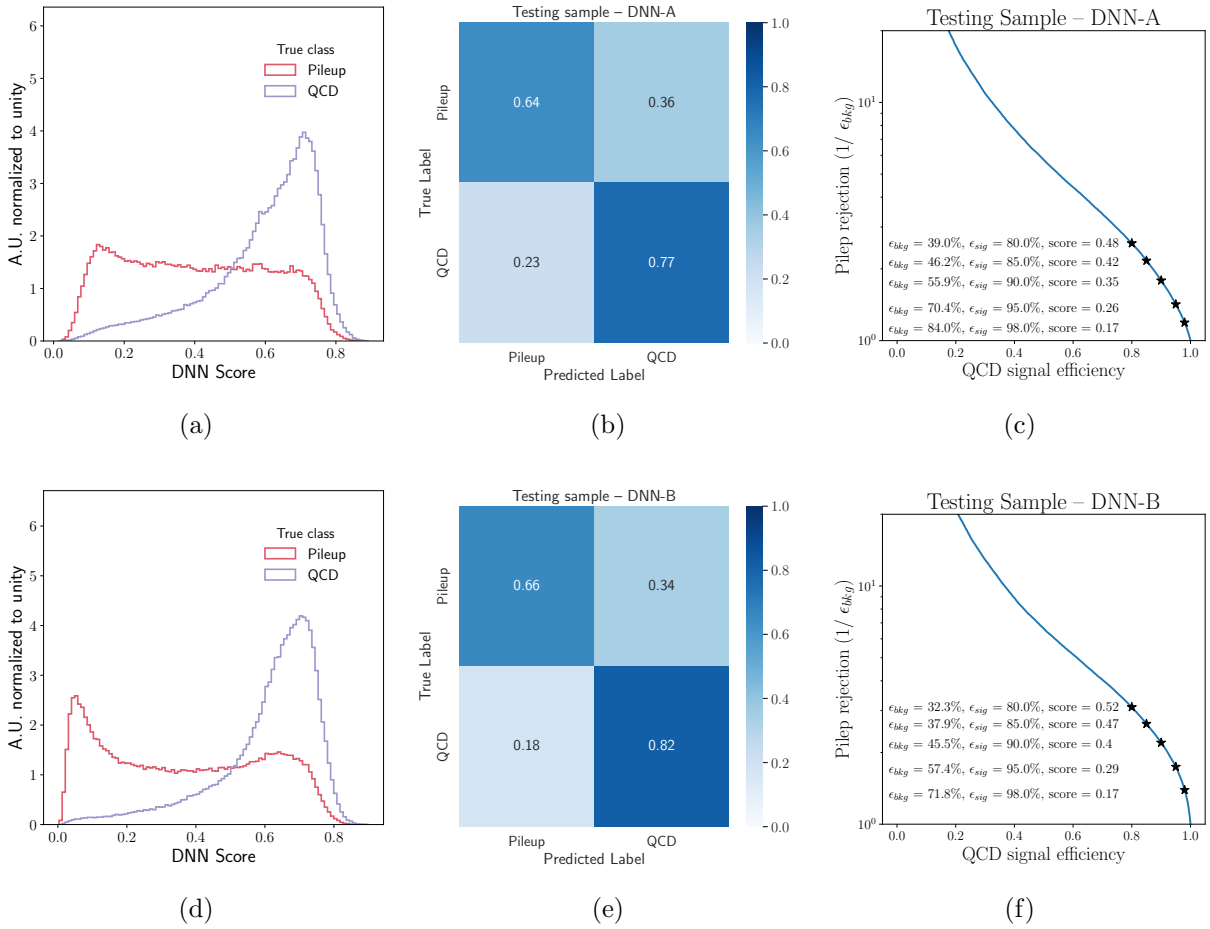


Figure 8.24: Comparison of testing performance in terms of output score (left), confusion matrix (center), and ROC curve (right) between DNN-A (top) and DNN-B (bottom).

the network learns that signal jets have on average fewer constituents and fewer pileup suppressed constituents. The network also correctly learns the correlation between the E_T of the leading topocluster and the p_T of the jet. In particular, Fig. 8.28 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. The source of false negatives (true signal classified as background) identified in the previous discussion comes from low transverse momentum signal jets. These low energy signal jets are the ones with a less collimated energy profile and the mis-classification is

more pronounced for DNN-A. Because of the weaker discrimination power, without the pileup suppression information a greater number of these jets is wrongly identified as pileup.

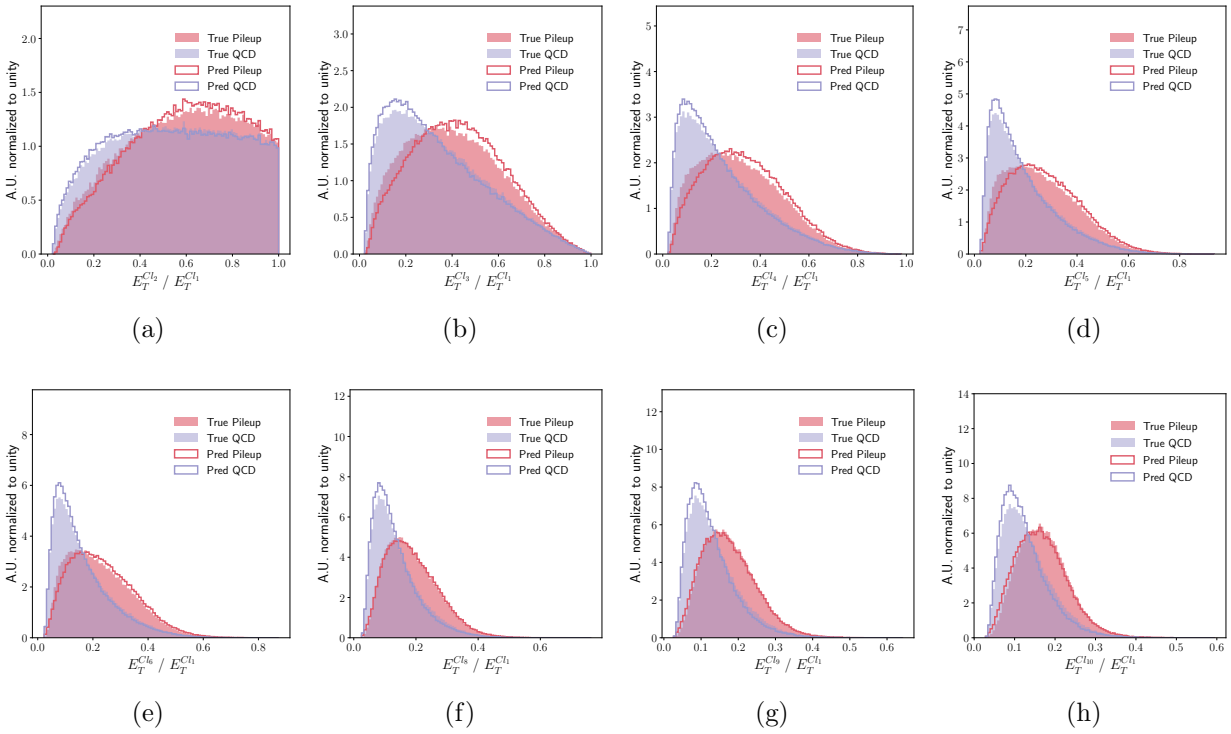
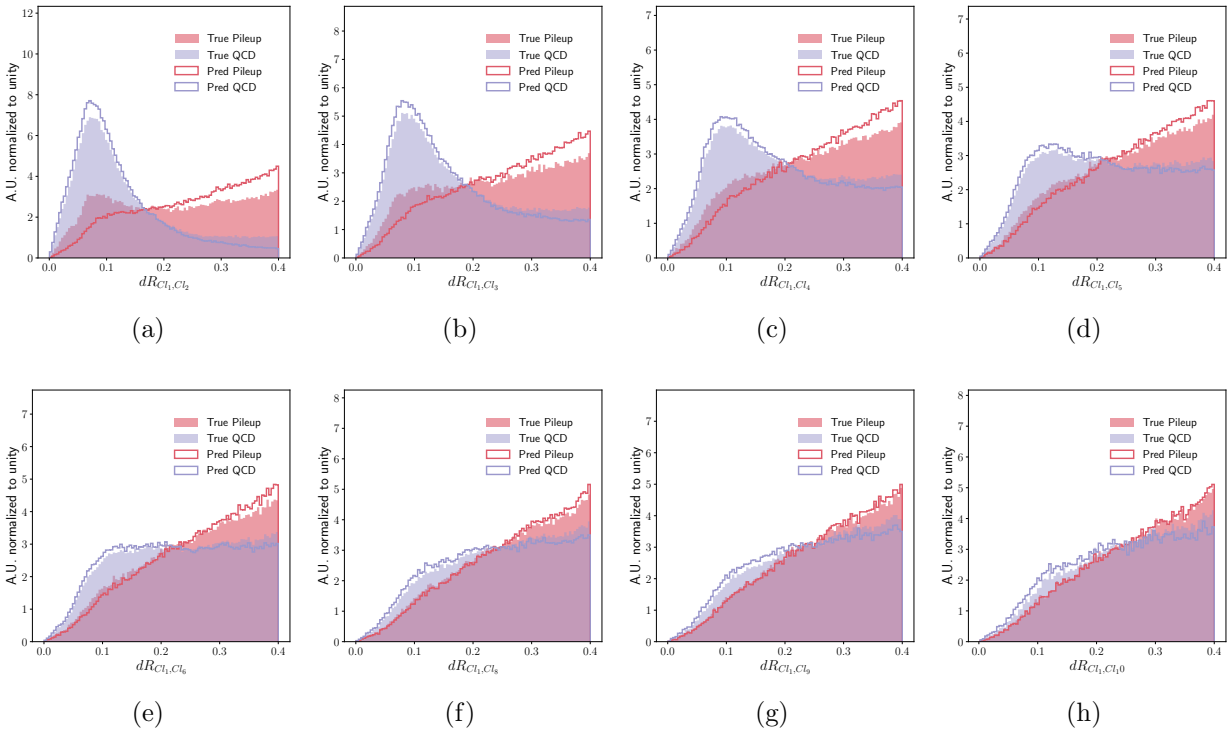


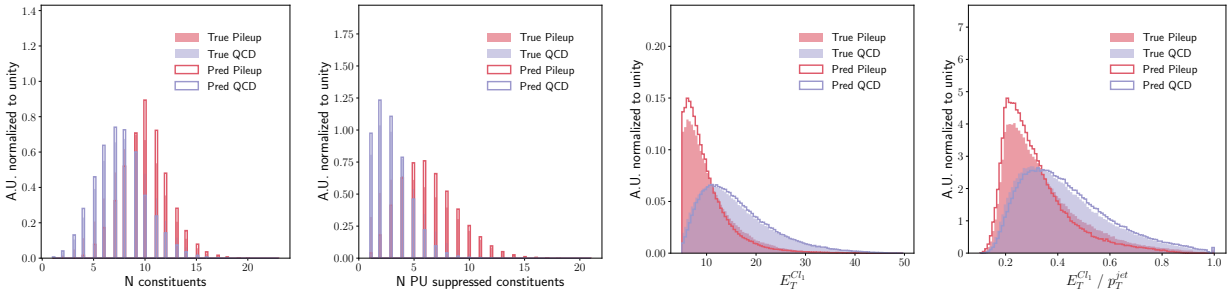
Figure 8.25: Transverse energy of input topoclusters normalized to transverse energy of the leading constituent. The distribution are shown tagged according to the true and predicted label.



(a) (b) (c) (d)

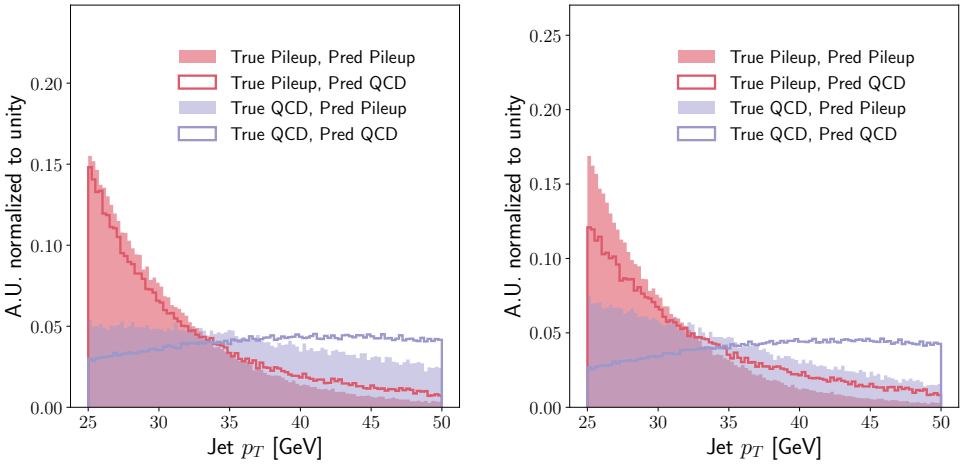
(e) (f) (g) (h)

Figure 8.26: Input variables distribution tagged according to true and predicted label.



(a) (b) (c) (d)

Figure 8.27: Selected out-of-training variables distributions tagged according to true and predicted label.



(a) (b)

Figure 8.28: Out-of-training jet p_T distribution tagged according to true and predicted label for DNN-A (left) and DNN-B (right).

8.4.5 Trigger performance

The neural networks are deployed in the Global Trigger framework to study their effect at the trigger level. The same `ConeTopo` jet collection used for training is used here, with the addition of energy overlap removal. Similar results are observed for the jet collection without EOR and for `AntiKt422` jets. For each model, the procedure is as follows. Each reconstructed jet that would normally enter the trigger performance workflow is passed through the NNs. In the following, a jet is said to be "in-training" if its transverse momentum is in the training range $[25, 50]\,\text{GeV}$. The scores for the four leading jets with in-training p_T are shown in Fig. 8.29 for the minimum bias, hh4b, and di-jet samples. Note that the neural networks correctly identify jets from the di-Higgs sample as signal-like. More confusion is present for the di-jet sample, where the fourth jet can be soft. Note that the two leading jets in the di-jet sample have very few statistics as they populate higher p_T bins. A cut on the score at a fixed signal efficiency WP, as shown in Fig.s **??**, is applied to all jets that enter the study. Each WP produces a new jet collection of "DNN pileup suppressed" jets. The "baseline" jet collection without any DNN cut is also shown for comparison. The p_T spectra of the fourth leading jet after the cuts are shown in Fig.a **??** for model A and B, respecitively. The trigger rates are then built with the minimum bias sample for the different jet collections. These are shown in Fig. 8.32 for both models. Note that two strategies are compared in these plots. One strategy applies the DNN cut to all jets, another to only in-training jets. As these plots show, applying the DNN cut on all jets has a negligible impact on the rates, indicating that pileup jets with $p_T > 50\,\text{GeV}$ are not rejected. It was therefore decided to only apply the DNN cut on jets with $p_T < 50\,\text{GeV}$ in order to retain the maximum signal efficiency. As expected, the harder the cut on the NN scores, the larger the decrease in high energy

pileup jets, and the effect is larger for DNN-B. However, it was important to understand what effect this would have on the trigger efficiencies. Fig. 8.33 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. 8.34 shows the final comparison of the best WPs for DNN-A and DNN-B. The comparison includes the trigger efficiencies obtained by running the same cone jet reconstruction algorithms on pileup suppressed topoclusters using both SK alone and Voronoi+SK. From this plot it is clear that the lower true positive rate of DNN-A has a significant impact on the trigger performance, as the lower background rejection keeps the rates and the online p_T cut higher. On the other hand, DNN-B has a visible impact on the offline p_T threshold. However, when compared to applying SK on the input topocluster collection, it results in a similar performance.

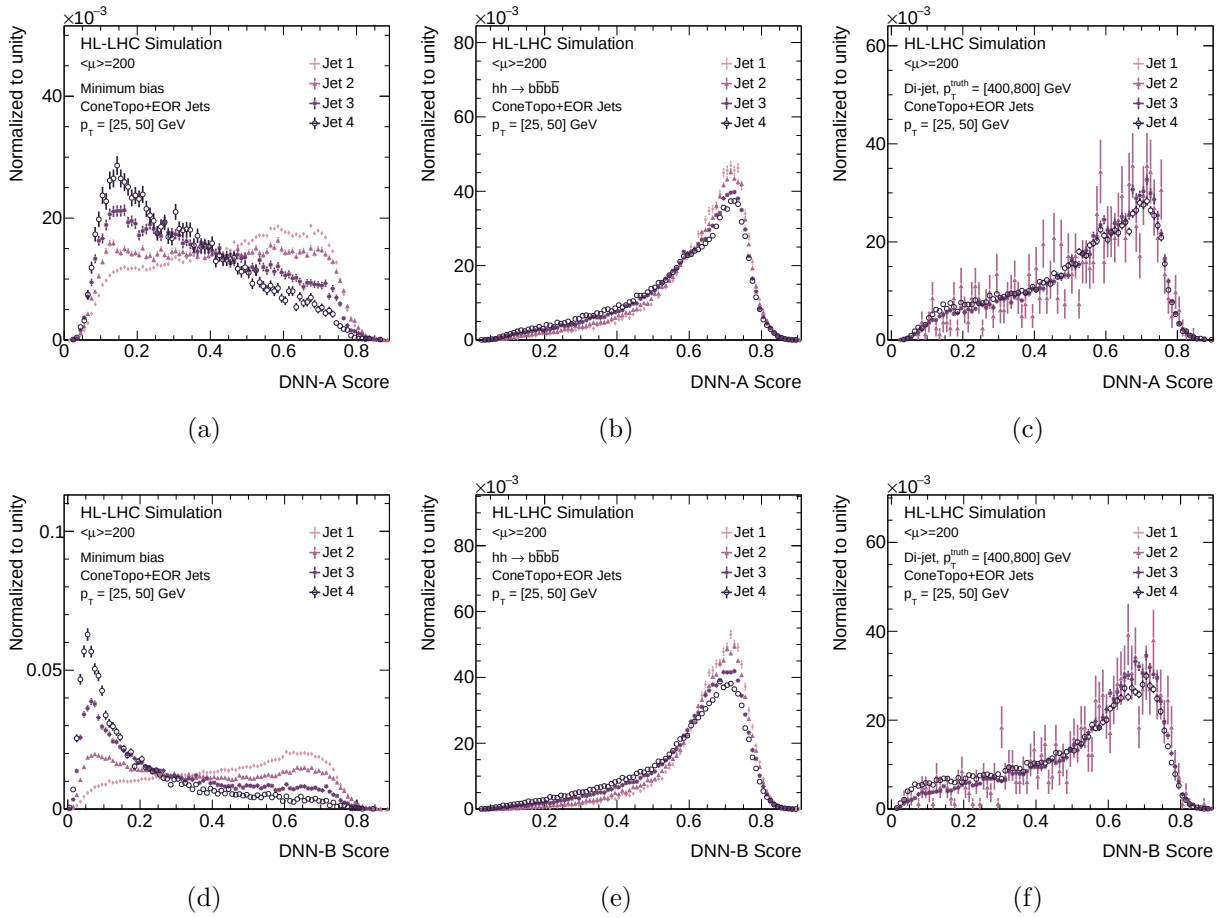


Figure 8.29: Output DNN-A (top) and DNN-B (bottom) scores for four leading jets for in-training jets with $p_T \in [25, 50]$ GeV. From left to right: minimum bias, di-Higgs, and di-jet ($[400 - 800]$ GeV truth p_T slice).

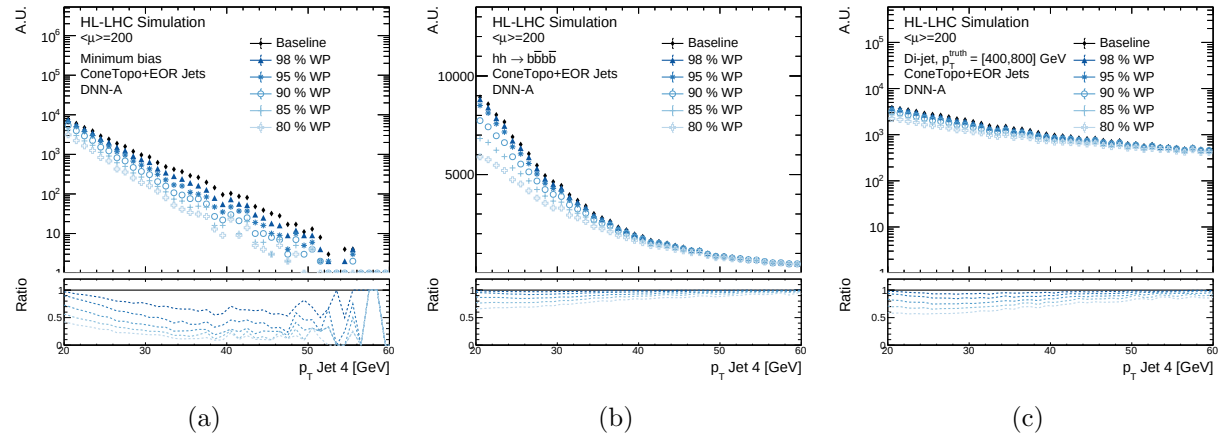
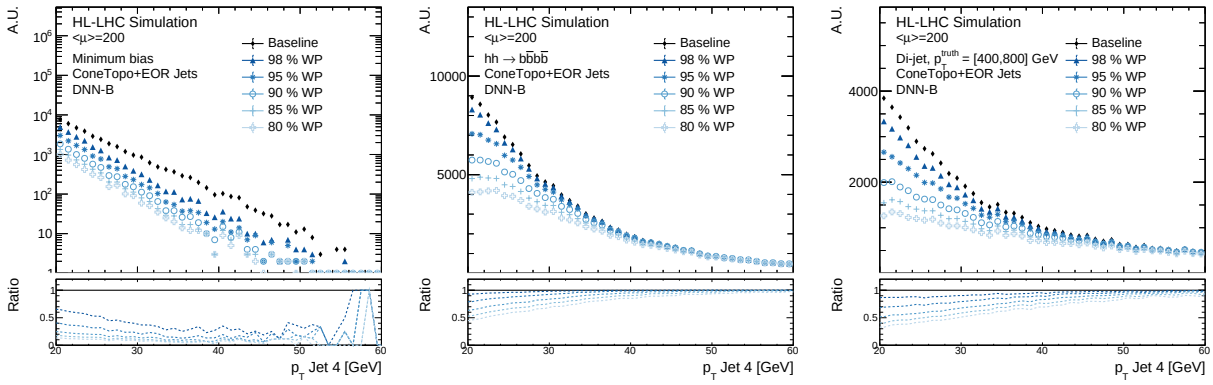
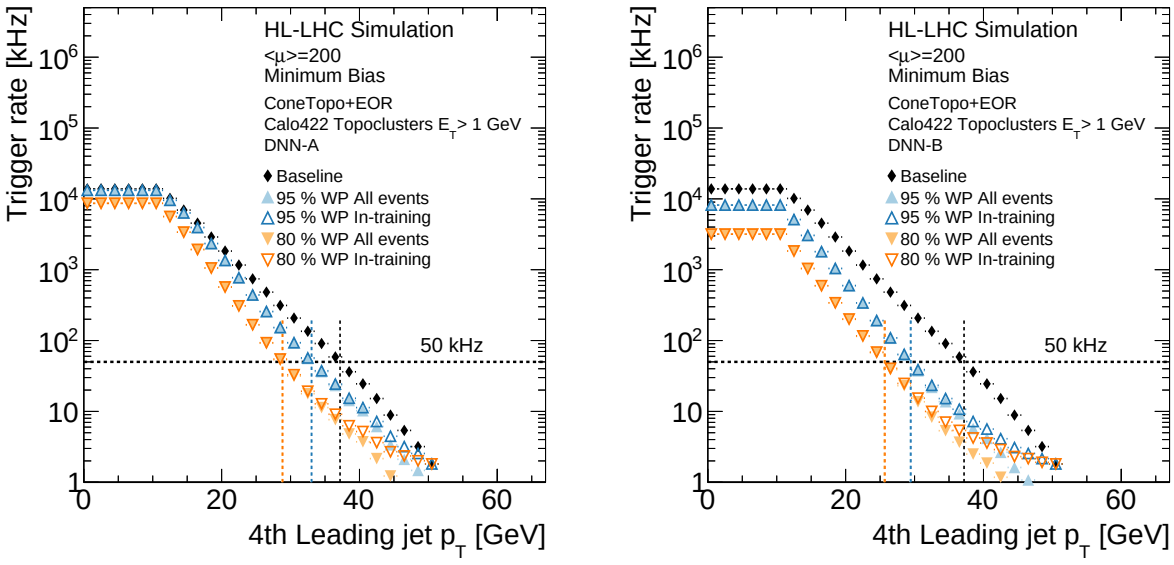


Figure 8.30: Effect of WPs cuts on fourth-leading jet p_T distribution for DNN-A. From left to right: minimum bias, di-Higgs, and di-jet ($[400 - 800]$ GeV truth p_T slice).



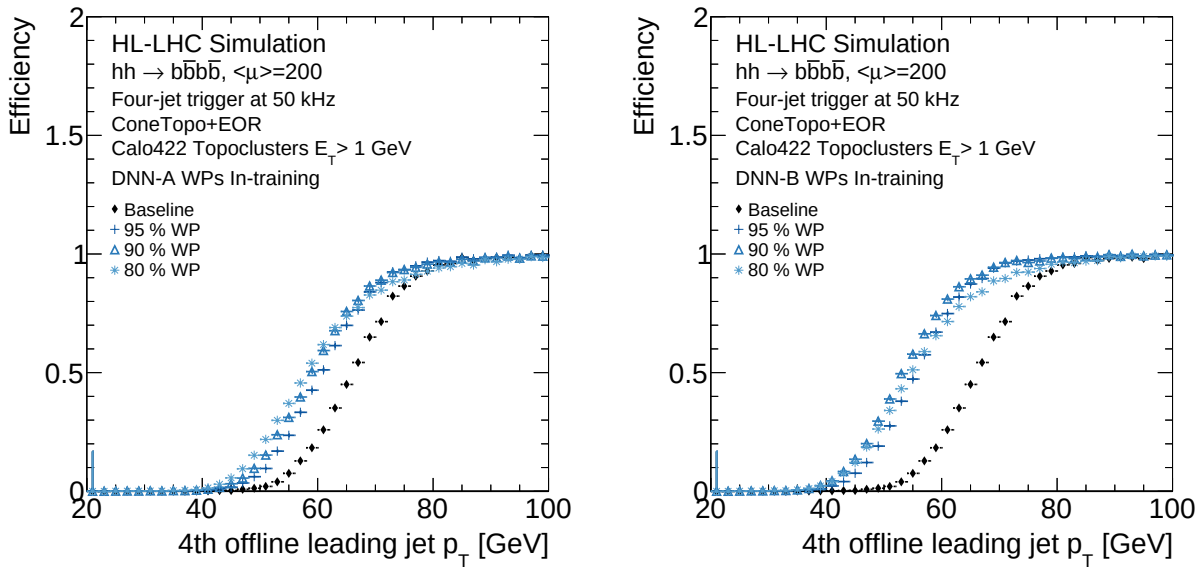
(a) (b) (c)

Figure 8.31: Effect of WPs cuts on fourth-leading jet p_T distribution for DNN-B. From left to right: minimum bias, di-Higgs, and di-jet ($[400 - 800]$ GeV truth p_T slice).



(a) (b)

Figure 8.32: Trigger rates comparing DNN-pileup-suppressed jet collections using 80% and 95% WPs with baseline jet collection with no DNN cut applied. For each WP, the rates are compared for 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).



(a) (b)

Figure 8.33: Trigger efficiencies at fixed rate threshold comparing DNN-pileup-suppressed jet collections with baseline jet collection with no DNN cut applied. The results are shown for DNN-A (left) and DNN-B (right). The DNN cut is applied only to in-training jets.

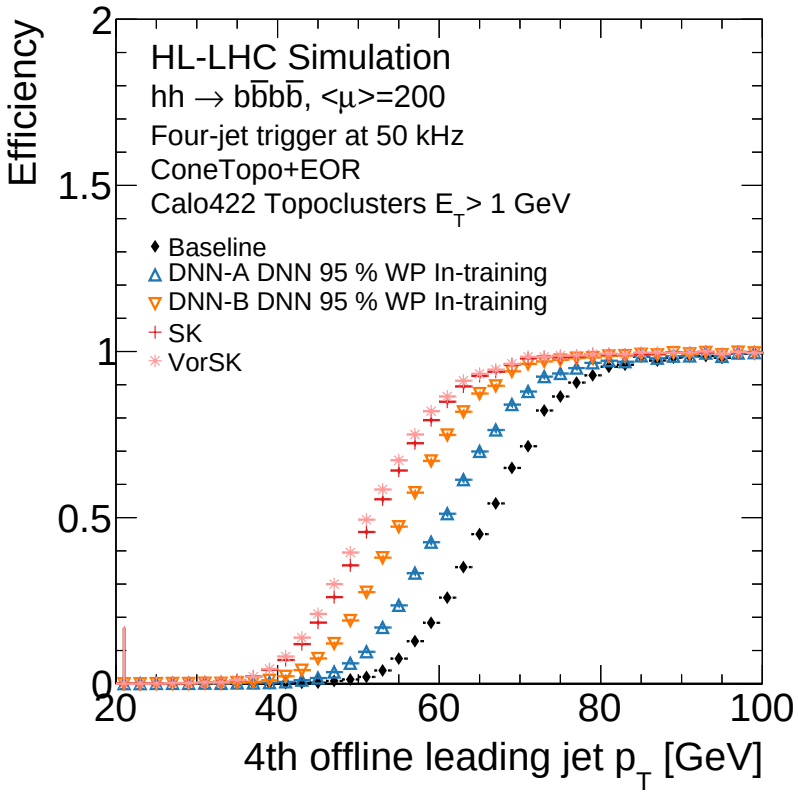


Figure 8.34: Trigger efficiencies at fixed rate threshold comparing DNN-pileup-suppressed jet collections with 95% WP for DNN-A and DNN-B, as well as the baseline jet collection, and the jet collection produced with the same `ConeTopo+EOR` algorithm run on SK and Vor+SK pileup suppressed topocluster collections.

Chapter 9

Conclusion and outlook

Search for heavy resonances

The search for heavy resonances has been the focus of intense efforts from the ATLAS Collaboration in looking for physics beyond the Standard Model. 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 in more than 60% of the cases W/Z/H bosons decay hadronically into a pair of quarks, boosted jet tagging plays an essential role in these searches.

The work performed in this thesis contributed to the search for such new heavy resonances decaying to 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 has been 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 QCD. 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 both the training of the DNN, 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 and allows to recover up to 20% loss in sensitivity 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, as well as in other similar searches for heavy resonances.

Phase II Trigger upgrade

The LHC will soon undergo a major upgrade of the accelerator that will raise the center-of-mass energy to $\sqrt{s} = 14\,\text{TeV}$ and will deliver an instantaneous luminosity of $5\times10^{34}\text{cm}^{-2}\text{s}^{-1}$. The High-Luminosity LHC (HL-LHC) will bring a ten-fold increase in the data collected by ATLAS, which will extend the experiment physics reach beyond the original design. However, with more luminosity comes more pileup and more radiation, so that the ATLAS detector and TDAQ system will have to be substantially upgraded to face the harsher conditions. A significant part of this work was to contribute to the Phase II upgrade of the hardware-based trigger system in preparation for the HL-LHC. This included the development and mainte-

nance of a software simulation framework for the development of new firmware algorithms, as well as the development of a new jet reconstruction 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, cone jets have been shown to provide equivalent performance both in terms of trigger rates and trigger efficiencies, as well as in terms of signal efficiencies in the offline analysis for the case of a di-Higgs signal. Several parameters had to be optimized and their effect and correlations had to be understood to arrive to this result. Several viable options have been identified with different advantages and drawbacks. The final decision 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. Therefore, while a finalized set of parameters will evolve with time, this work showed what choices and trade-offs will need to be addressed before arriving to the final version. The certain conclusion is that the cone jet algorithm is a viable option for the Global Trigger, with possible interesting extensions 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 $HH \to b\bar{b}b\bar{b}$. 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.

In this work, a novel deep-learning technique was studied for pileup mitigation at the jet level. Two neural networks were trained for the identification and removal of pileup

jets using only information about the constituent topological clusters, motivated by the fact that pileup jets have a more diffuse energy pattern uncorrelated from the hard scatter and a more even energy sharing among the jet constituents. One model, DNN-A, used only energy and spatial information of the topological clusters, while the second model, DNN-B, was provided with a Boolean output of the Soft-Killer pileup suppression algorithm for each constituent. Both DNNs were found to reduce the rate of background pileup jets. However, to improve the trigger efficiencies information from offline pileup suppression was needed. The conclusion therefore is that, with the current implementation, to observe a substantial improvement in trigger efficiencies an event-level characterization of the pileup density is needed. However, this study demonstrated that significant discrimination power is available in the energy profile of the jet constituents so this could be 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 DNN for pileup-jet rejection are just two examples of the many applications where deep learning algorithms can bring significant improvement.

APPENDICES

Appendix A

Analysis Appendix

Signal efficiencies in 0 and 1 lepton channel

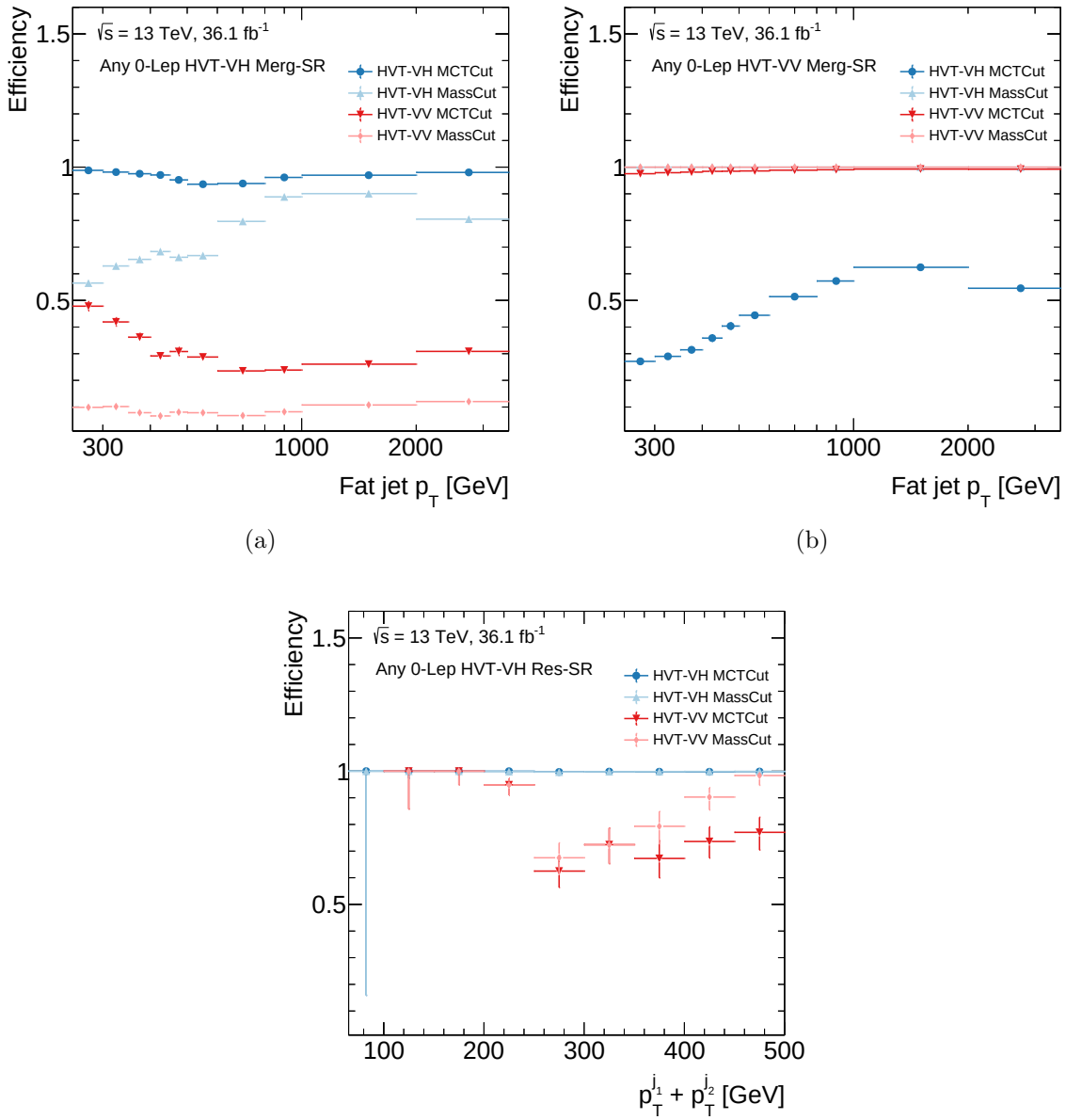


Figure A.1: Efficiency as a function of the two leading small-R jets of HVT-VH and HVT-VV signals in the 0-lepton channel in (c) HVT-VH Res SR, (a) HVT-VH Merg SR, and (b) HVT-VV Merg SR.

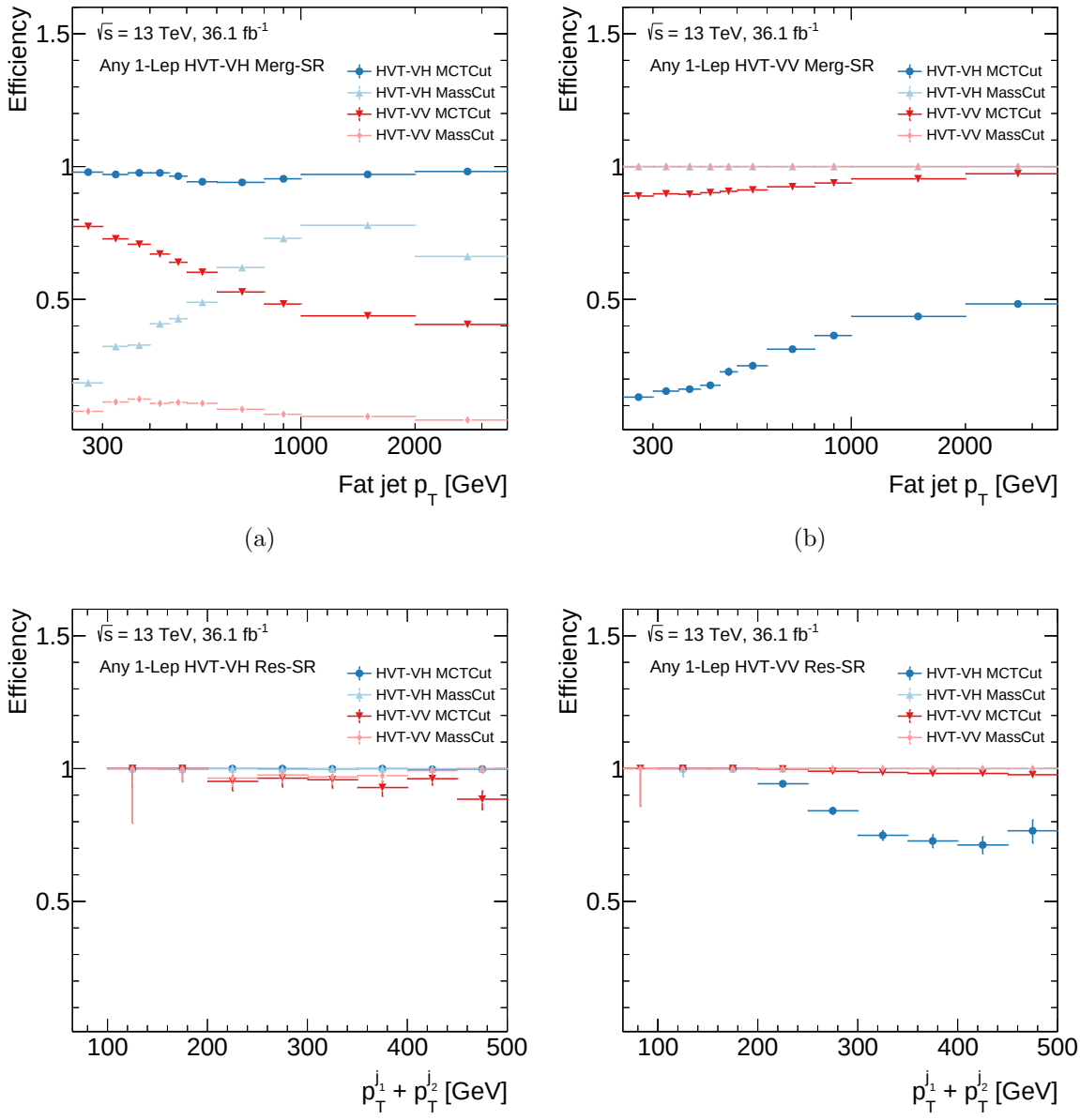


Figure A.2: Efficiency as a functionof the two leading small-R jets of HVT-VH and HVT-VV signals in the 1-lepton channel in (c) HVT-VH Res SR, and (d) HVT-VV Res SR, (a) HVT-VH Merg SR, and (b) HVT-VV Merg SR.

MCT Modeling in 0 and 1 lepton pre-selection regions

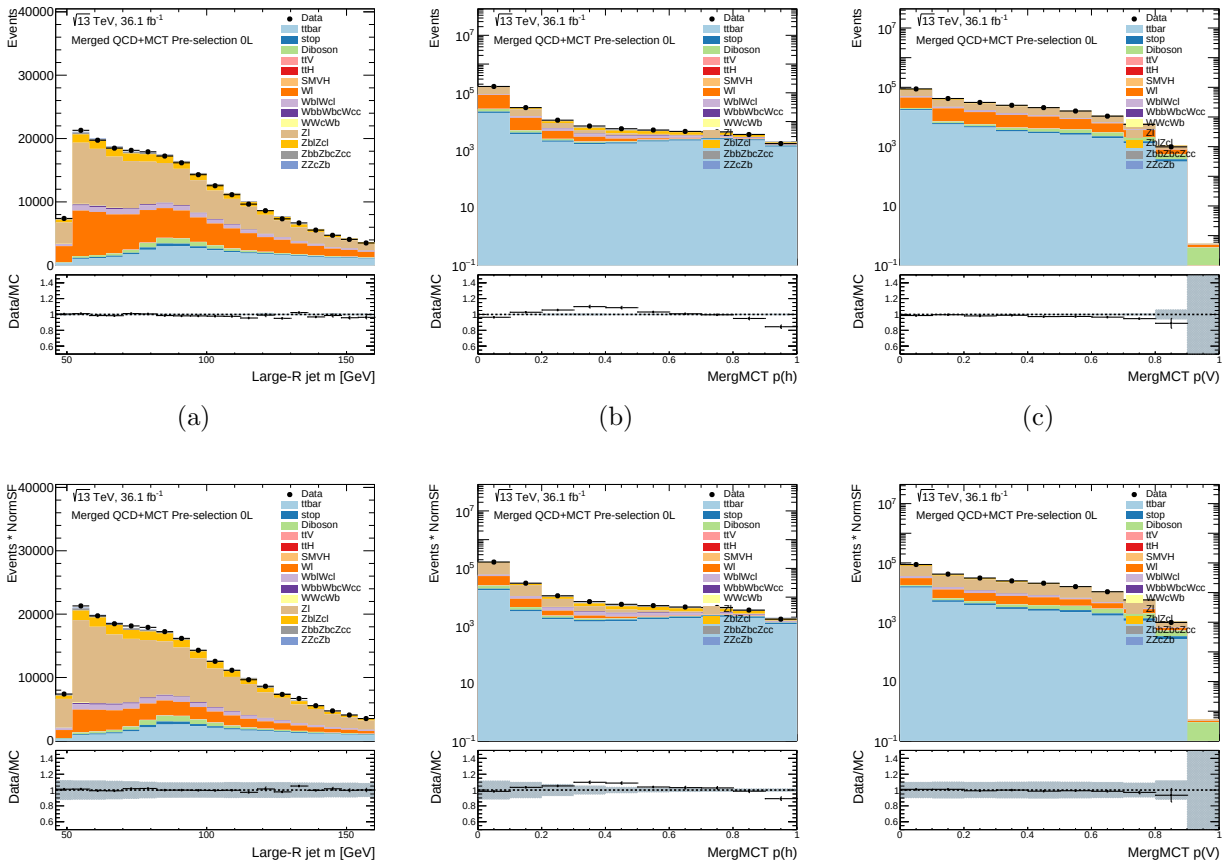


Figure A.3

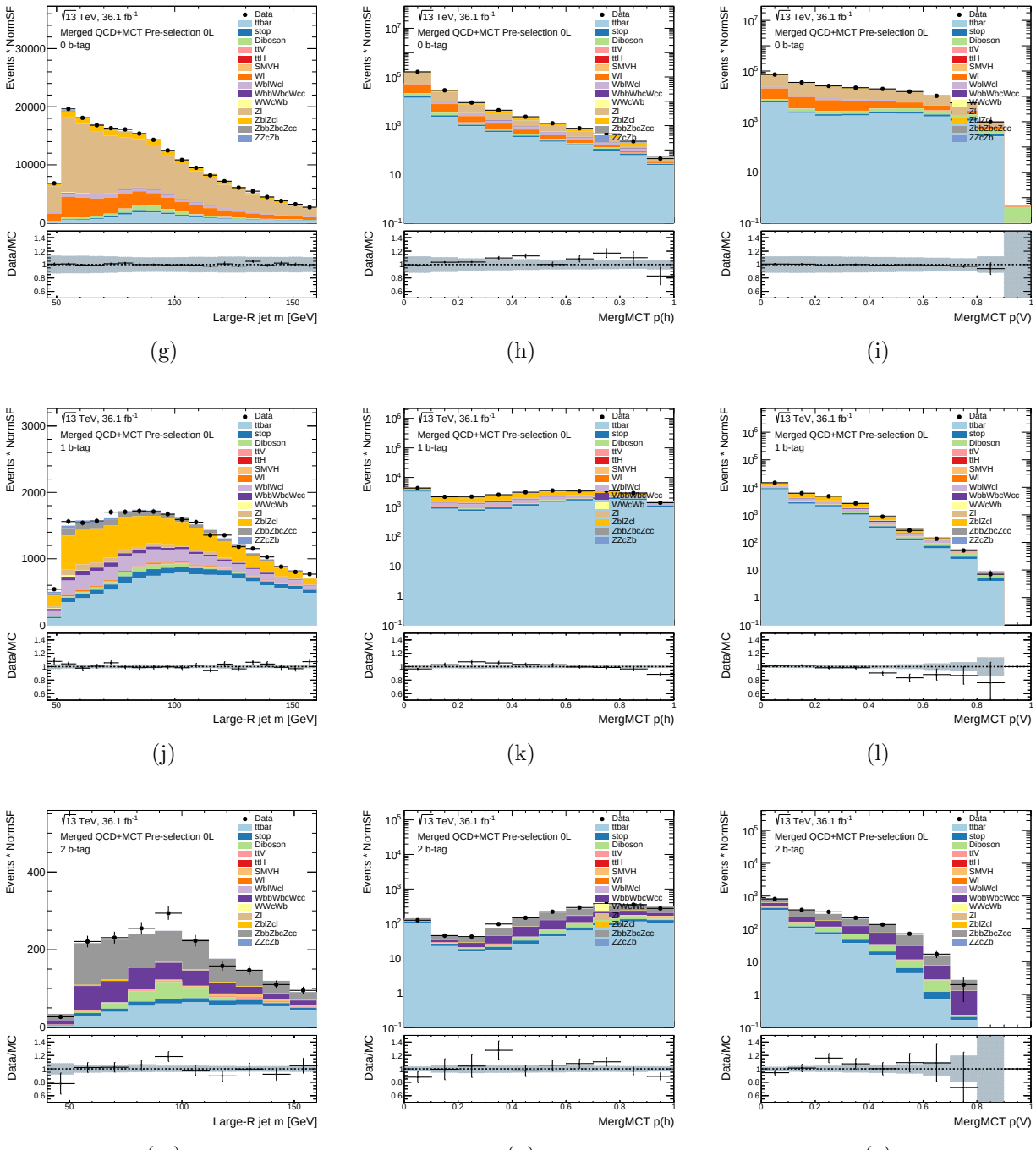


Figure A.3: Data and MC comparison in Merged pre-selection region in the 0-lepton channel. The large-R jet mass and the raw merged MCT scores p(h) and p(V) are shown before applying the normalization SFs (a), (b), (c) and after in the inclusive (d), (e), (f), 0-btag (g), (h), (i), 1-btag (j), (k), (l), and 2-btag (m),(n), (o) region.

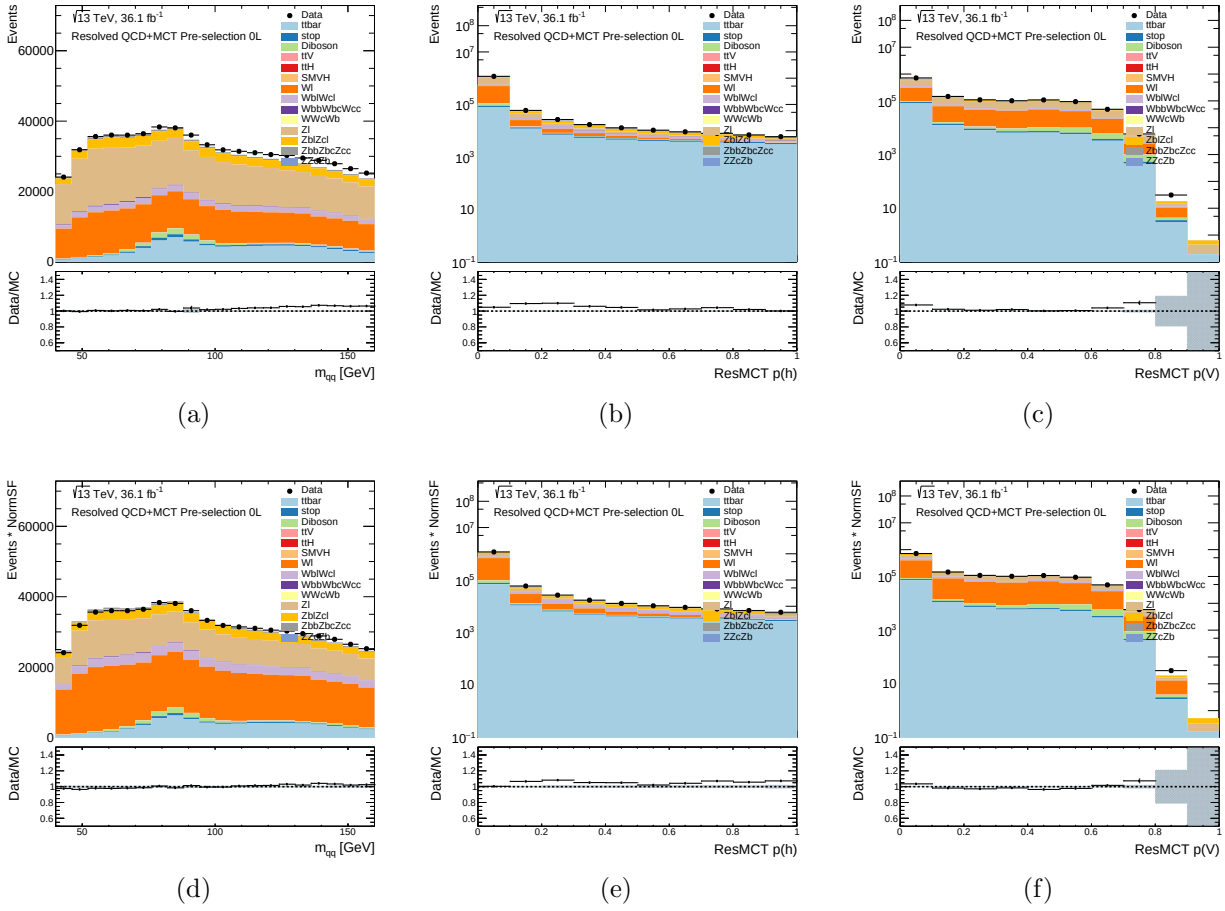
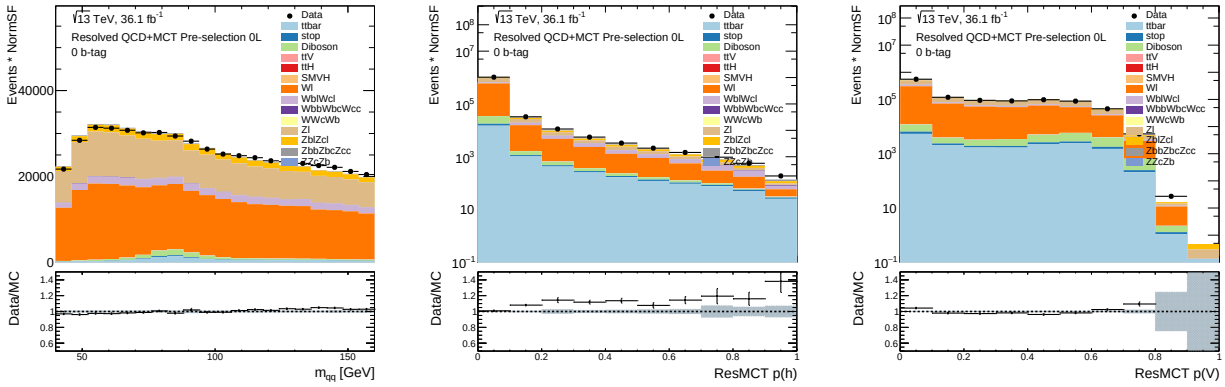
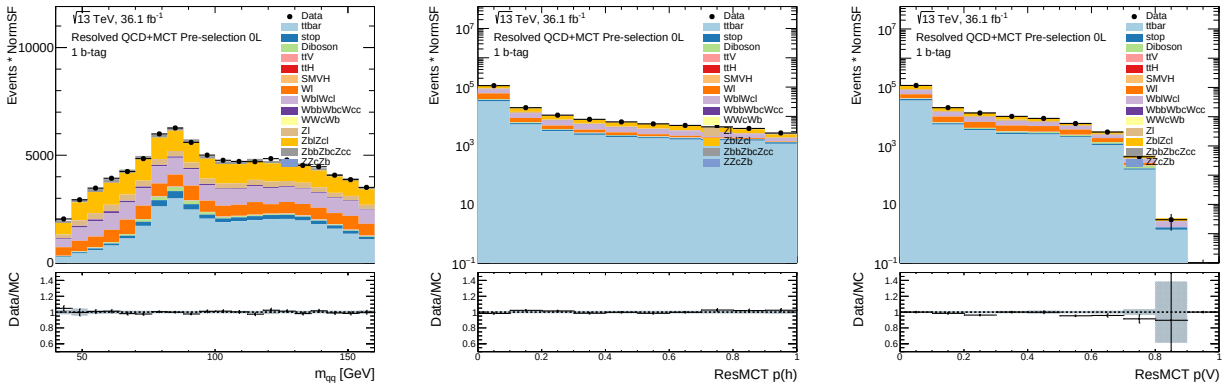

√13 TeV, 36.1 fb-1
Resolved QCD+MCT Pre-selection 0L
Data
ttbar
stop
Diboson
ttV
ttH
SMVH
Wl
WblWcl
WbbWbcWcc
WWcWb
Zl
ZblZcl
ZbbZbcZcc
ZZcZb
Events
Events * NormSF
Data/MC
mqq [GeV]
ResMCT p(h)
ResMCT p(V)
(a)
(b)
(c)
(d)
(e)
(f)

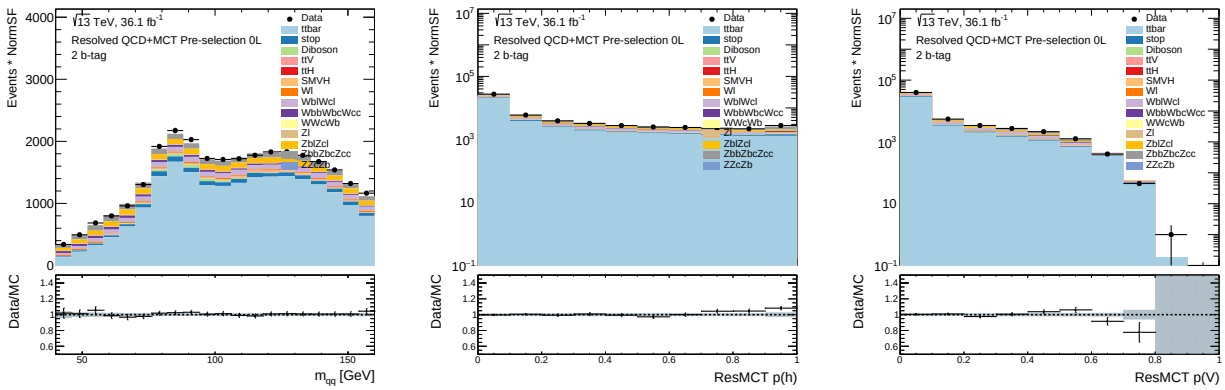
Figure A.4



(g) (h) (i)

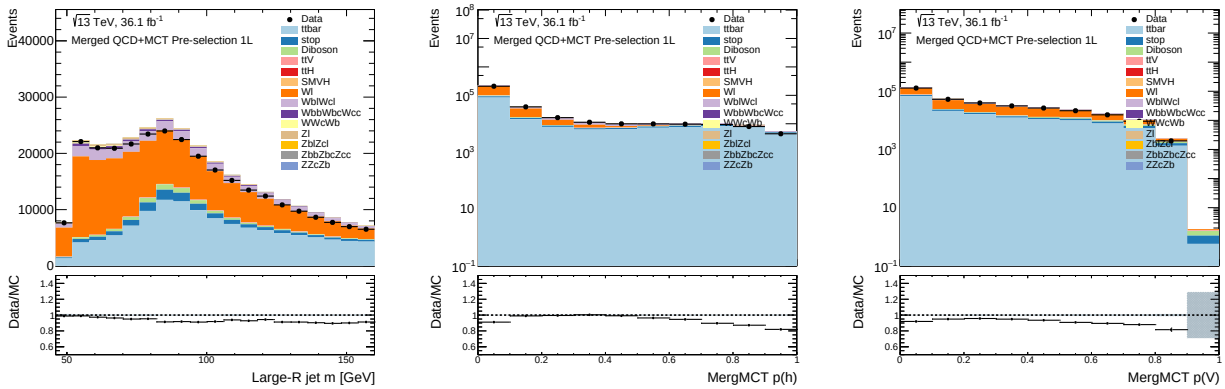


(j) (k) (l)

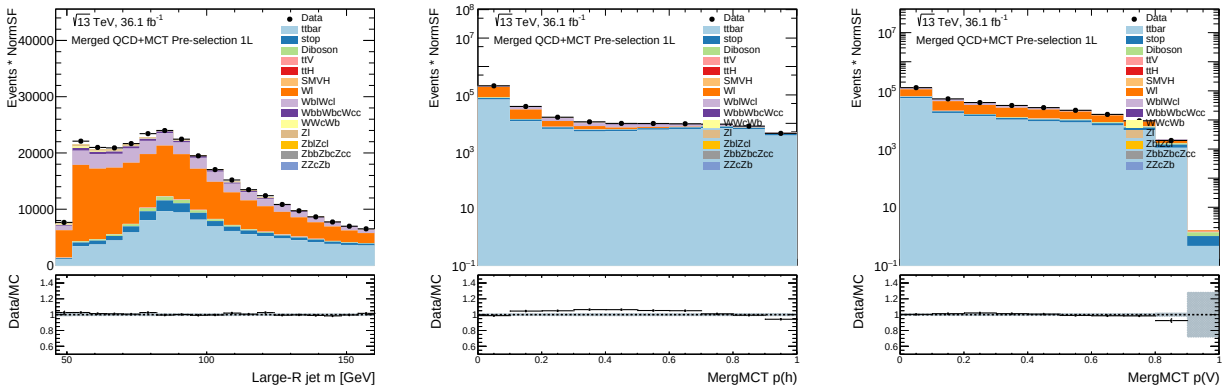


(m) (n) (o)

Figure A.4: Data and MC comparison in Merged pre-selection region in the 0-lepton channel. The large-R jet mass and the raw merged MCT scores p(h) and p(V) are shown before applying the normalization SFs (a), (b), (c) and after in the inclusive (d), (e), (f), 0-btag (g), (h), (i), 1-btag (j), (k), (l), and 2-btag (m),(n), (o) region.



(a) (b) (c)



(d) (e) (f)

Figure A.5

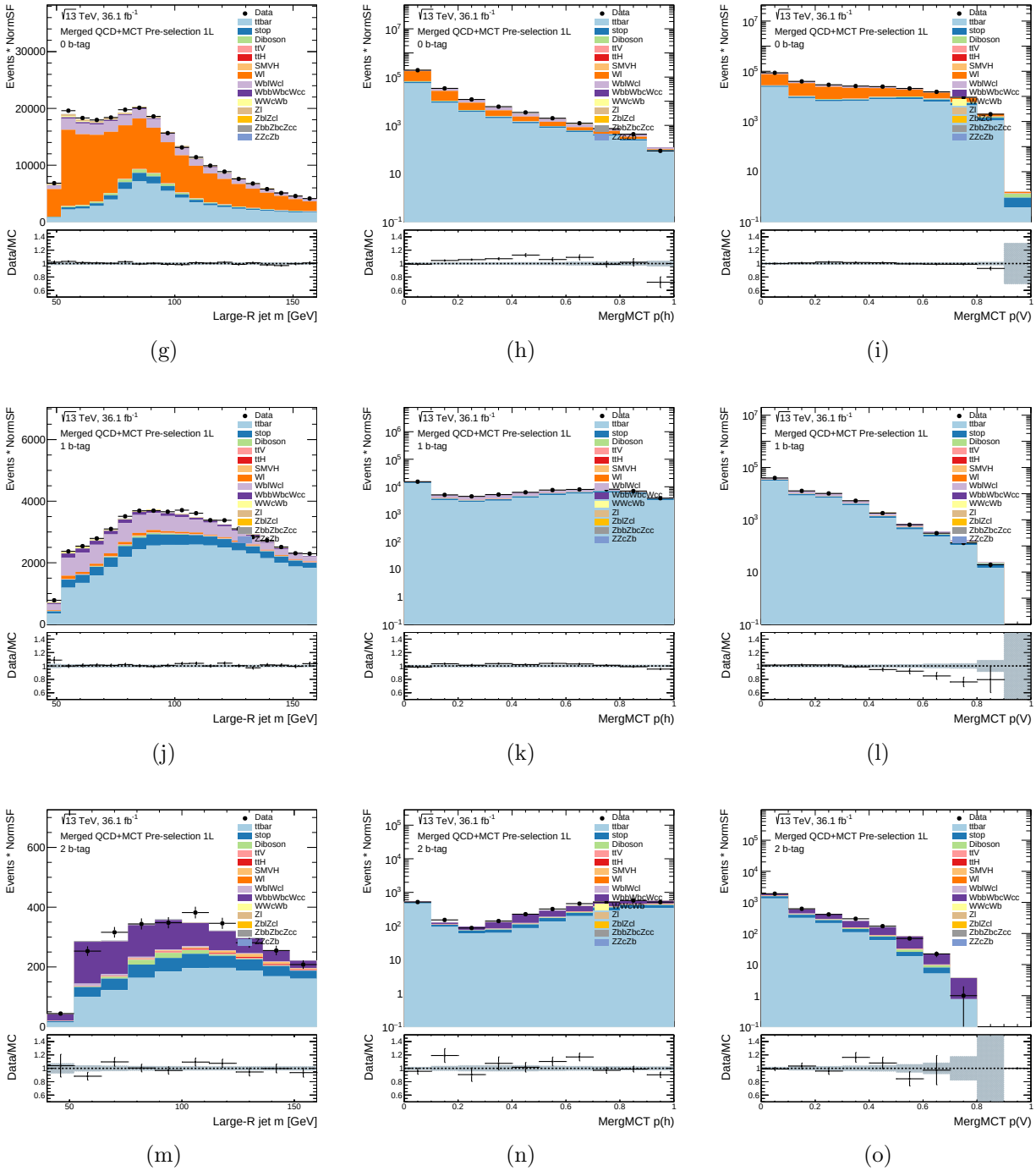
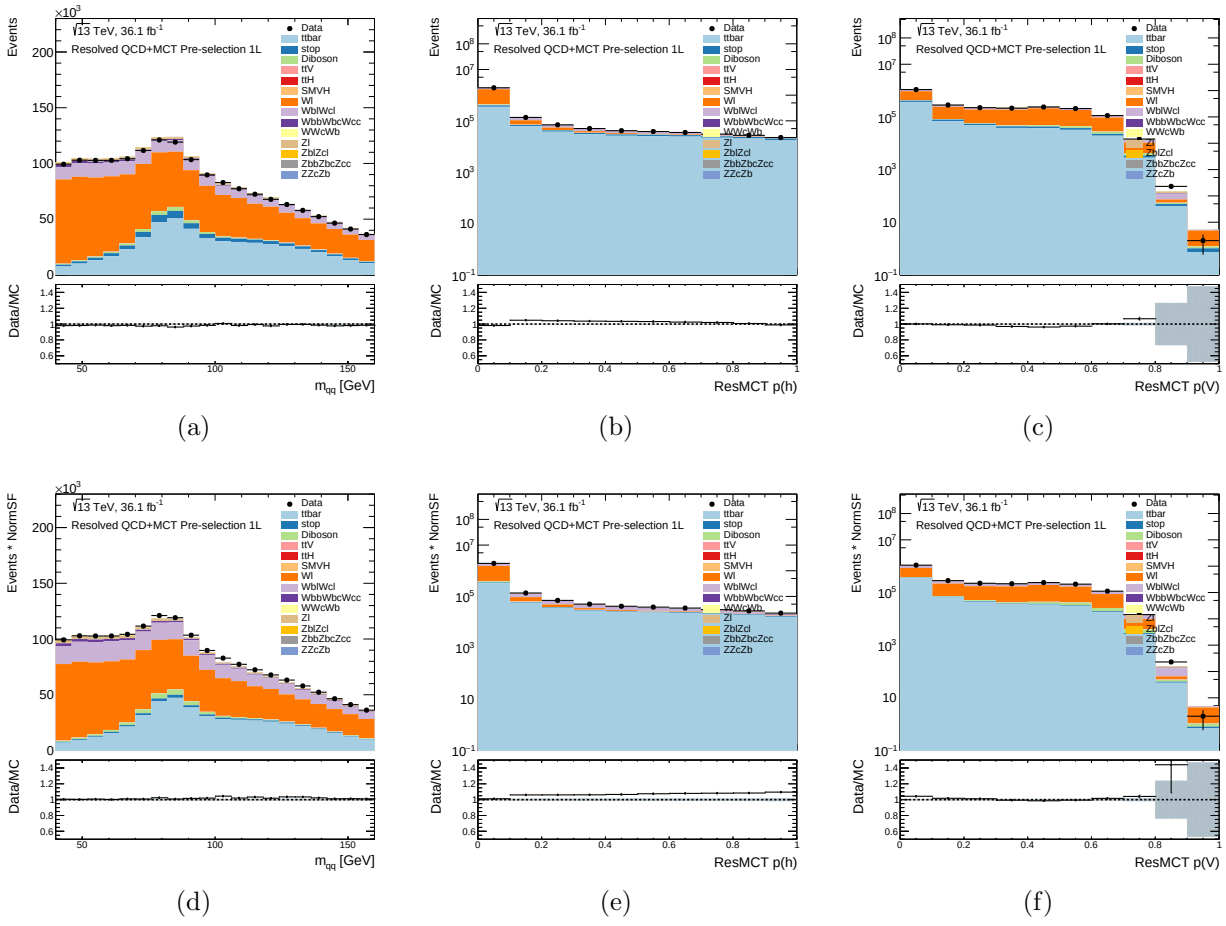


Figure A.5: Data and MC comparison in Merged pre-selection region in the 1-lepton channel. The large-R jet mass and the raw merged MCT scores p(h) and p(V) are shown before applying the normalization SFs (a), (b), (c) and after in the inclusive (d), (e), (f), 0-btag (g), (h), (i), 1-btag (j), (k), (l), and 2-btag (m),(n), (o) region.



(a) (b) (c)

(d) (e) (f)

Figure A.6

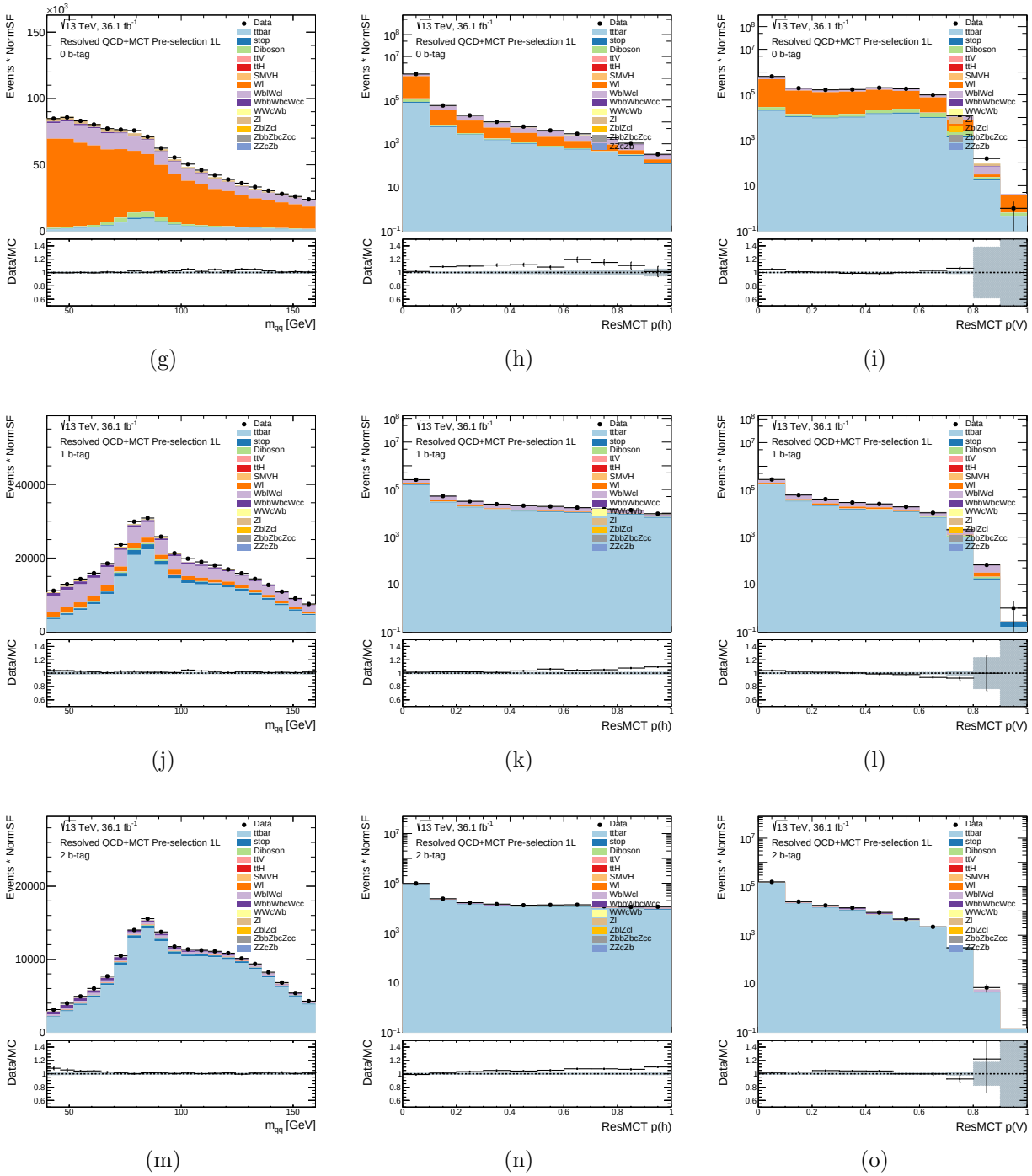


Figure A.6: Data and MC comparison in Merged pre-selection region in the 1-lepton channel. The large-R jet mass and the raw merged MCT scores p(h) and p(V) are shown before applying the normalization SFs (a), (b), (c) and after in the inclusive (d), (e), (f), 0-btag (g), (h), (i), 1-btag (j), (k), (l), and 2-btag (m),(n), (o) region.

Appendix B

Trigger Appendix

Cone jets performance with jFEX seeding

B.0.0.1 JFex seeds and nearby-jets

A short study checked the use of jFEX trigger objects (TOBs) as seeds for the cone jet algorithm. The jFEX TOB is taken as seed and the cone of radius $R = 0.4$ is built around it. The input topoclusters are thresholded according to the specified E_T cut. The resulting jet collection is referred to as `ConeJFEX`. The jFEX algorithm (see Sec. 4.3.4) is characterized by a minimum distance requirement between the towers seeding the jFEX objects. This makes them an interesting example of plateau inefficiency due to the inability to resolve nearby jets. The trigger efficiencies are built using a common online p_T cut of 30 GeV, which allowed to overlay the turn-on curves for better comparison, but with no particular physical meaning.

The performance of the `ConeJFEX` collection is compared to the `ConeTopo` and `AntiKt422` jets. Fig. B.1 shows trigger efficiencies for a di-Higgs and Z' signal sample for one jet and multi-jet triggers. While the performance is similar for the di-Higgs signal, the Z' sample shows a plateau inefficiency. The cause of this is identified by separating the events according to how isolated the n^{th}offline leading jet is. Fig. B.2 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 distance between any pair of the four leading jets is a dR

value between 0.4 and 0.6. Clearly, the plateau inefficiency is due to the first bin, which has jets closer than dR = 0.6. No plateau inefficiency is observed for cone jets.

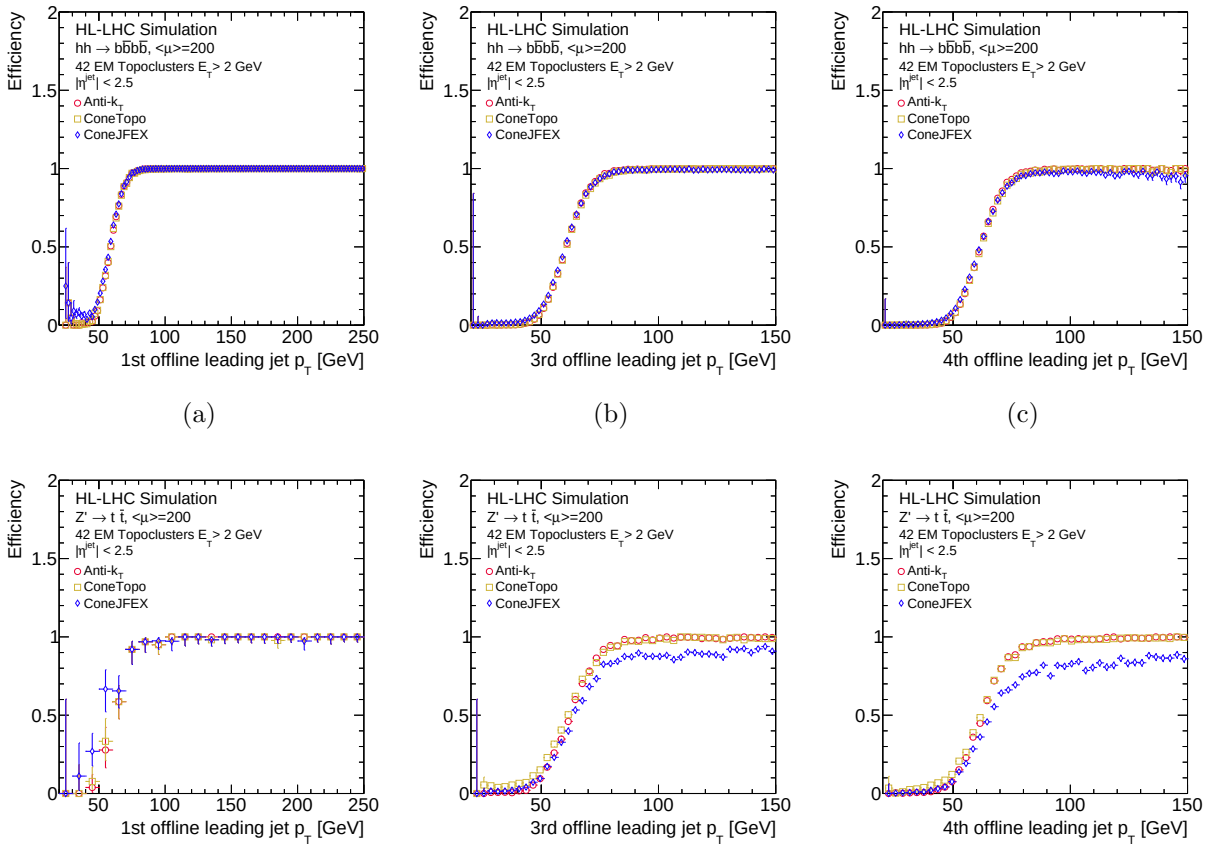
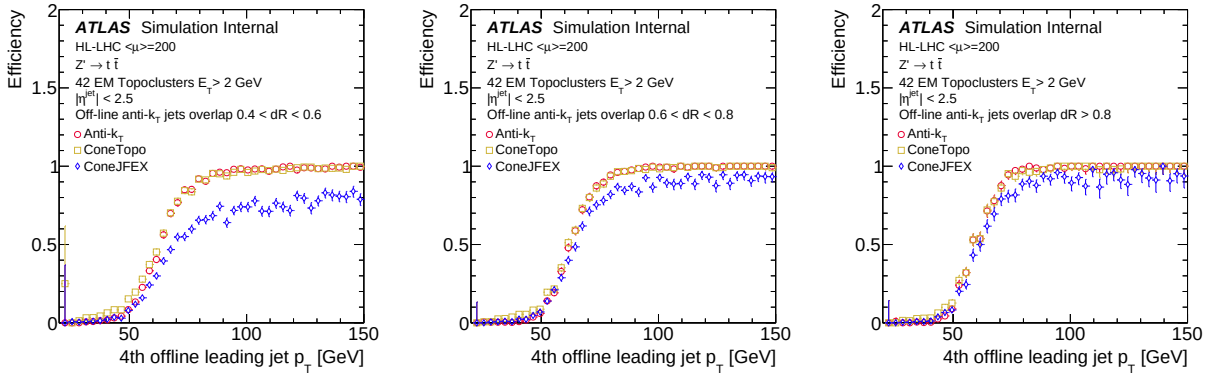


Figure B.1: Compare trigger efficiencies using `ConeTopo`,`AntiKt422`, and `ConeJFEX` for di-Higgs (top) and $Z' \to t\bar{t}$ (bottom) signals and an input topoclusters E_T threshold of $2GeV$.



(a) (b) (c)

Figure B.2: Compare trigger efficiencies in different offline jet isolation bins using `ConeTopo`,`AntiKt422`, and `ConeJFEX` for di-Higgs (top) and $Z' \rightarrow t\bar{t}$ (bottom) signals and an input topoclusters E_T threshold of $2GeV$.

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*, pages 155–178. WORLD SCIENTIFIC, jan 2008.

[3] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer. Heavy Vector Triplets: bridging theory and data. *J. High Energ. Phys.*, 09:060, 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 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 Collab. Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System. Technical report, CERN, Geneva, 2017.

[11] Wikipedia. Standard model of elementry particles. *Wikipedia, The Free Encyclopedia*.

[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*. Wiley, 2nd edition, 2010.

[14] A. Zee. *Quantum Field Theory in a Nutshell*. Princeton University Press, 2nd edition, 2010.

[15] D. Tong. Lectures on Quantum Field Theory.

[16] Ian J. R. Aitchison and Anthony J. G. Hey. *Gauge Theories in Particle Physics: A Practical Introduction, Volume 1 : From Relativistic Quantum Mechanics to QED, Fourth Edition*. Taylor & Francis, 2013.

[17] A. Djouadi. The anatomy of electro–weak symmetry breaking. Tome I: The Higgs boson in the Standard Model. *Phys. Rep.*, 457:1, 2008.

[18] S. L. Glashow. Partial-symmetries of weak interactions. *Nucl. Phys.*, 22:579, 1961.

[19] P. W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508, 1964.

[20] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321, 1964.

[21] S. Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264, 1967.

[22] A. Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367, 1968.

[23] 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.

[24] 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.

[25] P. A. M. Dirac. The quantum theory of the emission and absorption of radiation. *Proc. R. Soc. Lond. A*, 114:243, 1927.

[26] ATLAS Collab. Combined measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ decay channels with the ATLAS detector using $\sqrt{s} = 7$, 8, and 13 TeV pp collision data. 2023.

[27] N. Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531, 1063.

[28] M. Kobayashi and T. Maskawa. CP violation in the renormalizable theory of weak Interaction. *Prog. Theor. Phys.*, 49:652, 1973.

[29] M. Tanabashi et al. Review of Particle Physics*. *Phys. Rev. D*, 98:030001, 2018.

[30] ATLAS Collab. A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery. *Nature*, 607:52, 2022.

[31] R. Sancisi and T. S. van Albada. Dark matter. *Observational Cosmology.*, 124:699, 1987.

[32] R. Massey, T. Kitching, and J. Richard. The dark matter of gravitational lensing. *Rep. Prog. Phys.*, 73:086901, 2010.

[33] Y. Fukuda and Super-Kamiokande Collaboration. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81:1562, 1998.

[34] 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.

[35] LHCb Collab. Test of lepton universality in beauty-quark decays. *Nat. Phys.*, 18:277, 2022.

[36] Muon g 2 Collab. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.*, 126:141801, 2021.

[37] J. Ellis. Higgs Physics. *arXiv*, 2015.

[38] 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.

[39] J. M. Lizana and M. Pérez-Victoria. Vector triplets at the LHC. *EPJ Web of Conferences*, 60:17008, 2013.

[40] J. de Blas, J. M. Lizana, and M. Pérez-Victoria. Combining searches of Z' and W' bosons. *J. High Energ. Phys.*, 2013:166, 2013.

[41] M. Perelstein. Little Higgs models and their phenomenology. *Prog. in Part. and Nucl. Phys.*, 58:247, 2007.

[42] M. J. Dugan, H. Georgi, and D. B. Kaplan. Anatomy of a composite Higgs model. *Nucl. Phys. B*, 254:299, 1985.

[43] K. Agashe, R. Contino, and A. Pomarol. The minimal composite Higgs model. *Nucl. Phys. B*, 719:165, 2005.

[44] A. Djouadi. The anatomy of electro–weak symmetry breaking. Tome II: The Higgs bosons in the Minimal Supersymmetric Model. *Phys. Rep.*, 459:1, 2008.

[45] J. E. Kim. Light pseudoscalars, particle physics and cosmology. *Phys. Rep.*, 150:1, 1987.

[46] M. Joyce and T. Prokopec and N. Turok. Nonlocal electroweak baryogenesis. Part 2: The Classical regime. *Phys. Rev.*, D53:2958, 1996.

[47] S. L. Glashow and S. Weinberg. Natural conservation laws for neutral currents. *Phys. Rev. D*, 15:1958, 1977.

[48] E. A. Paschos. Diagonal neutral currents. *Phys. Rev. D*, 15:1966, 1977.

[49] ATLAS Collab. Constraints on new phenomena via Higgs boson couplings and invisible decays with the ATLAS detector. *J. High Energ. Phys.*, 2015:206, 2015.

[50] 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.

[51] P. J. Bryant. A brief history and review of accelerators. In *CERN Accelerator School: Course on General Accelerator Physics*, pages 1–16, 1992.

[52] Report of the Long Range Planning Committee to the CERN Council. 83rd Session of Council. 1987.

[53] R. Assmann, M. Lamont, and S. Myers. A brief history of the LEP collider. *Nucl. Phys. B Proc. Suppl.*, 109:17, 2002.

[54] C. L. Smith. Genesis of the Large Hadron Collider. *Phil. Trans. R. Soc. A*, 373:20140037, 2015.

[55] L. Evans and P. Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[56] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[57] The CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[58] The LHCb Collaboration. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[59] The ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[60] J. P. Blewett. 200-GeV Intersecting Storage Accelerators. *eConf*, C710920:501, 1971.

[61] F. Landua. The CERN accelerator complex layout in 2022. 2022. General Photo.

[62] D. Boussard and T. P. Linnecar. The LHC Superconducting RF System. Technical report, CERN, Geneva, 1999.

[63] S. Baird. Accelerators for pedestrians; rev. version. Technical report, CERN, Geneva, 2007.

[64] S. van der Meer. Calibration of the effective beam height in the ISR. Technical report, CERN, Geneva, 1968.

[65] 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, Geneva, 2004.

[66] G. Soyez. Pileup mitigation at the LHC: A theorist's view. *Physics Reports*, 803:1, 2019.

[67] O. Aberle et al. High-Luminosity Large Hadron Collider (HL-LHC): Technical design report. 2020.

[68] 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.

[69] B. Schmidt. The High-Luminosity upgrade of the LHC: Physics and technology challenges for the accelerator and the experiments. *J. Phys.: Conf. Ser.*, 706:022002, 2016.

[70] 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(1):353–356, 2000.

[71] J. Pequenao. Computer generated image of the ATLAS inner detector. 2008.

[72] The ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider: A description of the detector configuration for Run 3.

[73] ATLAS Collab. ATLAS pixel detector electronics and sensors. *JINST*, 3:P07007, 2008.

[74] ATLAS Collab. The silicon microstrip sensors of the ATLAS semiconductor tracker. *Nucl. Instrum. Meth. A*, 578:98, 2007.

[75] B. Mindur. 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.

[76] M. Capeans, G. Darbo, K. Einsweiller, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet. ATLAS Insertable B-Layer Technical Design Report. Technical report, 2010.

[77] ATLAS Collab. Production and integration of the ATLAS Insertable B-Layer. *JINST*, 13:T05008, 2016.

[78] ATLAS Collab. Technical Design Report for the ATLAS Inner Tracker Strip Detector. Technical report, CERN, Geneva, 2017.

[79] ATLAS Collab. Technical Design Report for the ATLAS Inner Tracker Pixel Detector. Technical report, CERN, Geneva, 2017.

[80] C. Grupen and B. Shwartz. *Particle detectors*. Cambridge University Press, 2nd edition, 2008.

[81] C. W. Fabjan and F. Gianotti. Calorimetry for particle physics. *Rev. Mod. Phys.*, 75:1243, 2003.

[82] ATLAS Collab. *ATLAS tile calorimeter: Technical Design Report*. CERN, Geneva, 1996.

[83] S. Palestini. The muon spectrometer of the ATLAS experiment. *Nucl. Phys. B - - Proc. Suppl.*, 125:337, 2003.

[84] J. Sterling. Private communication.

[85] ATLAS Collab. Performance of the ATLAS trigger system in 2015. *Eur. Phys. J. C*, 77:317, 2017.

[86] ATLAS Collaboration. Operation of the ATLAS Trigger System in Run 2. *JINST*, 15:P10004, 2020.

[87] ATLAS Collab. Performance of the ATLAS muon triggers in Run 2. *JINST*, 15:P09015, 2020.

[88] R. Achenbach et al. The atlas level-1 calorimeter trigger. *JINST*, 3:P03001, 2008.

[89] R. Simoniello. The ATLAS Level-1 Topological Processor: from design to routine usage in Run-2. Technical report, CERN, Geneva, 2019.

[90] ATLAS Collab. Athena, apr 2019.

[91] M. Cacciari, G. Salam, and G. Soyez. The anti-k$_t$ jet clustering algorithm. *J. High Energ. Phys.*, 04:063, 2008.

[92] ATLAS Collab. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C*, 77:490, 2017.

[93] ATLAS Collab. Trigger monitoring and rate predictions using enhanced bias data from the atlas detector at the lhc. Technical report, CERN, Geneva, 2016.

[94] ATLAS Collab. Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System. Technical report, CERN, Geneva, 2013.

[95] .

[96] A. Hoecker. Physics at the LHC Run-2 and beyond. In *Proceedings of the 2016 European School of High-Energy Physics*. CERN, 2016.

[97] J. Pequenao. Event Cross Section in a computer generated image of ATLAS detector. 2008.

[98] ATLAS Collab. Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2. *Eur. Phys. J. C*, 77:673, 2017.

[99] ATLAS Collab. Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC. *Eur. Phys. J. C*, 77:332, 2017.

[100] ATLAS Collaboration. Characterization of Interaction-Point Beam Parameters Using the *pp* Event-Vertex Distribution Reconstructed in the ATLAS Detector at the LHC. ATLAS-CONF-2010-027, 2010.

[101] ATLAS Collab. 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.

[102] ATLAS Collab. 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.

[103] ATLAS Collab. Performance of missing transverse momentum reconstruction with the atlas detector using proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C*, 78:903, 2018.

[104] ATLAS Collab. Object-based missing transverse momentum significance in the ATLAS detector. Technical report, CERN, Geneva, 2018.

[105] ATLAS Collab. Performance of *b*-jet identification in the ATLAS experiment. *JINST*, 11:P04008, 2016.

[106] P.A. Zyla et al. Review of Particle Physics*. *PTEP*, 2020:083C01, 2020.

[107] ATLAS Collab. Technical report.

[108] ATLAS Collab. ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset. *Eur. Phys. J. C*, 83:681, 2023.

[109] ATLAS Collab. Identification of jets containing *b*-hadrons with recurrent neural networks at the atlas experiment. Technical report, CERN, Geneva, 2017.

[110] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert3, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 2009:007, 2009.

[111] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020. and 2021 update.

[112] G. P. Salam. Elements of QCD for hadron colliders. In *2009 European School of High-Energy Physics*, 2010.

[113] P. Skands. Introduction to qcd. lectures given at tasi 2012.

[114] T. Carli, K. Rabbertz, and S. Schumann. *Studies of Quantum Chromodynamics at the LHC*, pages 139–194. Springer International Publishing, Cham, 2015.

[115] 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(5):145, 2011.

[116] S. Agostinelli et al. Geant4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 503:250, 2003.

[117] G. P. Salam G. Soyez. A practical seedless infrared-safe cone jet algorithm. *J. High Energ. Phys.*, 05:086, 2007.

[118] G. P. Salam. Towards jetography. *Eur. Phys. J. C*, 67:637, 2010.

[119] S. D. Ellis and D. E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160, 1993.

[120] M. Seymour S. Catani, Y. Dokshitzer and B. Webber. Longitudinally-invariant k_T-clustering algorithms for hadron-hadron collisions. *Nucl. Phys. B*, 406:187, 1993.

[121] Yu.L. Dokshitzer, G.D. Leder, S. Moretti, and B.R. Webber. Better jet clustering algorithms. *J. High Energ. Phys.*, 1997:001, 1997.

[122] M. Wobisch and T. Wengler. Hadronization corrections to jet cross-sections in deep inelastic scattering. 1999.

[123] M. Cacciari and G. P. Salam. Dispelling the N3 myth for the k_t jet-finder. *Phys. Lett. B*, 641:57, 2006.

[124] ATLAS Collab. Jet reconstruction and performance using particle flow with the ATLAS detector. *Eur. Phys. J. C*, 77:466, 2017.

[125] ATLAS Collab. Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector. *Eur. Phys. J. C*, 81:689, 2021.

[126] ATLAS Collab. 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.

[127] M. Cacciari, G. P. Salam, and G. Soyez. FastJet User Manual. *Eur. Phys. J. C*, 72:1896, 2012.

[128] ATLAS Collab. Improving jet substructure performance in ATLAS using TrackCaloClusters. 2017.

[129] ATLAS Collab. Optimization of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions. *Eur. Phys. J. C*, 81:334, 2021.

[130] ATLAS Collab. Variable Radius, Exclusive-k_T, and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS. Technical report, CERN, 2017.

[131] A. J. Larkoski, I. Moult, and B. Nachman. Hadron Collider: A review of recent advances in theory and machine learning. *Phys. Rep.*, 841:1, 2020.

[132] R. Kogler et al. Jet substructure at the Large Hadron Collider: Experimental review. *Rev. Mod. Phys.*, 91:045003, 2019.

[133] A. J. Larkoski, I. Moult, 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.

[134] J. Thaler and K. Van Tilburg. Identifying boosted objects with N-subjettiness. *J. High Energ. Phys.*, 2011:15, 2011.

[135] A. J. Larkoski, G. P. Salam, and J. Thaler. Energy correlation functions for jet substructure. *J. High Energ. Phys.*, 2013:108, 2013.

[136] A. J. Larkoski, I. Moult, and D. Neill. Power counting to better jet observables. *J. High Energ. Phys.*, 2014:9, 2014.

[137] A.J. Larkoski, S. Marzani, G. Soyez, et al. Soft drop. *J. High Energ. Phys.*, 2014:146, 2014.

[138] G. Soyez. Pileup mitigation at the LHC: A theorist's view. *Phys. Rep.*, 803:1, 2019.

[139] M. Cacciari, G. P. Salam, and G. Soyez. SoftKiller, a particle-level pileup removal method. *Eur. Phys. J. C*, 75:59, 2015.

[140] D. Bertolini, P. Harris, M. Low, and N. Tran. Pileup per particle identification. *J. High Energ. Phys.*, 2014:59, 2014.

[141] ATLAS Collab. Constituent-level pile-up mitigation techniques in atlas. Technical report, CERN, Geneva, 2017.

[142] P. Berta, M. Spousta, and D. W. Miller. Particle-level pileup subtraction for jets and jet shapes. *J. High Energ. Phys.*, 2014:92, 2014.

[143] I. J. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. `http://www.deeplearningbook.org`.

[144] 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.

[145] L. Randall and R. Sundrum. A large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.

[146] 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.

[147] ATLAS Collab. Search for heavy resonances decaying into a W or Z boson and a Higgs boson in final states with leptons and b-jets in $36\,\mathrm{fb}^{-1}\sqrt{s} = 13\,\mathrm{TeV}$ of pp collisions with the ATLAS detector. *J. High Energ. Phys.*, 2018:174, 2018.

[148] ATLAS Collab. Search for heavy resonances decaying into a Z or W boson and a Higgs boson in final states with leptons and b-jets in $139\,\mathrm{fb}^{-1}$ of pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector. *J. High Energ. Phys.*, 2023:016, 2023.

[149] ATLAS Collab. Search for WW/WZ resonance production in $l\nu qq$ final states in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector. *J. High Energ. Phys.*, 2018:42, 2018.

[150] ATLAS Collab. Searches for heavy ZZ and ZW resonances in the $llqq$ and $\nu\nu qq$ final states in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector. *J. High Energ. Phys.*, 2018:9, 2018.

[151] ATLAS Collab. 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.

[152] CMS Collab. 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.

[153] CMS Collab. 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.

[154] ATLAS Collab. 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.

[155] ATLAS Collab. 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.

[156] CMS Collab. Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at. *Eur. Phys. J. C*, 77:636, 2017.

[157] CMS Collab. 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.

[158] ATLAS Collab. Combination of searches for heavy resonances decaying into bosonic and leptonic final states using 36 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV with the atlas detector. *Phys. Rev. D*, 98:052008, 2018.

[159] ATLAS Collab. Combination of searches for heavy resonances using $139\,\text{fb}^{-1}$ of proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector. ATLAS-CONF-2022-028, 2022.

[160] CMS Collab. 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 $13\,\text{TeV}$. *Phys. Lett. B*, 774:533, 2017.

[161] CMS Collab. Combination of CMS searches for heavy resonances decaying to pairs of bosons or leptons. *Phys. Lett. B*, 798:134952, 2019.

[162] CMS Collab. Search for heavy Higgs bosons decaying to a top quark pair in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *J. High Energ. Phys.*, 2020:171, 2020.

[163] ATLAS Collab. Summary plots for beyond standard model higgs boson benchmarks for direct and indirect searches. Technical report, CERN, Geneva, 2022.

[164] A. Sherstinsky. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: nonlinear Phenomena*, 404:132306, 2020.

[165] ATLAS Collab. Performance of W/Z taggers using UFO jets in ATLAS. ATL-PHYS-PUB-2021-029, 2021.

[166] A.M. Sirunyan et al. Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques. *JINST*, 15:P06005, 2020.

[167] W. D. Goldberger and M. B. Wise. Modulus Stabilization with Bulk Fields. *Phys. Rev. Lett.*, 83:4922, 1999.

[168] W. D. Goldberger and M. B. Wise. Phenomenology of a stabilized modulus. *Phys. Lett. B*, 475:275, 2000.

[169] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni. Warped gravitons at the LHC and Beyond. *Phys. Rev. D*, 76:036006, 2007.

[170] J. Alwall, M. Herquet, F. Maltoni, et al. MadGraph 5 : Going Beyond. *J. High Energ. Phys.*, 2011:128, 2011.

[171] NNPDF Collab. Parton distributions for the LHC run II. *J. High Energ. Phys.*, 2015:40, 2015.

[172] T. Sjöstrand, S. Mrenna, and P. Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852, 2008.

[173] ATLAS Pythia 8 tunes to 7 TeV data. Technical report, CERN, Geneva, 2014.

[174] 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.

[175] ATLAS Collab. Tagging and suppression of pileup jets with the ATLAS detector. Technical report, CERN, Geneva, 2014.

[176] ATLAS Collab. Measurement of the tau lepton reconstruction and identification performance in the atlas experiment using pp collisions at $\sqrt{s} = 13$ tev. Technical report.

[177] 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.

[178] ATLAS Collab. Searches for heavy diboson resonances in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *J. High Energ. Phys.*, 2016:173, 2016.

Acronyms

2HDM Two-Higgs-Doublet Model. 2, 55–57, 59, 182, 184, 186, 195, 197

BC Bunch crossing. 281

BR branching ratio. 182, 186, 195, 208

BSM beyond the Standard Model. 2, 10, 45, 48, 50, 51, 120, 129, 141, 156, 218, 219

CERN European Organization for Nuclear Research. 10, 61, 62, 65, 73–75

CKM Cabbibo-Kobayashi-Maskawa. 42, 46

CoM center-of-mass. 127, 139

CR control region. 187–189, 191, 216, 217

CTP Central Trigger Processor. 102, 103, 111, 114, 282, 283

DNN deep neural network. 3, 4, 218, 220, 226, 235, 236, 338, 340

DY Drell-Yan. 191, 193, 195, 213

ECF energy correlation function. 159, 160

EL Euler-Lagrange. 14–16

EMCal electromagnetic calorimeter. 117

EMEC end-cap EM calorimeters. 87, 89, 91

EOR Energy Overlap Removal. 303–305

EWSB electroweak symmetry breaking. 47, 53

FCal forward calorimeter. 90, 91, 100, 101, 123, 124

FCNCs flavour changing neutral currents. 56

FPGA field programmable gate array. 2, 280, 282–284, 294, 296, 298, 313–315

GCM Global Common Module. 283

GEP Global Event Processor. 281–284

ggF gluon-gluon fusion. 191, 193–195, 213, 216, 288

GT Global Trigger. 278–280, 283, 284, 287–289, 291, 295, 301, 313–315

GWS Glashow-Weinberg-Salam. 9

HCal hadronic calorimeter. 85, 86, 89, 101

HL-LHC High Luminosity Large Hadron Collider. 1, 3, 39, 46, 49, 73, 81, 92, 106, 110, 161, 166, 280, 287, 307, 338, 340

HVT Heavy Vector Triplet. 47, 50, 51, 54, 55, 182, 184, 186, 193–197, 213, 216, 233, 247, 248, 251, 255, 257

IBL Insertable B-Layer. 79, 81

IP interaction point. 62, 69, 73, 95, 96

IRC infrared-collinear. 136, 144–147, 149–151, 159, 163, 294

LAr liquid argon. 85, 86, 89–92, 100, 106, 108, 110, 117, 200–202, 283, 284

LEP Large Electron Positron. 61, 62

MC Monte Carlo. 139, 141, 151, 152, 156, 188, 189, 196–198, 201, 266, 287, 289, 290, 307

MCT Multi-Class Tagger. 192, 193, 216, 218, 227, 233, 234, 251–255, 257, 258, 260, 261, 263–267, 272–274, 337, 338

MET Missing Energy Transverse. 200, 201

ML machine learning. 2

MPIs multi-parton interactions. 138, 139

MS muon spectrometer. 111, 119

MSSM Minimal Suppersymmetric Standard Model. 55, 184

MUCTPI Muon Central Trigger Processor Interface. 281–283

MUX Multiplexer Processor. 281, 283

NN neural network. 190, 222, 224

PDF parton distribution function. 128, 132–134, 137, 139

PV primary vertex. 116, 117

QCD quantum chromodynamics. 27, 127–132, 135, 142, 145, 146, 157, 158, 164, 165, 191, 193, 197, 209, 211, 224, 233, 239

QED quantum electrodynamics. 17, 19–21, 30, 31

RNN recurrent neural network. 191, 213

ROC Receiver Operating Characteristic. 227, 232, 239, 246

RS Randall-Sundrum. 182–184, 186, 194, 198

SCT SemiConductor Tracker. 79, 80, 116

SF scale factor. 263, 272

SK Soft-Killer. 165–168, 204, 288, 289, 309

SLR Super Logic Region. 284, 287, 314, 315

SR signal region. 187, 188, 191, 216

SSB spontaneous symmetry breaking. 21, 22, 25, 26, 33–36, 38, 40, 54, 56

TCC Track-CaloCluster. 154, 155, 204

TDAQ Trigger & Data acquisition. 278, 293

TOB Trigger Object. 108, 114, 281, 283, 301, 353

TRT Transition Radiation Tracker. 79–81

UE underlying event. 128, 138, 142, 164

UFO Unified Flow Object. 154, 155

VBF vector-boson fusion. 191, 193–195, 197, 213

vev vacuum expectation value. 34, 43, 58, 59

VR validation region. 188, 204, 205, 210, 214

WP working point. 115, 119, 120