

ALMA MATER STUDIORUM · UNIVERSITY OF BOLOGNA

School of Science
Department of Physics and Astronomy
Master Degree in Physics

**A Transformer and Novel Triggers
for the Search for Higgs Boson Pair
Production in the $b\bar{b}\tau^+\tau^-$ Final State
with the ATLAS Detector**

Supervisor:
Prof. Maximiliano Sioli

Submitted by:
Giovanni Rupnik Boero

Co-supervisors:
Dr. Brian Moser
Prof. Valerio Dao

Academic Year 2022/2023

Abstract

The Higgs boson self-interaction is a direct probe of the shape of the Higgs potential, one of the last predictions of the Standard Model awaiting experimental confirmation. A major goal of the ATLAS Experiment for the ongoing Run 3 of the LHC is to investigate the self-interaction via searches for Higgs boson pair production (HH). The $HH \rightarrow b\bar{b}\tau^+\tau^-$ analysis targets one of the most sensitive channels for this search but necessitates refining its strategy for Run 3. Two such refinements are presented in this thesis. In the first study, novel trigger chains are examined to improve the signal efficiency in data-taking, comparing them to baseline triggers used for the Run 2 analysis. The research considers new di- τ triggers with lower transverse momentum thresholds, the jet-only Delayed stream trigger from the $HH \rightarrow b\bar{b}b\bar{b}$ analysis, and proposed $b + \tau$ triggers. Together, the total trigger efficiency can be improved from 55% in Run 2 to up to 85% in Run 3. In the second study, a state-of-the-art machine learning algorithm, a transformer, is employed to separate ggF and VBF HH events. The study achieved a significant enhancement compared to the Run 2 analysis, based on a boosted decision tree. A 15% increase in VBF efficiency for the same ggF contamination, and a 100% increase in ggF rejection (inverted false positive rate) for the same VBF efficiency are found.

Table of Contents

Introduction	3
1 Theoretical Overview	6
1.1 The Standard Model of Particle Physics	6
1.1.1 Fermions and Bosons	6
1.1.2 The Standard Model as a Gauge Field Theory	8
1.2 Hadron Collider Physics	11
1.2.1 Parameters of Particle Accelerators	11
1.2.2 Hadron Colliders	12
1.3 Higgs Boson Physics	13
1.3.1 Single Higgs Boson Production and Decay Modes	14
1.3.2 Higgs Boson Pair Production and Decay Modes	16
2 The Large Hadron Collider and the ATLAS Detector	18
2.1 The Large Hadron Collider	18
2.2 The ATLAS Detector	19
2.2.1 ATLAS Coordinates	21
2.2.2 Inner Detector	21
2.2.3 Calorimeters	23
2.2.4 Muon Spectrometer	24
2.2.5 Trigger and Data Acquisition System	25
3 ATLAS Object Reconstruction	28
3.1 Object Reconstruction Overview	28
3.2 Jet Reconstruction	30
3.2.1 Flavour Tagging	31
3.3 τ -Lepton Reconstruction	32
3.3.1 Missing Mass Calculator	33
4 The $HH \rightarrow bb\tau\tau$ Analysis	34
4.1 The Run 2 Analysis Event Selection	34
4.1.1 Preselection	35
4.1.2 Event Categorisation	36
4.1.3 Signal/Background Classification	36
5 Novel Triggers for $HH \rightarrow bb\tau\tau$	38
5.1 Motivation	38
5.2 Methodology	39
5.2.1 The Offline Selections	40

5.2.2	The Trigger Chains	43
5.3	Results	44
5.3.1	The Event Kinematics	44
5.3.2	Low p_T τ Triggers	46
5.3.3	$HH \rightarrow 4b$ Delayed Stream Trigger	48
5.3.4	Proposed $b + \tau$ Triggers	50
6	A Transformer for $HH \rightarrow bb\tau\tau$ Event Classification	53
6.1	Motivation	53
6.2	Neural Networks Overview	54
6.2.1	Training	56
6.2.2	Architectures	56
6.2.3	GN2	59
6.3	Methodology	60
6.3.1	Data Preprocessing	60
6.3.2	Model Design and Training	63
6.4	Results	65
6.4.1	Comparison to the BDT	66
	Conclusion and Outlook	72
	Bibliography	74
	(A) Additional Studies on Novel Triggers for $HH \rightarrow bb\tau\tau$	80
	Acknowledgements	85

Introduction

The Standard Model of particle physics (SM) is the scientific theory that most accurately defines the building blocks of the universe. It classifies all known elementary particles and precisely describes three of the four known fundamental interactions: the *electromagnetic*, the *weak* and the *strong* interaction. Theoretically developed in the second half of the 20th Century and finalised in the 1970s, the SM has withstood rigorous tests and made predictions that have been experimentally confirmed even after several decades. One of the most noteworthy confirmation of what the SM anticipated occurred in 2012, when the ATLAS [1] and CMS [2] Experiments discovered the Higgs boson (H), predicted almost 50 years earlier [3, 4, 5].

However, despite its successes, it is known that the SM cannot be the final answer. It leaves deep questions unaddressed, such as the experimentally observed neutrino masses, dark matter and dark energy. Above all, the theory does not incorporate the forth fundamental interaction: *gravity*.

In order to find hints towards a new theory, the SM can be tested by experiments through searches for deviations from its predictions. A good candidate for this search is the Higgs boson, since it stands at the heart of the SM: 15 out of the 19 free parameters of the theory are connected to it. In particular, the Higgs boson potential (see Figure 1) is one of the major SM predictions that is not yet experimentally confirmed: its shape remains experimentally undetermined. Probing the potential is relevant to tackle deep unanswered questions in physics: it determines the stability of the vacuum state of the universe, and it could explain the observed matter-antimatter asymmetry if discrepancies with the SM were found.

The current best experimental probe to test the potential shape is the Higgs boson pair production (HH), since this process is sensitive to the Higgs boson self-coupling λ , one of the free parameters of the potential. Deviations from the shape predicted by the SM would be reflected in variations of λ , which in turn would be manifest in both the HH production rate and the invariant mass spectrum of the HH system (m_{HH}).

The ATLAS detector at the Large Hadron Collider (LHC) is one of two detectors worldwide that is able to study the rare HH process. Measuring proton-proton (pp) collisions at a high centre-of-mass energy, it looks for $pp \rightarrow HH$ production to compare the experimental rate and the m_{HH} distribution with the SM predictions. The third operational run of the LHC, termed Run 3, started in 2022 and is expected to (at least) double the available collision data set produced in the previous run (Run 2), amounting to 140 fb^{-1} . Moreover, Run 3 offers a record centre-of-mass energy

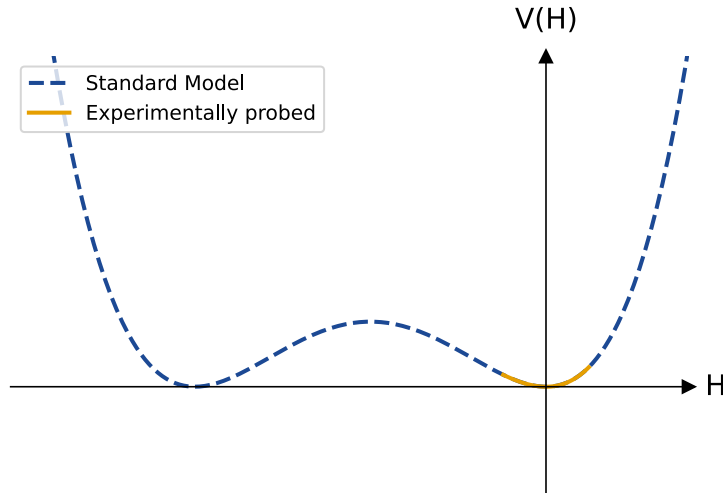


Figure 1: The Standard Model predicts a sombrero-shaped Higgs potential $V(H)$, fundamental to the theory. Experiments up to now have only probed the parabolic part around the minimum via single- H measurements. Refining the HH analyses would allow to explore the whole shape more precisely, potentially finding deviations from the SM expectations [6].

of 13.6 TeV. These unprecedented conditions present ATLAS with a unique opportunity to significantly improve the sensitivity of HH searches. However, due to the low HH production rate, analyses are still limited by the available data statistics. To address this constraint, the analysis strategy requires further refinement, with a particular emphasis on improving the efficiency to the HH signal.

One of the most sensitive channels to test the self-coupling is the $HH \rightarrow b\bar{b}\tau\tau$ channel, in which one of the Higgs bosons decays into two b -quarks¹ and the other decays into two τ -leptons. It provides a compromise between expected signal event yield and background contamination. In particular, the final state in which the τ -leptons decay hadronically ($b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$), meaning that all objects result in *jets*, has set one of the most stringent limits on the HH signal strength during Run 2 [7].

The work performed in the course of this thesis specifically studies enhancements to two crucial aspects of the $HH \rightarrow b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ signal detection, in preparation for the upcoming Run 3 analysis.

The first project deals with improving the signal efficiency of the analysis triggers. In ATLAS, the trigger system is responsible for the selection of physical processes during data-taking. When the decay products of a collision are measured, the system takes a decision whether to record the event or disregard it, due to limitations in computing resources. A trigger chain is a set of requirements that the detected collision must satisfy in order to be recorded. The all-hadronic $b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ final state is one of the most challenging to select, and requires highly performing triggers. In this thesis, novel trigger chains for the Run 3 $HH \rightarrow b\bar{b}\tau\tau$ analysis are exam-

¹Particles and antiparticles will be referred to in the same way throughout the text.

ined. The Run 2 triggers, taken as benchmark, imposed requirements only on the τ -leptons. Novel designs include *(i)* τ -leptons triggers with lower transverse momentum thresholds, *(ii)* a trigger consisting only of requirements on jets and *(iii)* a proposed chain tagging both a τ -lepton and a b -jet. The triggers are examined under different offline selections, i.e. criteria applied independently of data-taking, to provide a comprehensive understanding of their effect. To efficiently compare different combinations of the chains under multiple scenarios, a python framework is developed in this thesis. Furthermore, the event kinematics is analysed in support of the efficiency studies.

The second project deals with improving the signal event categorisation of the analysis via state-of-the-art machine learning algorithm, a transformer-based graph neural network. The $HH \rightarrow bb\tau\tau$ signal events can be categorised based on the HH production modes (named ggF and VBF), to enhance the sensitivity for testing certain SM predictions. For this purpose, a flavour tagging transformer is adapted to handle event classification with a preprocessing toolkit developed in this thesis. In particular, the flavour tagging algorithm concerns the classification of jets from variables related to the jet itself and to each internal constituent of the jet. To perform event classification, the information related to each event is re-structured to fit into an analogous input format. The model is then optimised and trained to perfect the classification, and the results are compared to the Run 2 $HH \rightarrow bb\tau\tau$ classifier, a boosted decision tree.

Outline

The structure of the thesis is the following. Chapter 1 sets the theoretical foundation of the thesis, introducing the Standard Model, the physics at hadron colliders and the Higgs boson proprieties. Chapter 2 presents an overview of both the Large Hadron Collider and the ATLAS experiment. Chapter 3 introduces the algorithms adopted by ATLAS to reconstruct physical objects from the raw measurements of the detector, with a particular focus on jets and τ -leptons. Chapter 4 summarises the $HH \rightarrow bb\tau\tau$ analysis. Chapter 5 presents the first of the two researches of this thesis, studying novel triggers for $HH \rightarrow bb\tau\tau$. Chapter 6 illustrates the second project, dealing with the transformer for $HH \rightarrow bb\tau\tau$ event classification. For both researches, each chapter provides the motivation, the methodology and the evaluation of results.

Chapter 1

Theoretical Overview

The fundamental physical theory that is probed at particle accelerators is called the *Standard Model of particle physics* (SM). An overview of this theory is presented in Section 1.1, highlighting its aspects most relevant to the analysis; Section 1.2 is dedicated to physics at hadron colliders; finally, Section 1.3 focuses on the Higgs boson, a central element of both the Standard Model and the analysis of this thesis.

The Einstein summation convention, as well as Lorentz-Heaviside units (setting $c = \hbar = 1$) are utilised throughout the text. General references for this chapter can be found in Ref. [8, 9].

1.1 The Standard Model of Particle Physics

1.1.1 Fermions and Bosons

The Standard Model classifies all known elementary particles as either *fermions*, with half-integer spin, or *bosons*, with integer spin: the bosons are in turn grouped between the *vector* bosons (with spin 1) and a *scalar* boson (spin 0).

Fermions

The elementary fermions (all of spin 1/2) are divided into *quarks*, which feel the strong interaction, and *leptons*, which do not. All fermions are grouped in three *generations*, each including two quarks of opposite charge sign and two leptons: one with non-zero electric charge, and one with neutral charge (the *neutrinos*). The three generations are identical in all their attributes except their mass, which increases from generation to generation. All stable matter is made of particles of the first generation: the subsequent generations are unstable and eventually decay to the lowest ones, but they can be produced in particle collisions.

An overview of the SM fermions is shown in Table 1.1, together with their electric charge and their experimentally measured masses.

Vector Bosons

In the SM, the vector bosons (spin 1) are the mediators of the fundamental interactions, i.e. are *gauge* bosons. These bosons only couple to particles that carry an

Gen.	Leptons			Quarks		
	Flavour	Electric charge [e]	Mass [GeV]	Flavour	Electric charge [e]	Mass [GeV]
1 st	Electron (e)	-1	$511 \cdot 10^{-6}$	Up (u)	2/3	$2.16 \cdot 10^{-3}$
	e -neutrino (ν_e)	0	0 in SM	Down (d)	-1/3	$4.67 \cdot 10^{-3}$
2 nd	Muon (μ)	-1	$105 \cdot 10^{-3}$	Charm (c)	2/3	1.27
	μ -neutrino (ν_μ)	0	0 in SM	Strange (s)	-1/3	$93.4 \cdot 10^{-3}$
3 rd	Tau (τ)	-1	1.777	Top (t)	2/3	173.1
	τ -neutrino (ν_τ)	0	0 in SM	Bottom (b)	-1/3	4.18

Table 1.1: Overview of the Standard Model fermions, together with the symbols used in this thesis. The electric charge is expressed in units of the elementary charge e . The masses were rounded from Ref. [10, 11]. Note that, while the SM does not predict neutrino masses, their existence has been experimentally confirmed from oscillation measurements, so an extension of the SM is required to take them into account.

associated conserved charge:

- The **photon** (symbol: γ) mediates the electromagnetic interaction between particles carrying the electric charge;
- The 3 **W^+ , W^- and Z bosons** mediate the weak interaction between particles carrying weak isospin;
- The 8 **gluons** (g) mediate the strong interaction between particles carrying color charge.

These conserved charges are what distinguish particles from *antiparticles*: for each elementary particle outlined in this section, the SM prescribes the existence of an antiparticle with identical features but opposite charges. In some cases, such as the photon and the gluon, the particle itself *is* its own antiparticle; as of now it is unclear if this is the case also for the neutrinos. Fermions and anti-fermions are denoted with the same notation throughout the text.

Boson	Electric charge [e]	Mass [GeV]	Mediated interaction
Photon (γ)	0	0	Electromagnetic
$W^\pm$ boson	± 1	80.379	Weak
Z boson	0	91.188	Weak
8 gluons (g)	0	0	Strong
Higgs boson (H)	0	125.25	-

Table 1.2: Overview of the Standard Model bosons, together with the symbols used in this thesis. The electric charge is expressed in units of the elementary charge e . The masses were rounded from Ref. [11]. The mediated interaction is also indicated when applicable (i.e. gauge bosons).

Scalar Boson

Finally, in the SM there is a single scalar boson: the Higgs boson. Its field is responsible for the generation of the mass of all the elementary particles through the Higgs mechanism (see Section 1.1.2).

An overview of all the SM bosons is shown in Table 1.2, together with their electric charge and their masses.

1.1.2 The Standard Model as a Gauge Field Theory

The Standard Model is a relativistic quantum field theory: as such, it considers fields to be the building blocks of reality.

The dynamics of a field are described by *equations of motion*, which give rise to the predictions that are tested by experiments. The equations of motion of a system of fields¹ ϕ_i can be encoded in the *Lagrangian density* $\mathcal{L}$, which is a single function of the fields and their spacetime derivatives:

$$\mathcal{L} = \mathcal{L}(\phi_i, \partial_\mu \phi_i, \dots). \quad (1.1)$$

In the Standard Model, the Lagrangian contains terms with derivatives no higher than the first order. To derive the equations of motion of a field ϕ from a given Lagrangian $\mathcal{L}$, one uses the Euler-Lagrange equations:

$$\frac{\partial \mathcal{L}}{\partial \phi} = \partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)}. \quad (1.2)$$

A key concept in field theories are *symmetries*, which are field transformations that are mathematically described as elements of an algebraic *group*. A theory is said to be symmetric (or *invariant*) under a group, if its equations of motion (or alternatively the Lagrangian) are unchanged as the field transform under any element of the group².

In the case of special relativistic theories, such as the SM, the dynamics of the fields are unchanged as the fields transform under the *Poincaré* group: i.e. under Lorentz boosts, spacial rotations and spacetime translations.

As a quantum theory, the SM fields can only be excited in a discrete manner: the aforementioned elementary particles are the quanta of excitation. They can interact with each other and are able to form bound states, giving rise to all the particles known up to now.

Gauge Symmetries

The theory of the SM follows the *gauge principle*: the Lagrangian is invariant under a *local gauge* transformation, meaning that the fields are allowed to vary independently in each spacetime point under the given group. There are two corollaries which justify the existence and coupling behaviour of the SM vector bosons outlined in Section 1.1.1: on one hand, the local requirement leads to the introduction

¹The spacetime dependence of each field (i.e. $\phi = \phi(x)$) is implicitly assumed throughout the text.

²More precisely, the Lagrangian is allowed to change as long as it does so by only adding a *total derivative*; the Euler-Lagrange equations still lead to invariant equations of motion.

of *connection* terms in the spacetime derivatives, i.e. *gauge fields*, one for each generator³ of the given group; their quanta are the gauge bosons. On the other hand, the invariance under a continuous symmetry implies the existence of a conserved charge (*Noether's theorem*): the gauge fields interact only with fields that carry the charge associated to the gauge symmetry. A field that does not carry a given conserved charge is said to transform *trivially* under the gauge group associated to the charge.

In the SM, the gauge group is:

$$\text{SU}(3)_C \otimes \text{SU}(2)_L \otimes \text{U}(1)_Y \quad (1.3)$$

- The non-abelian $\text{SU}(3)_C$ group describes the strong interaction, arising between fields possessing the conserved quantity C (color charge); the mediators are eight gluons because $\text{SU}(3)$ has eight generators. The theory describing the behaviour of the fields under this group is called Quantum Chromodynamics (QCD).
- The non-abelian $\text{SU}(2)_L$ group has three generators, corresponding to three gauge bosons W^1, W^2, W^3 . The conserved charge is called *weak isospin* ($\vec{T}$). The subscript L means that the gauge fields are interacting only with left handed⁴ fields.
- The abelian $\text{U}(1)_Y$ group has one generator, corresponding to the gauge boson B . The conserved charge is the *hypercharge* (Y).

The subgroup $\text{SU}(2)_L \otimes \text{U}(1)_Y$ is called the *electroweak* group.

As mentioned in the Introduction, the Standard Model accurately describes three fundamental interactions. Since interactions are associated with gauge fields, a potential association may be considered between the experimentally measured SM interactions (strong, weak, electromagnetic) and the gauge fields of the $\text{SU}(3)_C \otimes \text{SU}(2)_L \otimes \text{U}(1)_Y$ group. While it is indeed the case for the strong interaction, it is not for the other two: the electroweak bosons W_1, W_2, W_3, B cannot have a mass term associated to them in the Lagrangian, because it is not invariant under the gauge group. This contrasts with the experimentally measured mass of the particles mediating the weak interaction: the inconsistency is solved by the *Higgs mechanism*.

Higgs Mechanism

The Lagrangian describing a physical system can be invariant under a symmetry group even if the states describing the systems are not: this condition is called *spontaneous symmetry breaking* (SSB), and it arises when a state is expressed as a perturbation above a non-zero degenerate vacuum state. In the Standard Model, a major example occurs in the *electroweak* symmetry breaking.

For this purpose a new field Φ is introduced, consisting of a weak isospin doublet of complex fields:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} \quad (1.4)$$

³The set of generators of a group is the set of elements that can, through repeated application, generate all the other elements of the group.

⁴I.e. fields with eigenvalue -1 when acted upon by the chirality operator $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, where γ^i are the Dirac matrices.

with its Lagrangian $\mathcal{L}_H$:

$$\mathcal{L}_H = \underbrace{(D_\mu \Phi)^\dagger (D_\mu \Phi)}_{\text{kinetic term}} \underbrace{-\mu^2 \Phi^\dagger \Phi - \lambda(\Phi^\dagger \Phi)^2}_{\text{potential term } (V(\Phi))}. \quad (1.5)$$

In the kinetic term, the covariant derivative D_μ must take into account the local invariance under the gauge group, which in this case is the electroweak group. This implies the presence of connection terms, one for each symmetry generator: the gauge fields W^1, W^2, W^3, B .

The parameters λ and μ^2 , when the former has positive values and the latter has negative values, describe the so-called *sombrero* potential $V(\Phi)$ (Figure 1.1). The minimum is not located at 0, but at an infinite number of points given by $\Phi^\dagger \Phi = -\frac{\mu^2}{2\lambda}$. When the field transitions to its ground state, it must select one of these minima, thereby breaking the symmetry even if $V(\Phi)$ and $\mathcal{L}_H$ are still symmetric. The field's value v at the chosen minimum is termed the *vacuum expectation value*.

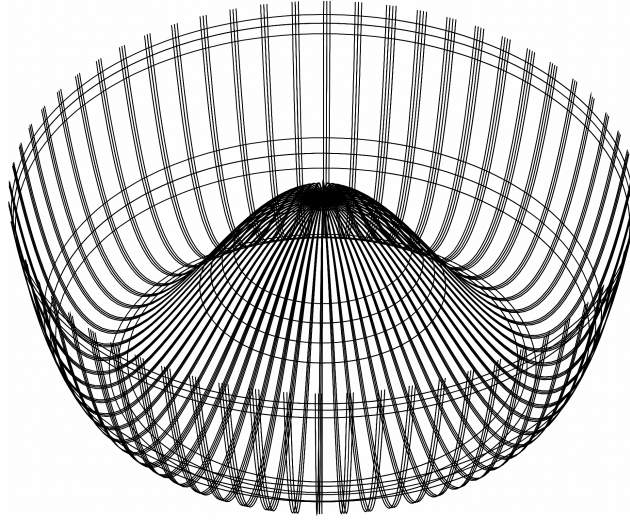


Figure 1.1: The sombrero potential $V(\Phi)$, visualised in the case of a field Φ with 2 degrees of freedom [12]. The set of minima is clearly visible as the ring at the bottom.

Selecting an arbitrary minimum ($v = \sqrt{-\mu^2/\lambda}$) and re-parametrising the Lagrangian⁵, it is possible to re-write the field expanding from v as the perturbation H :

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \quad (1.6)$$

From here, a re-parametrisation of the electroweak gauge boson fields (in the kinetic term) can be done, as such:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \quad (1.7)$$

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp i W_\mu^2) \quad (1.8)$$

⁵Including *gauge fixing*, here choosing the *unitary* gauge.

with θ_W named the weak mixing angle, and A_μ being the electromagnetic field (whose quanta are the photons γ). In this way, the weak and electromagnetic vector bosons are recovered. Now $\mathcal{L}_H$ has mass terms, i.e. coefficients of $W_\mu^- W^{+\mu}$ and of $Z_\mu Z^\mu$, given by:

$$m_{W^\pm} = \frac{gv}{2} \quad m_Z = \frac{v\sqrt{g^2 + g'^2}}{2}, \quad (1.9)$$

with g being the coupling constant of the $SU(2)_L$ gauge fields, and g' being the one of the $U(1)_Y$ field.

Out of the initial four degrees of freedom of the original doublet field Φ , three have been re-distributed to the longitudinal components of the massive $W_\mu^\pm, Z_\mu$ bosons. On the contrary, A_μ remains massless and allows $\mathcal{L}_H$ to be locally invariant under a new $U(1)_{EM}$ group; the conserved charge is the well-known electromagnetic charge Q , given by:

$$Q = T_3 + \frac{Y}{2}. \quad (1.10)$$

This formula (Gellman-Nishijima) relates Q to the quantum numbers of the original electroweak group, with T_3 being the projection of the weak isospin $\vec{T}$ along an arbitrary axis. Considering the fact that the color symmetry was untouched by the SSB, the Standard Model symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ has been broken to $SU(3)_C \otimes U(1)_{EM}$ (electroweak symmetry breaking).

The remaining degree of freedom of the original field Φ is the scalar field H from the expansion around v (Equation 1.6). This is the physical Higgs boson, whose mass is given by:

$$m_H = \sqrt{-2\mu^2} = \sqrt{2\lambda\nu^2}. \quad (1.11)$$

It now is possible to express the potential as a function of the field H :

$$V(H) = \frac{1}{2}m_H^2 H^2 + \lambda\nu H^3 + \frac{1}{4}\lambda H^4. \quad (1.12)$$

It is worth mentioning that the weak gauge bosons are not the sole particles to acquire mass thanks to the Higgs fields. In fact, all the fermions are also massless in the SM Lagrangian, and their masses arise via their coupling to H : the stronger the coupling, the larger their masses.

1.2 Hadron Collider Physics

1.2.1 Parameters of Particle Accelerators

One of the ways in which the Standard Model can make predictions that are testable at particle accelerators is through perturbation theory. From the Lagrangian it is possible to:

- Infer predictive quantities via a set of rules⁶ under asymptotic conditions that are approximately valid in particle accelerators.
- Read the interactions between the fields (i.e., between the particles).

⁶Among which: the Feynman's rules and the Fermi Golden rule.

Among the quantities predicted by the Lagrangian, a major parameter of interest is the *cross-section* σ , which states the probability for a physical process to occur. This quantity is expressed in units of barns (b), with $1 \text{ b} = 10^{-28} \text{ m}^2$.

It is possible to measure the cross-sections experimentally in order to compare them to the ones theoretically calculated assuming the Standard Model: this allows to make deductions about the validity of the theory.

In particle accelerators, the cross-section of a given process can be extracted from the rate $R = dN/dt$ (the number N of occurrences of a given process per unit time) and from the instantaneous luminosity L , a propriety of the machine:

$$\sigma = \frac{R}{L} = \frac{N}{L_{int}}, \quad (1.13)$$

where L_{int} is the luminosity integrated over time, conventionally expressed in units of inverse femtobarns (fb^{-1}).

The instantaneous luminosity has units of $\text{cm}^{-2} \text{ s}^{-1}$ and is entirely dependent on the accelerator. For instance, the LHC (which is introduced in Chapter 2) is a circular accelerator that collides two beams of particles grouped in bunches. These bunches have approximately a Gaussian shape in the two transversal x,y directions, so the formula of L is given by:

$$L = \frac{N_b f_r N_1 N_2}{2\pi \sqrt{\sigma_{1x}^2 + \sigma_{2x}^2} \sqrt{\sigma_{1y}^2 + \sigma_{2y}^2}} \cdot F, \quad (1.14)$$

with N_b being the total number of bunches, f_r the revolution frequency, N_1 and N_2 the number of particles in each bunch, and $\sigma_{ix,y}$ the bunch widths in the x,y directions [13]. This formula is corrected by one or more reduction factor(s) F that take into account the crossing angle (at the collision point) and the transverse displacement of the beams. The luminosity and the centre-of-mass energy at the interaction point ($\sqrt{s}$) are the two main parameters used to characterise a particle accelerator.

In general, when the bunches are passed through each other, particles can interact and scatter: a generic scattering process is called an *event*; if in the same bunch-crossing other interactions of lower energy are occurring, they are referred to as *pile-up*. If the event results in a specific process that the experiment is interested in, it is called a *signal* event, while the other events are called *background*.

An example of cross-section is the *production* cross-section, related to the probability of creating a particle from the interaction or decays of other particles. Another parameter of interest in particle physics is the *branching ratio* (BR): the fraction of how often a particle decays to one or more given particles with respect to all its possible decays.

1.2.2 Hadron Colliders

The elementary particles can form bound states, i.e. composite particle. A bound state of quarks is called *hadron*, with protons (p) being an example (u, u, d -quarks). A hadron is called *meson* (*baryon*) if it is composed of 2 (3) quarks. The quarks are bound together by the strong interaction (i.e. by gluons) and as such they are subject to the laws of QCD, which result in the following: when hadrons collide, the

highly energetic quarks and gluons undergo *parton showering*, emitting additional quarks and gluons in a branching pattern. As these particles separate from each other they *hadronise*, meaning that new color-neutral hadrons are formed from the vacuum; *jets* (collimated sprays of particles) are formed and emerge as the final-state product of the interaction. An illustration of the process is shown in Figure 1.2.

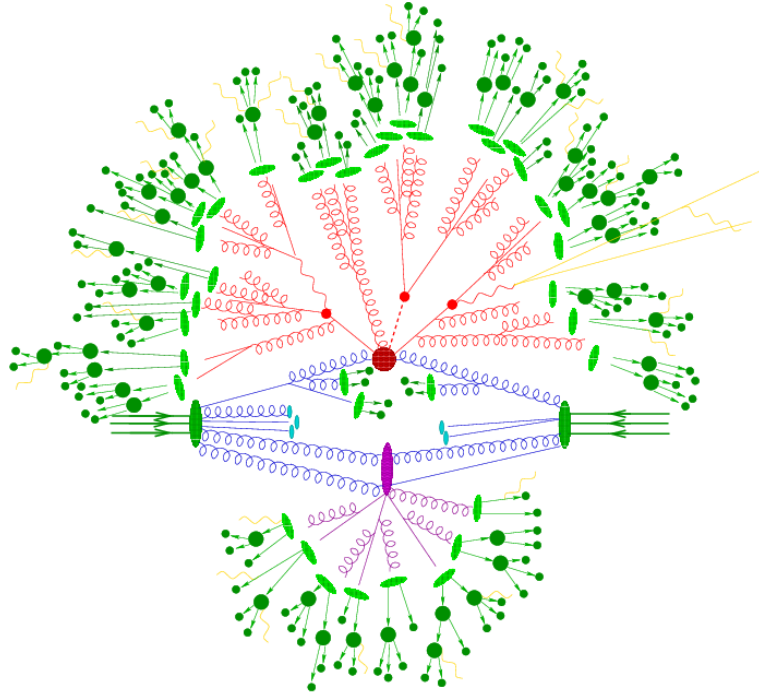


Figure 1.2: Illustration of a pp collision [14].

Due to the interaction terms in the SM Lagrangian, in theory all kinds of SM particles can (directly or indirectly) be produced as a consequence of a pp collision. Some of these particles have a low lifetime, meaning that they decay before they can be directly observed, while others might have a very low production cross-section, meaning that it could be necessary to perform a large number of collisions before being able to detect them.

While lepton colliders might be easier to deal with for precision measurements, hadron colliders have different advantages:

- All charged particles going at relativistic velocities in a curved trajectory are subject to energy loss by *synchrotron radiation*, which is highly dependent on the inverse of the mass: more massive particles, such as protons, are therefore more efficient to accelerate than leptons;
- A larger energy range can be probed, due to the momentum distribution of the quarks and gluons inside the protons.

1.3 Higgs Boson Physics

The Higgs boson is the quantum of the Higgs field. As discussed in Section 1.1.2, its mass m_H depends on μ , which is a free parameter of the SM. Once the mass is

experimentally known it is possible to calculate from the SM Lagrangian both the production cross-sections and the branching ratios, and compare the SM predictions to the experimental results. Currently m_H is measured to be 125.10 ± 0.14 GeV [11].

One of the major theoretical prediction of the Standard Model is the Higgs potential, whose particular shape is crucial for the electroweak symmetry breaking to take place (see Section 1.1.2). The potential depends on the parameter λ (Equation 1.12) known as the Higgs boson *self-coupling*, so establishing its existence experimentally represents an important test of the Standard Model. The way in which this validation can be performed at particle colliders is via the measurements of Higgs boson pair (HH , *di-Higgs*) production: in fact, the self-coupling is linked to the strength of the trilinear and quadrilinear interaction between Higgs bosons. The trilinear coupling can be manifest in H decaying into HH .

A possible observed deviation of λ from the theoretical prediction is quantified by the *coupling modifier* κ_λ , defined as:

$$\kappa_\lambda = \frac{\lambda_{HHH}}{\lambda_{SM}}, \quad (1.15)$$

where λ_{HHH} is λ as measured experimentally and λ_{SM} is λ according to the Standard Model. Deviations in λ are tied to deviations in di-Higgs production rates: for example, for $\kappa_\lambda = 0$, the observed HH production rates doubles with respect to the SM prediction, while for $\kappa_\lambda = -1$ it quadruples [15]. Therefore the investigation of HH production rates at a hadron collider provides a mean to explore the Higgs potential, offering the opportunity to test a central SM prediction.

1.3.1 Single Higgs Boson Production and Decay Modes

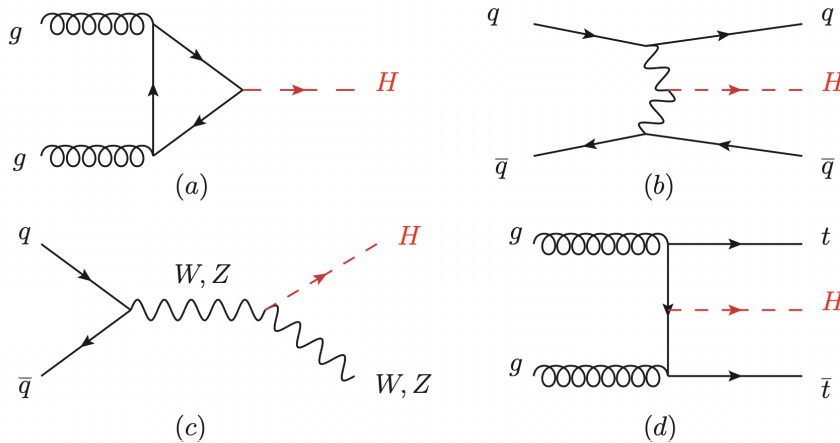


Figure 1.3: The main production modes of Higgs bosons at hadron colliders [16].

Production Modes

The most common production modes of the Higgs boson according to the SM are visualised in Figure 1.3 through their main Feynman diagrams:

- Diagram (a) displays **gluon fusion** (ggF): it is the mode with the largest cross-section at high centre-of-mass energy collisions (almost 90%); in this process two gluons (from inside the proton) interact and, through a loop of top or b -quarks, produce a Higgs boson.
- Diagram (b) displays **vector boson fusion** (VBF): the sub-leading mode, it consist of two weak gauge boson (generated by the colliding quarks) interacting and producing a Higgs boson. This mode has a distinctive signature characterised by two energetic jets in the forward region and a relatively hadron-free central region; this fact represent an analysis advantage compared to ggF, as it can be used to select Higgs boson events more easily.
- Diagram (c) displays **Higgs-strahlung** (VH): in this process, a Higgs boson is produced in association with a weak boson (Z or W); the presence of the gauge boson (radiated off incoming quarks) in the final state can be used as a signature to select signal events. The process is called in this way because in its main Feynman diagram the Higgs boson is seen to radiate off the gauge boson.
- Diagram (d) displays the **associated production of H with heavy quarks** ($ttH/bbH/tH/bH$): the Higgs is produced with a couple (or single) top or bottom quarks in the final state. Two gluons (from inside the proton) create two pairs of heavy quarks and antiquarks; an (anti)quark then annihilates with the oppositely-charged quark from the other gluon, and they emit a Higgs boson. This mode has the lowest production cross-section.

Figure 1.4 (left) displays the values of the production cross-sections as a function of the centre-of-mass energy of the collision, for a m_H value of 125 GeV.

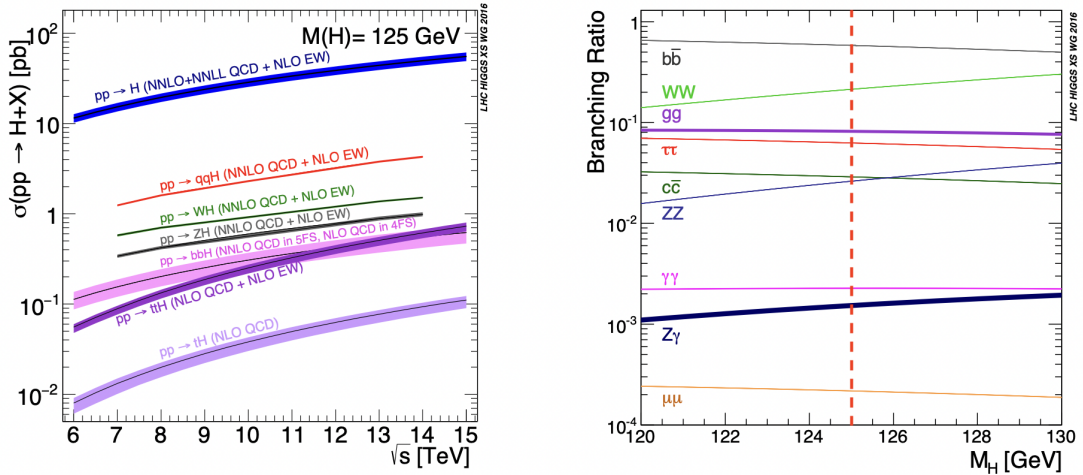


Figure 1.4: Left: the SM production modes of the Higgs boson as a function of the energy. Diagram (a) in Figure 1.3 corresponds to the uppermost curve with final state H ; diagram (b) to the second-uppermost, with final state qqH ; diagram (c) to the WH , ZH final state; diagram (d) to the bbH , ttH , tH final states. Right: branching ratio as a function of the Higgs boson mass, with the final states labelled. The red vertical line indicates the experimental value (125 GeV). Adapted from [16].

Decay Modes

As stated in Section 1.1.2, the mass of a particle depends on the strength of its interaction with the Higgs field. As the Higgs boson has stronger couplings to heavy particles, it most likely decays into the heaviest particles kinematically allowed. Since H has a mass of ~ 125 GeV, the b -quark (~ 4.18 GeV) is the most massive particle that can be pair-produced at rest: $H \rightarrow b\bar{b}$ has therefore the largest branching ratio ($\sim 58\%$). The second most prominent decay is the one to a pair of W bosons ($\sim 21\%$). The decay to a pair of τ -leptons is the fourth most likely (following the gg decay) with a branching fraction of $\sim 6.3\%$.

Figure 1.4 (right) displays the values of the SM branching ratios as a function of the Higgs boson mass in the range around 125 GeV.

1.3.2 Higgs Boson Pair Production and Decay Modes

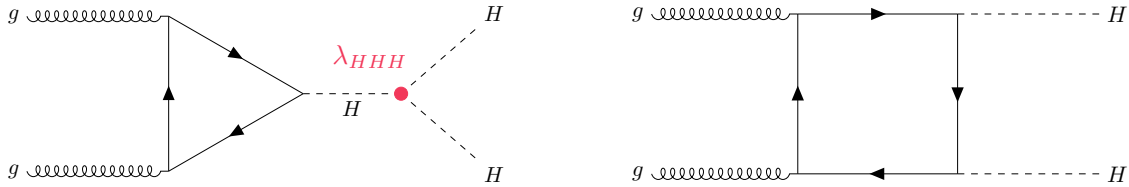


Figure 1.5: Feynman diagrams at leading order of HH production in the ggF mode: the *triangle* diagram (left) and the *box* diagram (right) [17].

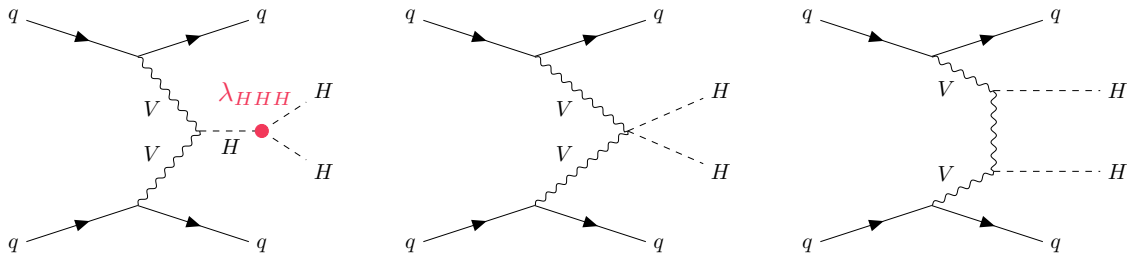


Figure 1.6: Feynman diagrams at leading order of HH production in the VBF mode [17].

Production Modes

According to the Standard Model the main HH production mode is gluon fusion (ggF). Figure 1.5 displays the leading order Feynman diagrams of the process, consisting of the so-called *triangle* and *box* diagrams. They interfere destructively, resulting in a very small production cross-section: $\sigma = 31.1^{+2.1}_{-7.2}$ fb for $m_H = 125$ GeV and $\sqrt{s} = 13$ TeV [17]; this means that, for ggF, the HH production cross-section is three orders of magnitude smaller than the single- H one [18]. The second most dominant HH production mode is vector boson fusion (VBF, Figure 1.6), with an even smaller cross-section of 1.73 ± 0.04 fb [17]. Since particle couplings are represented by vertices in Feynman diagrams, it can be seen that the VBF mode allows

to study not only λ but also the Higgs couplings to two weak gauge bosons. The observed deviation from the SM predictions of this couplings is parameterised by the modifier κ_{2V} . Differences in this coupling are manifested in variations in production rates: for $\kappa_{2V} = 2$, the VBF production mode increases by two orders of magnitude.

Decay Modes

In Figure 1.7, the branching ratios of the SM HH process are depicted. The most likely HH decay modes can be inferred from the combinatorics of the single- H decay branching ratios. The dominant decay mode involves the Higgs boson pair decaying into four b -quarks (referred to as the $4b$ channel), with a branching ratio of 34%. The second most frequent is the decay to two W bosons and two b -quarks (25%). The third largest branching ratio is the one of the decay into two b -quarks and two τ -leptons ($bb\tau\tau$ channel, 7.3%), the focus of this thesis. A significantly smaller channel, but relevant for the HH analysis, is the decay into two b -quarks and two photons ($bb\gamma\gamma$ channel, 0.26%).

	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	34%				
WW	25%	4.6%			
$\tau\tau$	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
$\gamma\gamma$	0.26%	0.10%	0.028%	0.012%	0.0005%

Figure 1.7: Branching ratios of different HH decay modes according to the SM [19].

From an experimental point of view, at hadron colliders the most sensitive channels to HH production are the $4b$, the $bb\gamma\gamma$ and the $bb\tau\tau$ channels. While the first has the largest branching ratio of all decays, it also faces challenges in colliders since a large number of non- HH events exhibits b -hadron signatures in the final state (QCD background). On the contrary, $bb\gamma\gamma$ has a very low branching ratio but can exploit the $\gamma\gamma$ signature in detectors to enhance background rejection. The $bb\tau\tau$ channel offers a balanced compromise, with a moderate branching ratio and a higher background rejection than $4b$ thanks to the presence of the 2 τ -leptons.

Chapter 2

The Large Hadron Collider and the ATLAS Detector

This chapter introduces the experimental setup: as typical of high energy physics experiments, it is composed of the particle accelerator that generates the collisions, and the detector that measures the decay products to perform the analysis. Specifically, the setup utilised in this thesis involves the Large Hadron Collider [20] and the ATLAS¹ Detector [21]. An overview of both apparatuses is presented in this chapter, respectively in Section 2.1 and Section 2.2.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a particle accelerator built by the European Organization for Nuclear Research (CERN) close to Geneva, Switzerland. It is located in a ring-shaped tunnel with a circumference of 26.7 km, approximately 100 m underground. Its purpose is to collide particles at a high center-of-mass energy and with a high luminosity, allowing new particles with larger masses to be produced and facilitating observations of rare processes.

As its name suggests, the LHC collides hadrons, either protons or heavy ions. The particles are accelerated in bunches via 8 radiofrequency cavities, guided in circle by 1232 dipole magnets, and focused via 474 quadrupole magnets. The ultra low temperature of the magnets (1.9 K), achieved through contact cooling with superfluid helium, makes them superconducting and allows them to reach a magnetic field of up to 8.3 T. The configuration of the magnets allows to accelerate particles in two beams travelling in opposite directions. The particles are accelerated even before entering the LHC, in the pre-existing accelerator complex of CERN (displayed in Figure 2.1). For protons, the acceleration chain is currently composed of the linear accelerator LINAC 4, the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS).

The operational periods in which the LHC actively collides particles and collects data are called *Runs*. They are alternated with periodical shut downs in which maintenance and upgrades of the collider (and its experiments) take place. The LHC is currently undergoing its third Run.

¹A Toroidal LHC ApparatuS.

other with an increasing distance from the interaction point, located at the centre. The subdetectors are either *trackers*, responsible to determine the momentum (and charge sign) of charged particles, or *calorimeters*, specialised on determining particle energies.

Starting from the collision point, there are:

- **Inner Detector:** a precise tracker of charged particles.
- **Electromagnetic Calorimeter:** responsible for measuring the energy of particles that mostly interact electromagnetically.
- **Hadronic Calorimeter:** responsible for measuring the energy of hadrons that mostly interact strongly.
- **Muon Spectrometer:** a tracker for muons; it is located at the outermost portion of the detector because muons are the only detectable particles able to pass through the calorimeters.

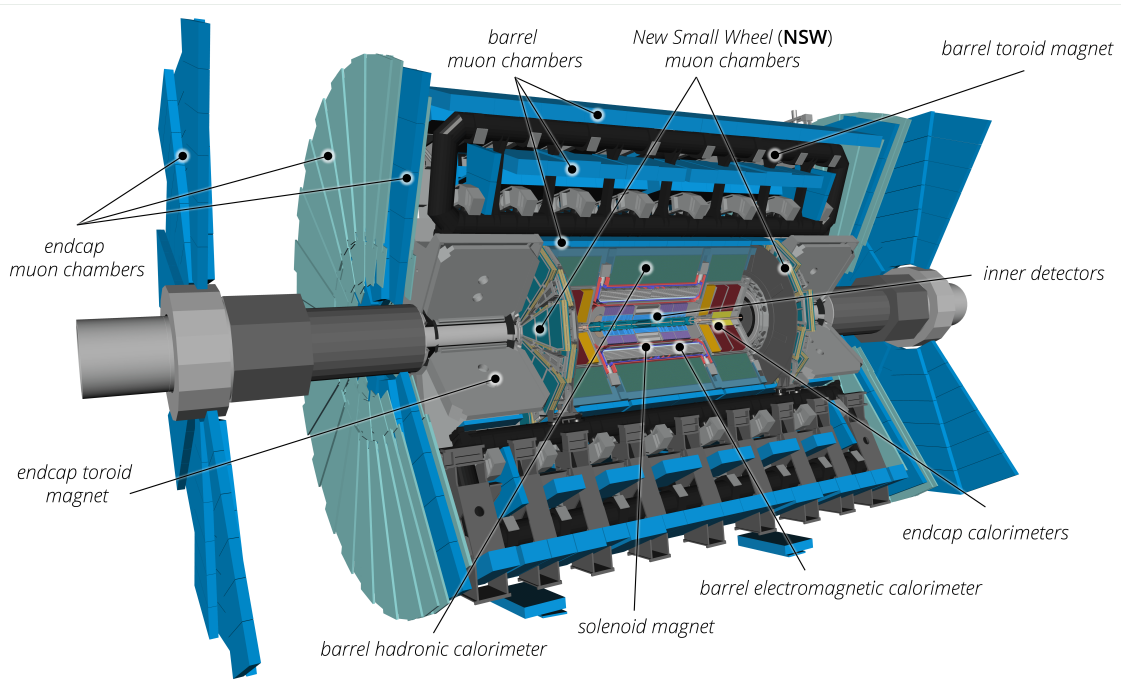


Figure 2.2: Cut-away view of the ATLAS Detector [27].

These subdetectors are divided into the barrel region and the endcap regions: in the former they are organised as concentric cylinders surrounding the beam axis, while in the endcaps they are positioned on discs perpendicular to it. To study the kinematics of the particles, the detector's structure also includes large superconducting magnets: a thin central solenoid surrounds the inner detector, while a set of large toroidal magnets (with 8 coils each) are incorporated by the Muon Spectrometer. Finally, ATLAS also includes a sophisticated **Trigger and Data Acquisition System** which rapidly identifies and records interesting events.

The whole detector is 44 m long and 25 m high, and weights approximately 7000 tons.

This section refers to Ref. [27] for the upgrades of the ATLAS detector that have

been performed in the time between its original construction and the configuration used in this thesis. The general source for the Trigger System is Ref. [28].

2.2.1 ATLAS Coordinates

The ATLAS experiment takes the interaction point as the origin of its right-handed coordinate system. The x-axis is pointing towards the centre of the ring, the z-axis is oriented along the direction of the particle beam and the y-axis is pointing upwards. The xy-plane is referred to as the *transverse* plane, and the variables defined in it are labelled by the subscript _T (e.g. p_T for the transverse momentum).

Angular variables

The polar angle θ describes the angle from the z-axis, as measured in the laboratory frame. While differences in azimuthal angle are invariant under Lorentz boosts along the beam axis, because they refer to the plane orthogonal to it, differences in θ are not. It is therefore common to describe the angle relative to the beam axis via the pseudorapidity η , which is defined as:

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right). \quad (2.1)$$

In the high momentum regime of the particle collision², differences in η are invariant under Lorentz boosts along the beam axis.

This allows measurements to obtain independence from the frame of reference of the laboratory, an advantage that the polar angle θ does not grant. From the definition of pseudorapidity, the y-axis is $\eta = 0$ and the z-axis is $\eta = \infty$; being logarithmic, at $\theta = \pi/2$ radians η is equal to just 0.88.

The azimuthal angle ϕ measures the angle in the transverse plane from the x-axis. The detector's design is symmetrical around ϕ , with almost full coverage in this angle.

When using polar coordinates, the radial distance from the beam pipe is denoted by r . Finally, from the pseudorapidity η and the azimuthal angle ϕ , it is possible to define a general geometrical distance ΔR between any two objects, as:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \quad (2.2)$$

2.2.2 Inner Detector

The Inner Detector (ID) is situated within the 2 T axial magnetic field generated by the central solenoid. This field bends the trajectories of charged particles: the task of the ID is to measure discrete energy deposits (*hits*) so that the trajectories can be reconstructed (as *tracks*); furthermore, the magnetic field enables the calculation of particle momenta based on the magnitude of their curvatures; their charge sign, on the other hand, can be inferred from the direction of their curvature. The charged tracks must have $p_T > 0.5 \text{ GeV}$ in order to be reconstructed using hit measurements from the ID. The designed momentum resolution of the reconstructed tracks is $\sigma_{p_T}/p_T = 0.05\%p_T [\text{GeV}] \oplus 1\%$.

²Where the mass of the particles can be neglected with respect to their kinetic energy.

The ID is composed of three subdetectors: starting from the centre, they are the **Pixel Detector**, the **Semiconductor Tracker** and the **Transition Radiation Tracker**. While the first two cover the region within pseudorapidity $|\eta| < 2.5$, the latter covers the region within $|\eta| < 2.0$.

A cut-away view of the ID is shown in Figure 2.3.

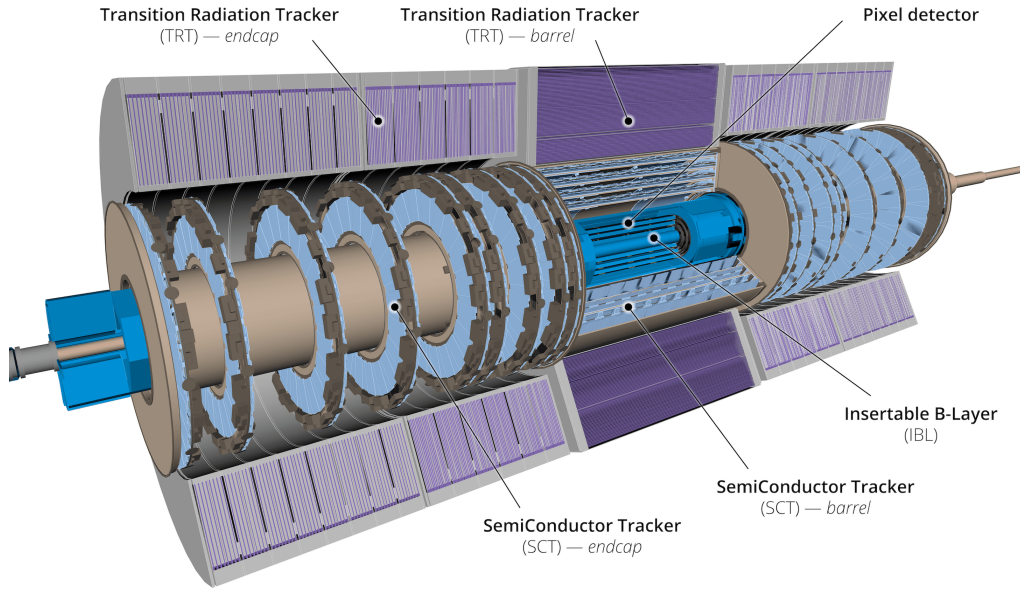


Figure 2.3: The ATLAS Inner Detector and its subcomponents [27].

Pixel Detector

The Pixel Detector is the closest to the interaction point and has the highest granularity. Entirely made of silicon, it consists of 3 barrel layers and 3 discs on each endcap, plus an additional inner layer called **Insertable *b*-layer** (IBL) [29, 30]. The IBL was installed before Run 2 with the specific purpose of enhancing the reconstruction of secondary vertices, such as those arising from *b*-hadron decays. The pixels' accuracy in the barrel layers is $10\ \mu\text{m}$ ($r\phi$) and $115\ \mu\text{m}$ (z), while for the discs it is $10\ \mu\text{m}$ ($r\phi$) and $115\ \mu\text{m}$ (r). The Pixel detector provides on average 4 hits per track, with the first occurring in the IBL.

Semiconductor Tracker

The Semiconductor Tracker (SCT) is a silicon microstrip tracker, providing on average 8 hits per track. In the 4 layers of barrel region, the SCT employs stereo strips (at a small angle of 40 mrad) to measure both $r\phi$ and z , with a set of strips parallel to the beam pipe; the 9 discs at the endcaps have a set of stereo strips and a set of strips running radially to the beam pipe. Each module has an accuracy of $17\ \mu\text{m}$ ($r\phi$) and $580\ \mu\text{m}$ (z) in the barrel and $17\ \mu\text{m}$ ($r\phi$) and $580\ \mu\text{m}$ (r) in the endcaps.

Transition Radiation Tracker

The Transition Radiation Tracker (TRT) complements the tracking of the silicon detectors in the outermost region of the ID. It consists of 300,000 gas-filled straw

tubes (4 mm in diameter) with a gold-plated tungsten wire in the middle. A potential difference is maintained between the wire and the walls of the tube. As the charged particles cross the mixture, they ionise the gas, freeing other charges that drift under the electric field. This drifting in turn induces a current in the wire, detected as the signal. In addition to its tracking power, the TRT has also particle identification capabilities thanks to the phenomenon of *transition radiation*: the emission of light from a charge passing through inhomogeneous material. In fact, in the barrel region the straw tubes are immersed in a matrix of polypropylene fibres, while in the end-caps they are interleaved with foils. The TRT can for instance distinguish electrons from pions, based on the detection of photons in the X-ray energy range.

While the resolution ($130\ \mu\text{m}$) is worse than the one of the silicon subdetectors, the TRT provides a larger number of hits per track on average (36).

2.2.3 Calorimeters

The **Electromagnetic** and the **Hadronic Calorimeters** are responsible for measuring the energies of particles that interact electromagnetically or strongly. With the exception of muons, the detectable particles resulting from a collision are forced to stop entirely within these detectors, depositing all of their energy so that it can be measured. Precisely measuring the energy of the particles is crucial for calculating the missing transverse energy, necessary to reconstruct the undetectable neutrinos (see Chapter 3).

These calorimeters are of the *sampling* type: they alternate active detection materials with dense absorbers. The interaction between the high-energy particles and the latter components results in a cascade of secondary particles known as *showers*, whose energies can be measured by the former components.

In ATLAS, the calorimeters surround the central solenoid. A schematics is displayed in Figure 2.4.

Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECal) primarily measures the energy deposits of photons and electrons.

It is divided between a barrel region ($|\eta| < 1.475$, with a characteristic accordion shape) and the endcaps ($1.375 < |\eta| < 3.2$). It adopts liquid argon (LAr) as the active detection material³ due to its radiation hardness and energy resolution [31]; lead is instead used as the absorber.

The mean distance travelled by an electron before reducing its energy by $1/e$ is called radiation length (X_0); the ECAL is $22 X_0$ deep. The designed relative energy resolution is $\sigma_E/E = 10\%/\sqrt{E[\text{GeV}]} \oplus 0.7\%$.

Hadronic Calorimeter

The Hadronic Calorimeter (HCal) primarily measures the energy deposits of hadronic showers.

It is divided between the Tile Calorimeter (TileCal), covering pseudorapidities reaching up to $|\eta| < 1.7$, and two subdetectors in the endcaps: the LAr Hadronic Endcap

³Together with Kapton electrodes.

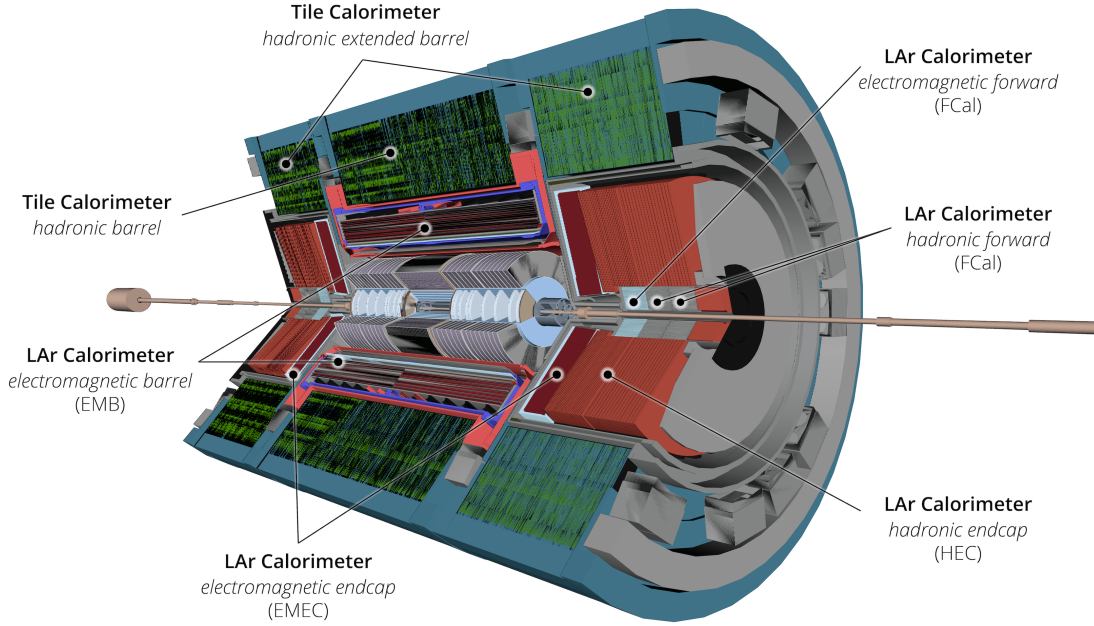


Figure 2.4: The ATLAS Calorimeters and their subcomponents [27].

(HEC, $1.5 < |\eta| < 3.2$) and the LAr Forward Calorimeter (FCal, $3.1 < |\eta| < 4.9$). The TileCal adopts steel as the absorber, while using 500,000 scintillating plastic tiles arranged in pseudo-projective towers as the active material [32]. Both the HEC and the FCal use LAr as the detection material, but while the former adopts only copper as the passive absorber, the FCal includes also modules made of tungsten. The mean distance travelled by a hadron before a nuclear interaction is called nuclear interaction length (λ_{int}). The HEC is $7.4 \lambda_{\text{int}}$ deep, to be summed to the $1.5 \lambda_{\text{int}}$ of the ECAL. This value is enough to stop any collision products other than muons (and neutrinos) in reaching the Muon Spectrometer. The designed relative energy resolution is $\sigma_E/E = 50\%/\sqrt{E[\text{GeV}]} \oplus 3\%$ for the TileCal and HEC, which worsens to $\sigma_E/E = 100\%/\sqrt{E[\text{GeV}]} \oplus 10\%$ for the FCal.

2.2.4 Muon Spectrometer

As the Inner Detector, the Muon Spectrometer (MS) is a tracker: it precisely reconstructs the trajectories of charged particles and measures their momentum via their curvature. While the ID's magnetic field is provided by the central solenoid, for the MS it is created by the toroidal magnets that are intertwined with its chambers. The MS has a fundamental particle identification capability: in fact, the muons are the only detectable particles that are not stopped by the calorimeters, and that can therefore travel up to the MS. The MS covers the pseudorapidity range $|\eta| < 2.7$, and the region within $|\eta| < 2.4$ is also used for triggering. The designed resolution for standalone tracks is $\sigma_{p_T}/p_T = 10\%$ for 1 TeV muons.

As shown in Figure 2.5, the MS employs multiple detection methods: *monitored drift tubes chambers* for precisely measuring the bending coordinate, and either *resistive plate chambers* or *thin gap chambers* (TGCs) for fast triggering and measurements of ϕ .

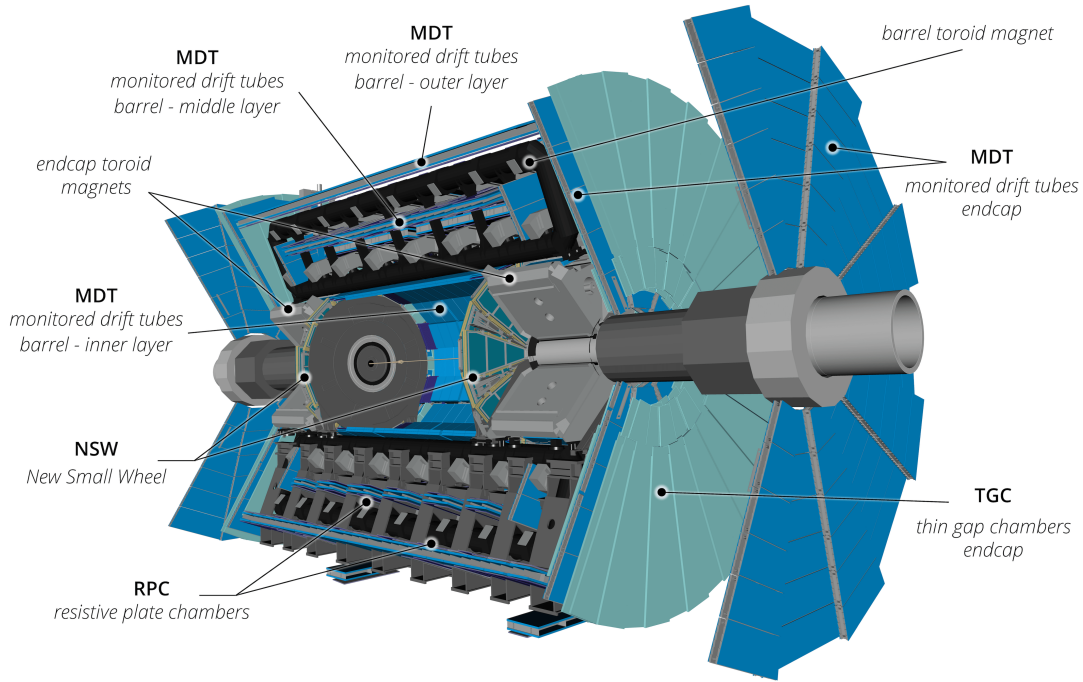


Figure 2.5: The ATLAS Muon Spectrometer and its subcomponents [27].

As for the other detectors, the MS is divided between the endcaps and the barrel region. There are three discs in each endcap, of which the innermost (known as the *Small Wheel*) has been completely upgraded during the shutdown between Run 2 and Run 3. The New Small Wheels adopt two chamber technologies: small-strip TGCs (around 357,000 channels) and micro-mesh gaseous structure (Micromegas, with up to two million channels) detectors. These upgrades are fast enough for Level-1 trigger and improve the resolution in ϕ and in p_T ; in addition, due to allowing higher background rejection, lower muon p_T thresholds are possible [33].

2.2.5 Trigger and Data Acquisition System

The ATLAS detector encounters a 40 MHz rate of bunch-crossings, each with an average of 60–70 interactions expected for Run 3. Due to computational limitations, this amount of data cannot be entirely recorded: for each bunch-crossing interaction (*event*), a decision must be made regarding whether to permanently record the data for further analysis or not. To address this, ATLAS employs the Trigger and Data Acquisition (TDAQ) system, responsible for event selection and recording. It aims to minimise the recorded event rate while maintaining high efficiency for the physical processes of interest. This system, illustrated in Figure 2.6, is divided into the **Level-1 Trigger** (L1) and the **High-Level Trigger** (HLT); supporting the two levels is the DAQ system, a second data pipeline performing readout, data processing and monitoring.

Level-1 Trigger

Each event undergoes initial processing by Level-1 (L1) through two subsystems, L1Calo and L1Muon, handling reduced-granularity information from the Calorime-

ters and the Muon Spectrometer, respectively. Subsequently, the L1 topological processor (L1Topo) applies topological selection to objects reconstructed by L1Calo and L1Muon. The Central Trigger Processor (CTP) forms a trigger decision based on inputs from these subsystems. L1 reduces the rate from 40 MHz to the maximum detector read-out rate of 100 kHz, making decisions with a latency of less than $2.5\,\mu\text{s}$.

High-Level Trigger

An event that is accepted by L1 is sent to the High-Level Trigger (HLT), which is software-based. Here a higher level of detail is available. The event objects are reconstructed and identified by *online* algorithms more similar to the ones used *offline* after data-taking (discussed in Chapter 3); for instance, the HLT has access to hits in the ID and is therefore able to reconstruct tracks. The event is reconstructed either in the full detector volume or in restricted Regions-of-Interest (RoIs) where candidate trigger objects have been identified by L1. The HLT imposes tighter selection criteria than L1, exploiting full detector information. The targeted physics output rate of the HLT for Run 3 is 3 kHz on average.

The selection criteria of the trigger system are known as *trigger chains*; they define the reconstruction and selection stages from L1 to HLT, including kinematic variables threshold, object type and multiplicity. An event is accepted if it satisfies the conditions of at least one trigger chain. The full list of chains used for data taking is called *trigger menu*. The menu also includes a *prescale* value for each chain, which is used to limit the trigger rates in order to better manage computational resources: applying a prescale value of k means that, out of k events passing the selection criteria, only one is accepted for data taking.

Accepted events are stored in distinct datasets (*streams*) to facilitate subsequent processing and analysis, with the trigger menu dictating the streams to which an event is written. Physics Streams, reserved for collision events of interest, consist of the *Main* stream, the *B-physics and light states* (BLS) stream, and the *Hadronic* stream. The Main stream events are promptly reconstructed after calibration and data quality assessment, while events in the other two streams (referred to as *Delayed*) are reconstructed when resources allow it. Additional streams can be introduced for specific data-taking configurations, such as the *Enhanced bias* stream, used for trigger rate predictions.

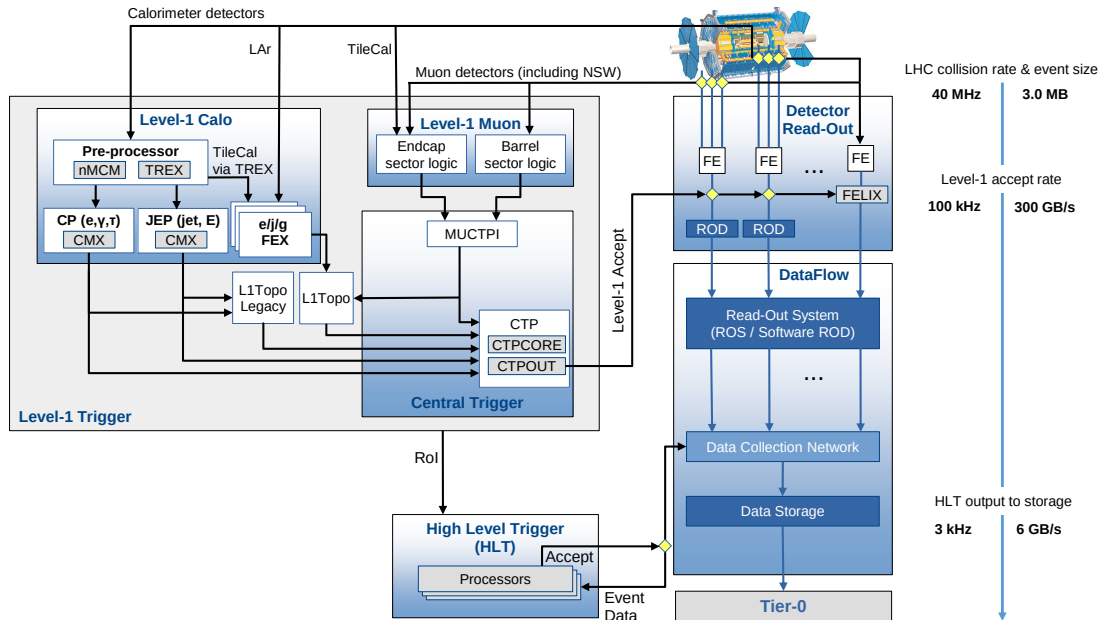


Figure 2.6: The ATLAS Trigger and Data Acquisition system [28].

Chapter 3

ATLAS Object Reconstruction

As seen in Chapter 2, the ATLAS detector directly collects only discrete measurements of the products of pp collisions, such as tracker hits and energy deposits in the calorimeters. From this raw data, meaningful information must be extracted about the physical objects (such as particles) in order to perform an analysis. This is the task of *offline object reconstruction* algorithms, presented in this chapter. In ATLAS, all the reconstruction algorithms are implemented in the ATHENA framework [34].

An overview of object reconstruction is presented in Section 3.1; a focus is then individually given to hadronic jets in Section 3.2 and to τ -leptons in Section 3.3. Jets from b -quarks and τ -leptons are the final state objects of the HH decay analysed in the thesis.

3.1 Object Reconstruction Overview

From the raw data, the starting point of object reconstruction is building *tracks*, *vertices* and *topological clusters*:

- **Tracks** are the trajectories of charged particles in the trackers. The algorithms must recognise the hits belonging to the same track, facing a complex pattern recognition task; the combinatorial Kalman Filter [35] is one of the main approaches. Tracks must have $p_T > 0.5$ GeV and be within the acceptance of the ID ($|\eta| < 2.5$).
- The **primary vertex** is the location of the most energetic pp interaction, while *pile-up* vertices are the locations of the other pp collisions during the bunch-crossing; from the collision's products, a **secondary** or even a tertiary vertex can occur (e.g. in b -quark decays); they are identified from the intersections of the tracks. The distance of closest approach from a track to the primary vertex is called **impact parameter**.
- **Topological clusters** are interconnected energy deposits in the calorimeters that exceed a noise threshold [36].

With these high-level elements it is then possible to reconstruct the final physical objects: the **charged leptons** (e , μ , τ), the **photons** and the hadronic **jets**. In

addition, while neutrinos are not directly detected by ATLAS, a part of their 4-vector can be reconstructed from the imbalance of energy-momentum conservation in the transverse plane: the **missing transverse energy** (MET).

The objects are reconstructed based on their expected signatures left in the different sub-detectors:

- **Electrons**, being charged, leave hits in the ID and shower in the ECal, where they are stopped; it is then possible to infer their momentum and charge sign from the ID tracks, and the total energy from the ECal clusters; the reconstruction is performed combining the information from the two sub-detectors, fitting the topological clusters together with the tracks [37].
- **Photons**, being chargeless, do not leave hits in the ID, but similarly to the electrons they shower in the ECal without reaching the HCal; the absence of a track serves as the primary distinction from electrons.
- **Muons** are *minimally ionizing* particles at the energies probed by ATLAS: they deposit very little energy in the calorimeters, and they are the only detectable particles able to reach the Muon Spectrometer. Being charged, they leave hits in both trackers (ID & MS), which most of the times are fitted together (*combined* muons) [38].
- **Hadrons** such as protons and neutrons shower in the calorimeters, leaving most of their energy in the HCal; while the former (being charged) leave a track in the ID, the latter do not.

The general treatment for hadronic jets and for τ -leptons is discussed in Sections 3.2 and 3.3 respectively.

The schematics of the signatures of different particles can be visualised in Figure 3.1.

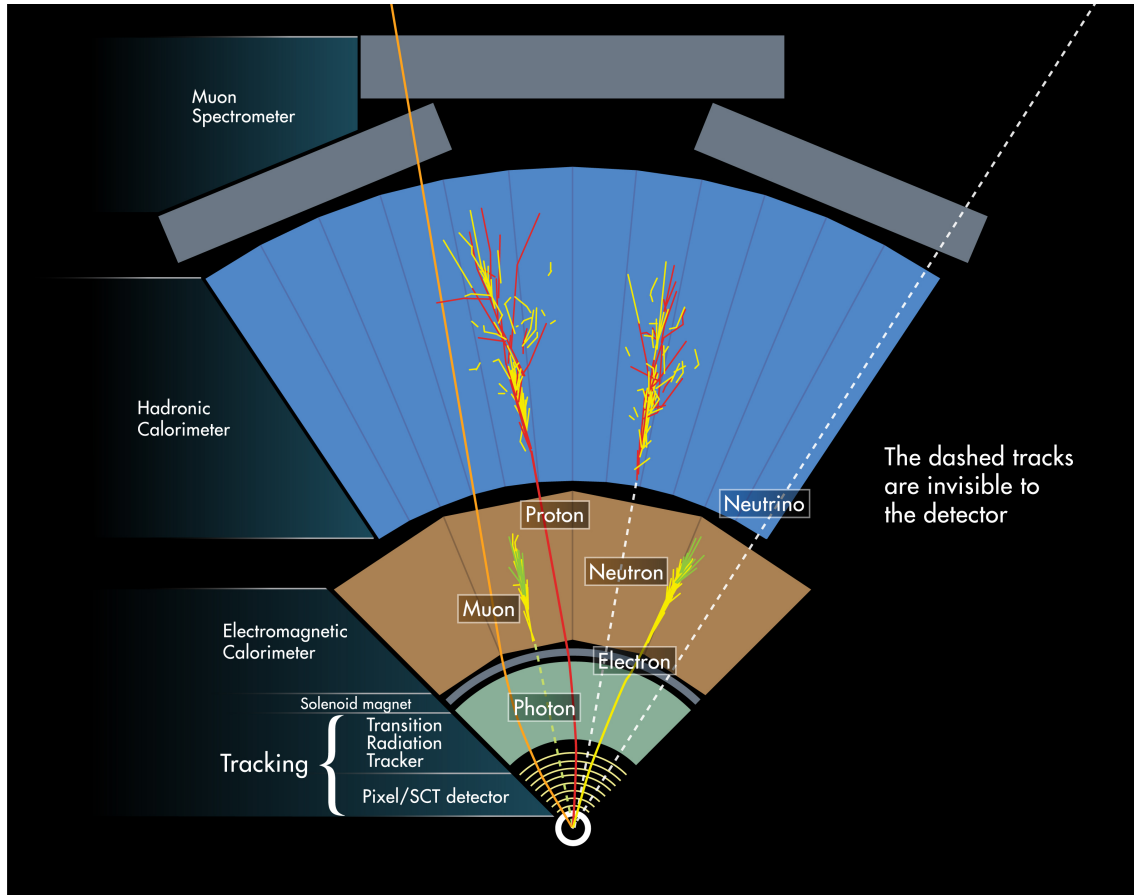


Figure 3.1: Particle signatures in the different ATLAS sub-detectors [39].

3.2 Jet Reconstruction

Jets (introduced in Section 1.2.2) result from quarks and gluons undergoing hadronisation. They are reconstructed by associating topo-clusters and tracks in an approximate cone shape that expands outward.

To perform jet reconstruction, 4-vectors are required as inputs; in ATLAS, these 4-vectors are calculated by the *Particle Flow* [40] algorithm, which combines information from the calorimeters and from the ID. The reconstruction algorithm considered in the $HH \rightarrow b\bar{b}\tau\tau$ analysis is called *anti-kt* [41]. It calculates a weighted distance d_{ij} between any pair i, j of 4-vectors associated to the constituents (tracks and topo-clusters):

$$d_{ij} = \min \left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2} \right) \cdot \frac{\Delta R_{i,j}^2}{R^2}, \quad (3.1)$$

with $\Delta R_{i,j}$ being the geometric distance defined in Section 2.2.1; the radius R is a free parameter of the algorithm, and it is set to 0.4 for the jets used in this thesis. At the same time a threshold distance d_{iB} is calculated for any object i , defined as:

$$d_{iB} = \frac{1}{p_{T,i}^2}. \quad (3.2)$$

Then the minimum between these two distances is calculated: if $d_{ij} < d_{iB}$, the algorithm merges the constituents i, j ; otherwise, the constituent i is considered a

jet, and it is removed from the pairings of following iterations of the algorithm until all constituents have been merged to a jet.

Other algorithms and procedures adopted in jet reconstruction include vertex tagging (JVT), used to discriminate those that originated from pile-up vertices [42], jet calibration and flavour tagging, discussed in the next section.

3.2.1 Flavour Tagging

Flavour tagging (FTAG) is the task of identifying the flavour of the particle that initiated a jet. It is common to refer to jets according to their underlying initiating parton: those initiated by a gluon or a u, d, s -quark are termed *light-jets*, while those initiated by c - or b -quark are respectively called *c-jets* or *b-jets*. Top quarks decay even before they can hadronise, so they do not form jets.

The most common use case of flavour tagging is b -tagging, the identification of b -jets: it is crucial not only for physics involving b -quarks directly but also for top quarks, since they decay almost always as $t \rightarrow Wb$.

When b -quarks hadronise, they mainly form mesons composed of a b - and a light-flavoured quark. B -hadrons exhibit particular signatures that are used to identify them:

- Their significant lifetime of 1.5 ps [10] allows them to travel away from the primary vertex (4-5 mm for a momentum of 50 GeV) before decaying, therefore exhibiting a second decay vertex within the ID displaced from the first;
- Most hadrons from b -quarks are heavier than hadrons from c -quarks or light quarks, having a mass of ~ 5 GeV;
- Most of the b -hadrons decay into c -hadrons, potentially resulting in tertiary vertices from the c -hadron decay.
- The tracks associated to the b -jets have large impact parameters.
- b -jets have a greater number of charged tracks.

All these proprieties are exploited by b -tagging algorithms. Figure 3.2 illustrates the main characteristics of b -jets.

The b -tagging algorithm from the Run 2 analysis is the *DL1r tagger* [43], a multi-variate classifier. It is a deep neural network with a multidimensional output, giving the probabilities for a jet to be b -, c - or light-flavoured. It combines the outputs of lower level algorithms, including ones based on secondary/tertiary vertices (JetFitter [44], SV1 [45]) and others based on impact parameters (RNNIP [46]). DL1r was improved as *DL1d*: it substitutes RNNIP with DIPS [47], an algorithm based on the Deep Sets architecture [48] which is able to handle inputs in a permutation-invariant manner.

In 2022, a new flavour tagging algorithm based on graph neural networks (*GN1*) was developed [49]. It takes as inputs a variable number of tracks within the jet to predict its flavour, without the need of other low-level algorithms; GN1 is also able to carry on auxiliary tasks, such as grouping the tracks into vertices.

An improved version of GN1, termed *GN2*, is based on the transformer architecture [50] and is currently in the final stages of development; it is the focus of Chapter 6.

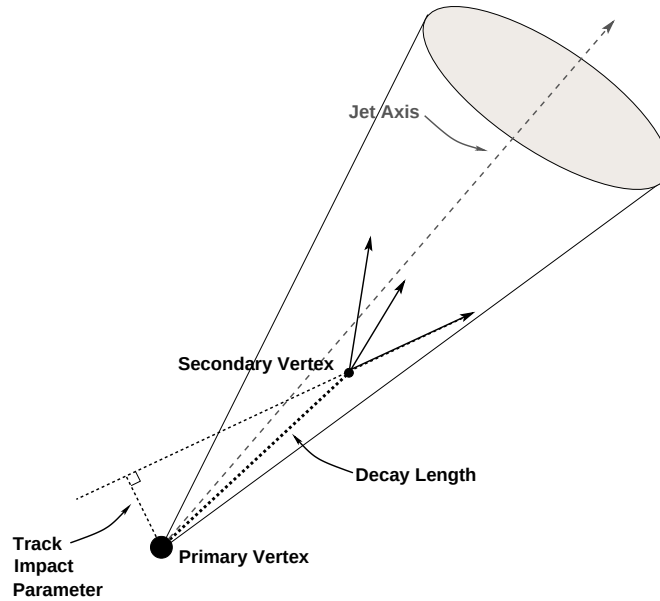


Figure 3.2: The most relevant variables for the identification of b -jets [51].

3.3 τ -Lepton Reconstruction

The τ -lepton decays into a τ -neutrino and a W boson, which in turn promptly decays fermionically: to an electron and an e -neutrino in 17.8% of the times; a muon and a μ -neutrino in 17.4% of the times; and hadronically (denoted as τ_{had}) in 64.8% of the times, the majority [10]. The τ is the only lepton capable of decaying hadronically, due to its large mass of $\sim 1.777 \text{ GeV}$. This decay occurs into either three charged hadrons (branching ratio of 14.55%) or a single charged hadron; additionally, neutral pions¹ can be present. The number of charged particles produced by the τ decay is named *prongness*, e.g. the 3-hadrons process termed *3-prong* decay. In this thesis only τ_{had} decays are considered. Refer to Figure 3.3 for an overview of the τ -lepton decays.

The heavy τ -lepton has a lifetime of $2.9 \cdot 10^{-17} \text{ s}$, very short compared to the other leptons; it therefore decays in the beam pipe, before reaching any sub-detector. In hadronic decays, the τ -lepton must first be reconstructed as a jet from its hadronic products and only subsequently can be identified.

Considering that among the τ_{had} products there are particles invisible to ATLAS, such as the τ -neutrino, the energy and momentum of the original τ -lepton cannot coincide with those directly measured from the jets; the measurable proprieties are usually referred to as belonging to $\tau_{\text{had-vis}}$ (the *visible* τ), which is what is computationally reconstructed.

The $\tau_{\text{had-vis}}$ candidates are first reconstructed as jets with the anti-kt algorithm, with radius parameter of 0.4. The tracks must be within the *core* region ($\Delta R < 0.2$ around the direction of motion, as opposed to the *isolation* region) and satisfy additional requirements for p_T and number of hits [52]; their multiplicity corresponds to the reconstructed prongness of the particle.

After the hadronic jet reconstruction, the object must be identified as the lepton:

¹a u -quark (or d -quark) bounded with its antiparticle.

for this classification task, a recurrent neural network (RNN) is employed [53].

3.3.1 Missing Mass Calculator

In processes involving pair production of τ -leptons (known as di- τ), such as the $HH \rightarrow b\bar{b}\tau\tau$ decay, the reconstruction of the mass of the original resonance (decaying into the two τ -leptons) is made difficult by the production of the undetectable neutrinos. The Missing Mass Calculator (MMC) [54] can be employed to address this issue, increasing the resolution of the di- τ mass compared to using only visible objects information. The algorithm is based on requiring that the relative orientations of the decay products, including the neutrinos, are consistent with the mass and decay kinematics of the τ -leptons, by minimising a likelihood function defined in the kinematically allowed phase space. MMC assumes that the missing p_T is exclusively originating from the di- τ system.

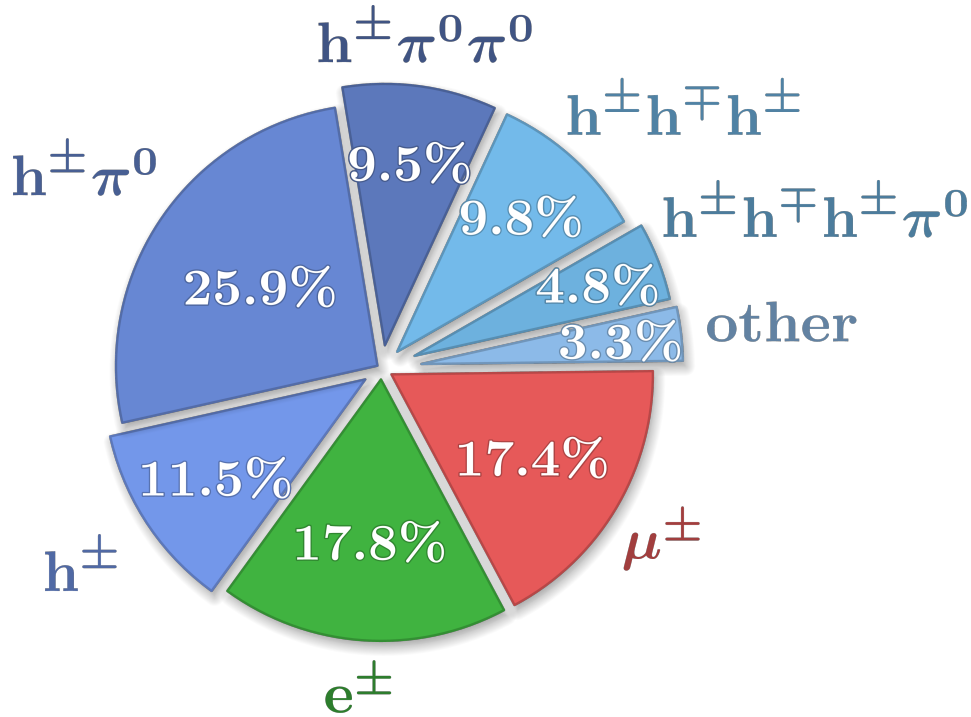


Figure 3.3: The τ -lepton decays: the green and the red sectors are the decays into an electron and a muon respectively, i.e. the leptonic decays. The blue sectors correspond to the hadronic decays, with h standing for a general charged hadron and π^0 standing for the neutral pion [55].

Chapter 4

The $HH \rightarrow bb\tau\tau$ Analysis

The two projects of this thesis, the novel triggers and the transformer-based classifier, are part of the ATLAS $HH \rightarrow bb\tau\tau$ analysis and specifically target the channel in which both τ -leptons decay hadronically, in the following referred to as $\tau_{\text{had}}\tau_{\text{had}}$. Both projects are exploratory studies for the Run 3 analysis, which is currently in its initial phase, and are built upon the Run 2 *Legacy* analysis.

This chapter presents an overview of the Legacy $\tau_{\text{had}}\tau_{\text{had}}$ channel **event selection**. Selection criteria (also known as *cuts*) are defined by an analysis to extract the events that it will study. These criteria can either be *online*, meaning triggers, or *offline*, meaning that they are applied after data has been collected.

The general source of this chapter is Ref. [7].

4.1 The Run 2 Analysis Event Selection

The $HH \rightarrow bb\tau_{\text{had}}\tau_{\text{had}}$ signal leads to a final state with (at leading order) two b -jets and two τ -leptons, or more precisely their (visible) hadronic decay products ($\tau_{\text{had-vis}}$). Many other non-signal physical processes (*backgrounds*) can lead to a similar signature in the detector, most notably those involving the production of a single top quark or pairs thereof ($t\bar{t}$), Z+jets¹, single Higgs boson, and *fakes*. Fakes occur when a poorly reconstructed jet is wrongly identified as a hadronically decaying τ -lepton: even though this probability is small, jet processes with large cross sections can make this contamination non-negligible: the most common fake sources are $t\bar{t}$ events and multi-jet production.

Background events involving $t\bar{t}$, Z+jets and single-top processes are simulated, while jet $\rightarrow \tau_{\text{had}}$ fakes are estimated from data.

The $HH \rightarrow bb\tau\tau$ Legacy analysis selects its events in signal regions (SR) designed to minimise the background contamination while simultaneously keeping the signal contribution high. There are three SRs: one associated to the $\tau_{\text{had}}\tau_{\text{had}}$ channel; two associated to the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, i.e. a hadronically decaying τ -lepton and a leptonically decaying one.

To control the Z+jets and $t\bar{t}$ backgrounds, a dedicated control region (CR) is used, denoted as Z+HF control region.

¹A Z-boson produced in association with two jets (mostly b -jets or c -jets).

4.1.1 Preselection

$\tau_{\text{had}}\tau_{\text{had}}$ events are selected *online* using a combination of single- τ trigger (STT) and di- τ triggers (DTTs), which are discussed in Chapter 5; they are required to have two $\tau_{\text{had-vis}}$ with opposite charges. At the same time, an electron and muon veto is employed to exclude $\tau_{\text{lep}}\tau_{\text{had}}$ events.

The *offline* p_{T} of STT $\tau_{\text{had-vis}}$ have a threshold in the 100-180 GeV range², while DTT (sub-)leading $\tau_{\text{had-vis}}$ have thresholds of 40 GeV (30 GeV). STT events demand a second $\tau_{\text{had-vis}}$ with a 25 GeV p_{T} threshold. Additional offline requirements for DTT are either one jet with an offline p_{T} threshold of 80 GeV, the two $\tau_{\text{had-vis}}$ reconstructed within $\Delta R = 2.5$, or two extra jets with offline p_{T} thresholds of 45 GeV. The STT offline requirements are applied to events passing both STTs and DTTs. The invariant mass of the τ -lepton pair ($m_{\tau\tau}^{\text{MMC}}$) is estimated with the Missing Mass Calculator (MMC); all SR events must have $m_{\tau\tau}^{\text{MMC}} > 60$ GeV to minimise the contamination from $\text{jet} \rightarrow \tau_{\text{had}}$ fakes. To target $H \rightarrow bb$ decays, the events are required to have exactly two b -tagged jets identified by the DL1r algorithm at a working point with an overall 77% efficiency³, with minimum p_{T} thresholds of 45 and 20 GeV respectively (within $|\eta| < 2.5$).

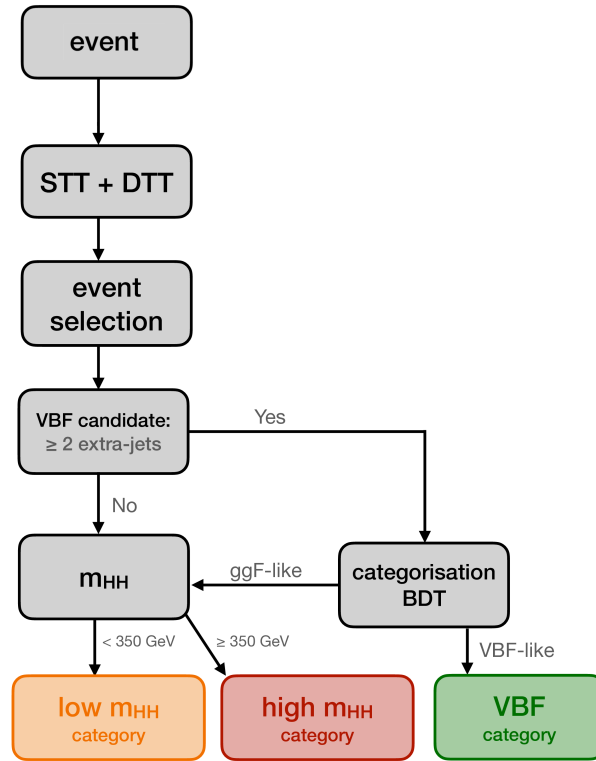


Figure 4.1: The workflow of event categorisation of the $HH \rightarrow bb\tau\tau$ Legacy analysis of Run 2, for the $\tau_{\text{had}}\tau_{\text{had}}$ signal region [7].

²Depending on the year of data-taking.

³Estimated on a semi-leptonic $t\bar{t}$ simulation.

4.1.2 Event Categorisation

Events in each SR are in turn divided between three categories. As seen in Section 1.3.2, the VBF HH production allows to study the coupling modifier κ_{2V} ; to enhance the sensitivity to this parameter and reduce its correlation to κ_λ , a dedicated **VBF category** is defined via a multivariate approach. In each SR, a boosted decision tree (BDT) is designed to select the events with the distinctive characteristics of VBF production mode. The remaining events, targeting ggF HH production, are split into two categories based on the invariant mass of the HH system (m_{HH}), which is significantly affected by the value of κ_λ : a **ggF high- m_{HH} category**, particularly sensitive to the ggF HH signal if κ_λ is SM-like; and a **ggF low- m_{HH} category**, which enhances the sensitivity to larger than SM values of κ_λ as these would lead to a significantly softer m_{HH} spectrum.

To assign an event to one of the three categories, the sequence of steps (depicted in Figure 4.1) is the following:

1. **VBF candidates.** If the event has least two jets in addition to the ones of the $H \rightarrow b\bar{b}$ decay (*extra jets*), it is a VBF *candidate* and gets passed to a classification BDT. If it does not, it is assigned to the ggF categories. Nearly half of the true ggF events pass this requirement, while 80% of the true VBF do.
2. **Categorisation BDT.** VBF candidates are passed to a boosted decision tree (BDT) trained on VBF candidate events to separate ggF from VBF production modes. If the BDT classifies the VBF candidate event as VBF, the event is assigned to the VBF category, otherwise to the ggF categories. The working point of the BDT is chosen to enhance the limit on the VBF signal strength without degrading the limit on the overall HH signal strength or on κ_λ , which were deemed to be priorities.
3. **Low/high- m_{HH} splitting.** Events that failed either the VBF candidate criteria or the BDT classification are in the ggF categories. The threshold of 350 GeV is chosen to split them between the high- m_{HH} and low high- m_{HH} categories; this values is optimised to increase the κ_λ constraint.

4.1.3 Signal/Background Classification

Based on the event categorisation, another set of BDTs is trained and evaluated in each SR to discriminate against the vast variety of background processes. The three categories have different definitions for the signal:

- In the **VBF category**, the signal corresponds to the VBF production of HH according to the Standard Model, i.e. all coupling modifiers are equal to 1, including both κ_λ and κ_{2V} .
- In the **ggF high- m_{HH} category**, the signal corresponds to the ggF production of HH according to the Standard Model, i.e. all coupling modifiers are equal to 1, including κ_λ .
- In the **ggF low- m_{HH} category**, the signal corresponds to the ggF production of HH with $\kappa_\lambda = 10$, significantly deviating from the SM prediction.

While the BDTs of the VBF category is trained on events from all three categories, the two BDT of the ggF categories are trained using events belonging only to their category. The main sources of background for the three categories are displayed in Figure 4.2.

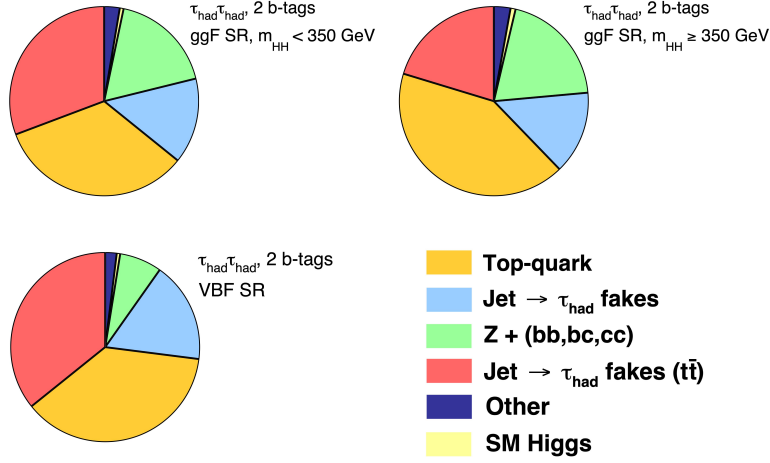


Figure 4.2: The main sources of background for the three categories of the $bb\tau_{\text{had}}\tau_{\text{had}}$ signal region, for a centre-of-mass energy of 13 TeV and a luminosity of 140 fb^{-1} [6].

The signal is extracted from a combined profile likelihood fit [56] to all SRs and the control region. The Run 2 $bb\tau\tau$ Legacy analysis is able to set the following limits⁴:

- κ_λ is constrained to an observed (expected) 95% confidence interval of $-3.2 < \kappa_\lambda < 9.1$ ($-2.5 < \kappa_\lambda < 9.2$).
- κ_{2V} is constrained to an observed (expected) 95% confidence interval of $-0.4 < \kappa_{2V} < 2.6$ ($-0.2 < \kappa_{2V} < 2.4$).
- The HH production signal strength μ_{HH} has an observed 95% confidence level upper limit of 5.9, to be compared with an expected limit of 3.1 in the background-only hypothesis ($\mu_{HH} = 0$).
- The ggF and VBF HH production signal strengths⁵ μ_{ggF} and μ_{VBF} have an observed 95% confidence level upper limit of 5.8 and 91 respectively, to be compared with the expected limits of 3.2 and 72 respectively in the background-only hypothesis ($\mu_{\text{ggF}} = \mu_{\text{VBF}} = 0$).

The constraints are mostly coming from the $\tau_{\text{had}}\tau_{\text{had}}$ channel. No significant excess above the expected SM predictions is observed.

⁴Assuming that all other Higgs boson couplings follow the SM predictions.

⁵The ggF and VBF limits are coming from a fit where both signal strenghts are free parameters at the same time. This works because of the separate VBF region.

Chapter 5

Novel Triggers for $HH \rightarrow bb\tau\tau$

This chapter presents a study involving the validation and optimisation of novel triggers to be employed for the search for Higgs boson pairs in the $bb\tau\tau$ final state during the Run 3 data-taking. This research shows that the total trigger efficiency can be improved from 55% in Run 2 to up to 85% in Run 3.

The motivation for these studies is exposed in Section 5.1; Section 5.2 discusses the experimental methodology; finally, the results are presented in Section 5.3.

5.1 Motivation

The $HH \rightarrow bb\tau_{\text{had}}\tau_{\text{had}}$ process is one of the most sensitive channels to test the Higgs boson self-coupling, and has set one of the most stringent limits on the HH signal strength. However, the all-hadronic final state is also one of the most challenging to select, requiring efficient τ and b -jet triggers.

Trigger optimisation is often a trade-off between increasing the signal efficiency, maintaining a manageable output rate and online computing budget. For this purpose, major help can come from the Run 3 ATLAS upgrades and from the increase usage of the Delayed trigger streams, which allows a larger rate to be stored and analysed when computing resources become available.

The τ -lepton and b -jet triggers need to be highly performing, particularly at low p_T , in both the Standard Model (SM) and beyond the SM scenarios. As illustrated in Figure 5.1, the invariant mass distribution of the HH system is notably influenced by the value of κ_λ . If a substantial number of HH events predominantly occurs in the region below $m_{HH} \approx 350$ GeV, the distribution is significantly deviating from the expectation of the SM. The novel triggers need therefore to be studied and optimised for both scenarios.

The benchmark for these studies are the *baseline* Run 2 triggers: the STT and the DTTs introduced in Chapter 4, which started running in 2018. They are:

- **Single- τ trigger:** requires a τ -lepton with HLT p_T of 160 GeV.
- **Di- τ + 2 jets trigger:** the HLT requires the leading¹ and subleading τ -leptons to have p_T larger than 35 and 25 GeV respectively; Level-1 selects

¹The *leading* and *subleading* objects are the most energetic and the second most energetic objects respectively.

four jets (all with $p_T > 12$ GeV), of which two are reconstructed as τ -leptons ($p_T > 20, 12$ GeV).

- **Di- τ + jet trigger:** it shares the same HLT p_T thresholds for the τ -leptons as the other di- τ trigger; Level-1 requires close-by τ -leptons ($p_T > 20, 12$ GeV) and a single additional jet ($p_T > 25$ GeV). It is employed by the ATLAS $H \rightarrow \tau\tau$ analysis.

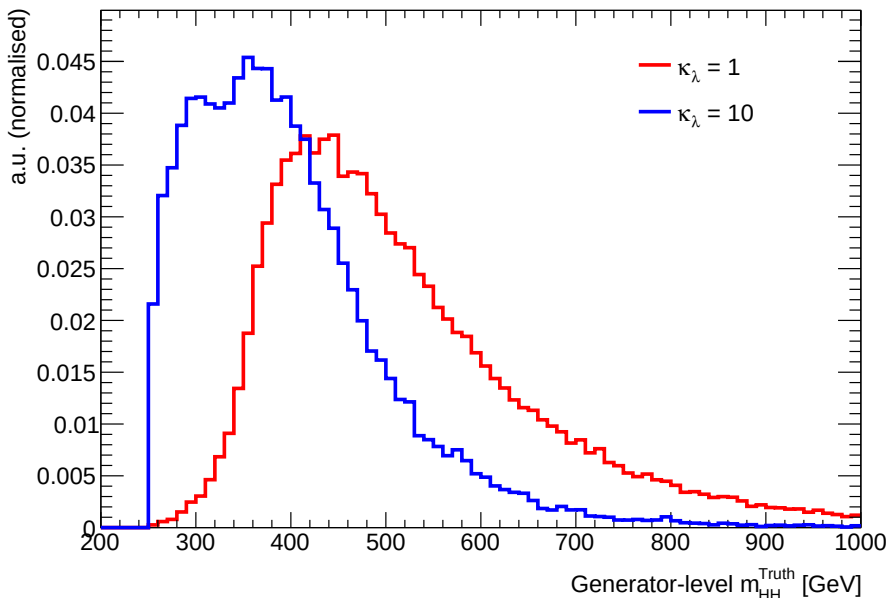


Figure 5.1: The distributions of the HH invariant mass at a centre-of-mass energy of 13 TeV for $\kappa_\lambda = 1$ and $\kappa_\lambda = 10$.

5.2 Methodology

To study the efficiency of novel triggers for the Run 3 $HH \rightarrow bb\tau\tau$ analysis, a python framework was developed as part of this thesis. The framework, named *TrigVsSelect*, offers versatile customisation options and allows to study different combinations of selection cuts. As seen in Chapter 4, cuts can either be online, meaning triggers, or offline, meaning applied by the analysis independently of the data-taking. *TrigVsSelect* allows to combine multiple cuts, calculate their ratio, and compare them. The ratios can also be plotted as a function of the kinematic variables, to examine how the efficiency changes in different kinematic regions, for example as a function of the di-Higgs invariant mass m_{HH} .

This customisability allows different definition of efficiencies to be explored. For instance, it is possible to study:

- The efficiency of a trigger with respect to an offline selection, another trigger or in absolute terms.
- The efficiency of an offline selection with respect to another offline selection or in absolute terms.

The framework enables the comparison of these ratios in the same plot by altering either the numerator (the quantity for which the efficiency is calculated) or the denominator (the reference for calculating the efficiency). Python classes have been built to rapidly calculate and plot these comparisons.

The framework is automatised to also calculate the combinations (in logic OR) of the configured cuts.

The datasets used for calculating the efficiencies are Monte Carlo (MC) based simulations of ATLAS Run 3 conditions ($\sqrt{s} = 13.6$ TeV, full detector simulation) under either $\kappa_\lambda = 1$ or $\kappa_\lambda = 10$ for the $HH \rightarrow bb\tau_{\text{had}}\tau_{\text{had}}$ process in the ggF production mode. In the following, the word *truth* is used to specify that a quantity is taken from MC-based simulation of a particle before any interaction with the detector or reconstruction is simulated.

In support of the studies on trigger performances, the kinematics of the signal event has also been examined. In particular, the most energetic objects in the $bb\tau\tau$ final state have been investigated, as they can be leveraged by triggers.

5.2.1 The Offline Selections

The trigger performances have been studied under different offline selections to provide a comprehensive understanding of their efficiency and sensitivity across varied criteria. The selections are the following:

- **Inclusive selection:** requires the visible p_T of the leading (subleading) τ -lepton to be larger than 25 (15) GeV.
- **Truth-matched selection:** requires reconstructed leading (subleading) τ -lepton with p_T greater than 25 (20) GeV and $|\eta| < 2.5$ and two reconstructed Particle Flow [40] jets with $p_T > 20$ GeV and $|\eta| < 2.5$; the τ -leptons must be geometrically matched² to the (truth-level) decay products of the Higgs boson, and the jets must be geometrically matched to a (truth-level) b -hadron; the τ -leptons must have 1 or 3 tracks.
- **Identified selection:** requires reconstructed leading (subleading) τ -leptons with p_T greater than 25 (20) GeV and $|\eta| < 2.5$ identified with the loose operating point of the RNN identification algorithm [57], and two reconstructed Particle Flow jets with $p_T > 20$ GeV and $|\eta| < 2.5$ identified as b -jets by the GN2 tagging algorithm [49] at the 82% efficiency operating point; the τ -leptons must have 1 or 3 tracks.

The three selections serve distinct purposes. The inclusive selection is the least biased among the three: it corresponds to the filter applied by the simulation generator to avoid objects that would hardly be reconstructed by ATLAS (e.g. due to a particularly low p_T). The truth-matched selection is the ideal limit of the identification algorithms: it conveys what would be technically possible if the analysis had perfect identification performance. Finally, the identified selection delineates what is at the moment realistically doable with the current algorithms; this selection, which does not have any truth-level information, is the closest to the ongoing $HH \rightarrow bb\tau\tau$

²Within a cone of $\Delta R = 0.3$ around the object axis.

analysis.

Each offline selection has its own efficiency, to be combined with the one of the triggers to fully understand their impact on the analysis. Figure 5.2 shows³ the efficiency of the three selections. The inclusive efficiency is marked as 100% since it is the maximum achievable value in the simulated dataset, being the generator preselection. The truth-matched selection halves this value since it comes with its own inefficiencies, in particular due to the p_T thresholds and the η acceptance. The identified selection (approximating the actual analysis offline selection) reaches 18%: this is an estimation of the analysis own efficiency, coming for instance from the identification algorithms efficiencies.

Figure 5.3 illustrates the varying impact of the triggers across samples with different offline selections. Considering the Run 2 τ triggers as an example, the plot demonstrates how *the tighter the offline selection, the larger the trigger efficiency with respect to it*.

³All the histograms of the chapter are displayed with *overflow bins*: the rightmost bin includes the data values that exceed the plotted range. In addition, all the bins are presented with vertical error bars indicating the statistical uncertainty.

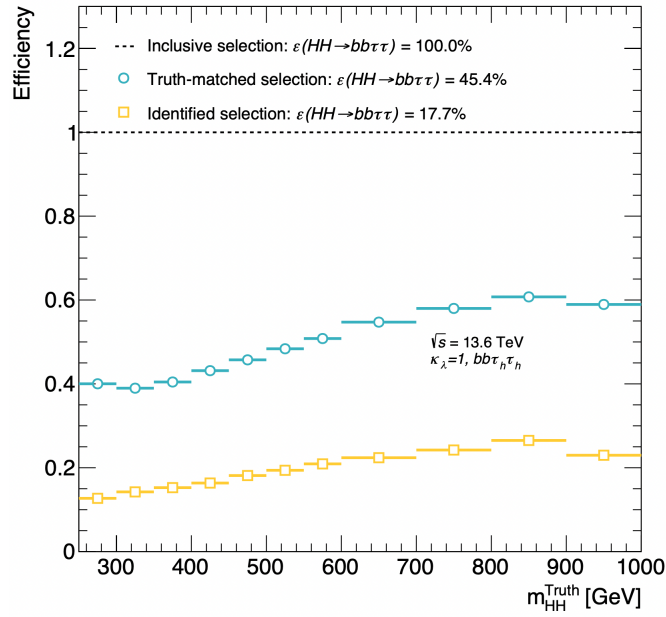


Figure 5.2: The efficiencies of the three offline selections, as a function of the truth-level m_{HH} . The upper limit of full efficiency (dotted line) corresponds to the generator-level preselection, i.e. the inclusive selection.

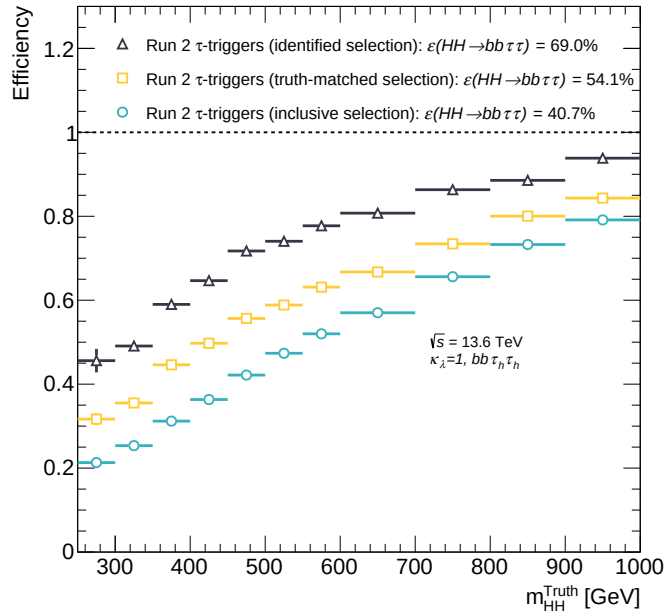


Figure 5.3: The influence of offline selections on trigger efficiency, depicted as a function of truth-level m_{HH} . Noticeably, triggers exhibit higher efficiency for samples selected with more stringent offline criteria. As an example, the impact of Run 2 τ triggers is displayed.

5.2.2 The Trigger Chains

The trigger chains examined in this project are written in full in Table 5.1, together with the names used to refer to them.

Trigger name	Trigger chain
Single- τ trigger	HLT_tau160_mediumRNN_tracktwoMVA_ L1TAU100
Run 2 di- τ + jet	HLT_tau35_mediumRNN_tracktwoMVA_tau25_ mediumRNN_tracktwoMVA_03dRAB30_ L1DR_TAU20ITAU12I_J25
Run 2 di- τ + 2 jets	HLT_tau35_mediumRNN_tracktwoMVA_tau25_ mediumRNN_tracktwoMVA_03dRAB_ L1TAU20IM_2TAU12IM_4J12p0ETA25
Run 3 di- τ + jet	HLT_tau30_mediumRNN_tracktwoMVA_tau20_ mediumRNN_tracktwoMVA_03dRAB30_ L1DR_TAU20ITAU12I_J25
Run 3 di- τ + 2 jets	HLT_tau30_mediumRNN_tracktwoMVA_tau20_ mediumRNN_tracktwoMVA_03dRAB_ L1TAU20IM_2TAU12IM_4J12p0ETA25
4 jets (2 b -tagged) ($HH \rightarrow 4b$ trigger)	HLT_j75c_020jvt_j50c_020jvt_j25c_020jvt_j20c_ 020jvt_SHARED_2j20c_020jvt_bgn177_pf_ftf_ presel2c20XX2c20b85_ L1J45p0ETA21_3J15p0ETA25
$b+\tau$ trigger	HLT_tau25_mediumRNN_tracktwoMVA_j75c_020jvt_ j50c_020jvt_j25c_020jvt_j20c_020jvt_SHARED_j20c_ 020jvt_bgn185_pf_ftf_presel3c20XX1c20b85_ L1J45p0ETA21_3J15p0ETA25

Table 5.1: The trigger chains analysed in this chapter.

In the trigger chains, the first portion in lowercase letters refers to the HLT requirements; the second portion in uppercase refers to the Level-1.

The tags inside the chains can be read in the following way:

- **TAUXX**: L1 τ -lepton with $p_T > XX$.
- **JXX**: L1 jet with $p_T > XX$.
- **pXETAY**: within $X < |\eta| < Y$ (L1).
- **tauxx_mediumRNN**: HLT τ -lepton with $p_T > xx$, identified by the medium efficiency operating point of the HLT RNN.
- **jxxc_020jvt**: HLT jet passing the JVT cut with $p_T > xx$.
- **xxdRAByy**: the two preceding objects must be within $xx < \Delta R < yy$ (HLT).
- **SHARED_xjyyc_020jvt_bgn1zz**: an x number of HLT jets with $p_T > yy$ passing the JVT cut and b -tagged by GN1 at the $zz\%$ efficiency operating point.

The tags representing objects can have a prefix indicating the number of objects the tag is referred to.

5.3 Results

5.3.1 The Event Kinematics

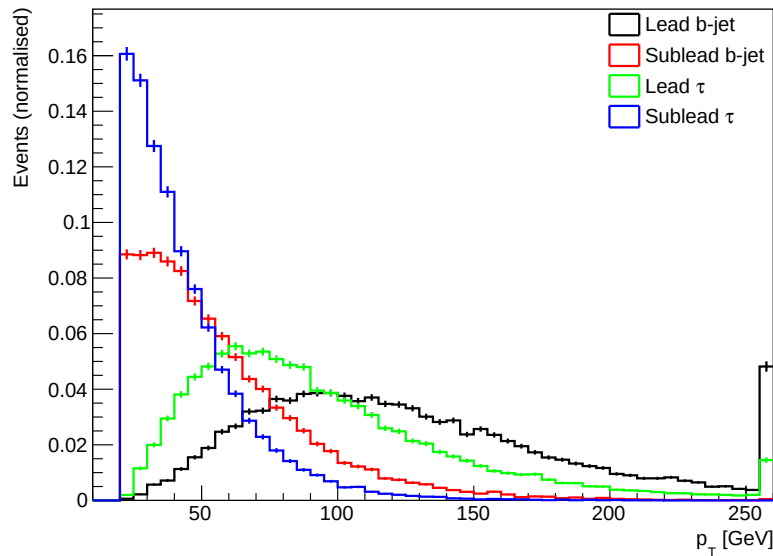


Figure 5.4: The p_T distribution of the four signature objects of the $HH \rightarrow bb\tau\tau$ process at $\sqrt{s} = 13.6$ TeV.

In support of the efficiency studies, an examination of the event kinematics was conducted. At leading order, the $HH \rightarrow bb\tau\tau$ signal has a final state composed of four objects, each with a distinct p_T distribution. They have been plotted for the truth-matched selection in Figure 5.4 for $\kappa_\lambda = 1$. On average the most energetic object is the leading b -jet, followed by the leading τ -lepton, the subleading b -jet and the subleading τ -lepton. From these distributions, the frequency of each permutation of the order of the four objects p_T was calculated. The results have been plotted in Figure 5.5, illustrating large variation between the different orders. Notably, in the majority of cases, the two most energetic objects belong to different types rather than the same type (see Figure 5.6). Analogous results are obtained for $\kappa_\lambda = 10$ (see Figure A.1 in the Appendix).

These plots hold significant relevance for trigger studies, as the most energetic objects are typically the ones targeted by triggers.

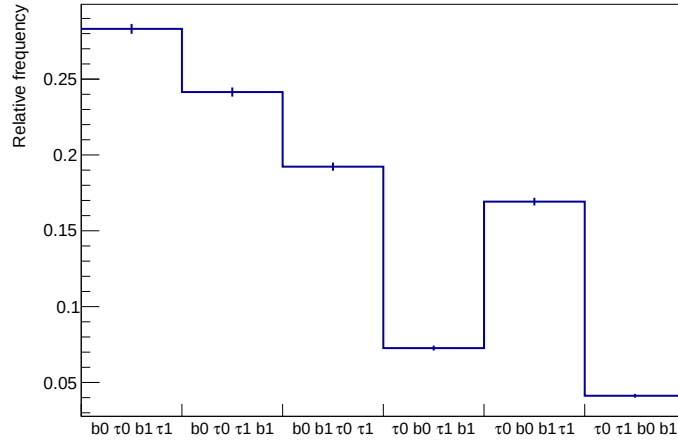


Figure 5.5: The relative frequency distribution of all potential p_T orders of the signature objects; the leading object of each type is denoted with 0, the subleading with 1.

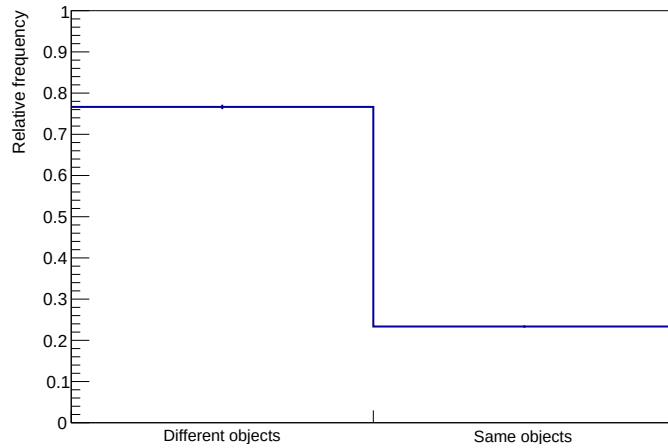


Figure 5.6: The occurrences where the two most energetic decay products are of different types, juxtaposed with instances where they share the same type.

5.3.2 Low p_T τ Triggers

To enhance the efficiency of the $HH \rightarrow bb\tau\tau$ signal for Run 3 triggers, a straightforward approach is to retain the same Run 2 chains while lowering the required p_T thresholds. As depicted in Figure 5.4, among the final state objects the leading τ -lepton and the leading b -jet rarely have p_T values below 20 GeV. This value can therefore serve as a lower bound when establishing the new trigger thresholds tailored to the $bb\tau\tau$ signal. For Run 3, the di- τ trigger chains have been modified to reduce the leading and subleading τ -lepton p_T from 35 and 25 GeV to 30 and 20 GeV at HLT, while maintaining the same Level-1 requirements. The single- τ chain remains unaltered.

By relaxing the thresholds, the new triggers fully encompass the baseline triggers, ensuring no signal events are lost compared to them. This results in increased efficiency but simultaneously leads to a higher trigger rate. In particular:

- The di- τ + 2 jet trigger rate increased by 65%, from 23.1 Hz to 38.0 Hz;
- The di- τ + jet trigger rate almost doubled, increasing by 95% from 41.3 Hz to 80.4 Hz.

The application of these chains has been facilitated by the ATLAS Delayed stream, capable of handling such rates, and they have been operational since May 2023. In contrast, the rates of the baseline chains allow them to run in the Main physics stream, where they are still employed for Run 3.

The combination of the new di- τ triggers with the single- τ trigger is referred to as *low p_T τ triggers* or *Run 3 τ triggers*. The improvement in efficiency compared to the baseline triggers has been calculated using *TrigVsSelect*.

Figure 5.7 displays the efficiency of the individual benchmark triggers from Run 2 as a function of m_{HH} for the truth-matched and identified selections (for $\kappa_\lambda = 1$). Figure 5.8 illustrates the improvement of the Run 3 τ triggers compared to the Run 2 triggers, including the ratio between the two in the subplots. The full set of plots, calculating the efficiency for the three offline selections and both $\kappa_\lambda = 1$ and $\kappa_\lambda = 10$, can be found in the Appendix in Figure A.2 and Figure A.3.

Table 5.2 summarises the gains for both $\kappa_\lambda = 1$ and $\kappa_\lambda = 10$ (identified selection). Significant improvements are observed for $\kappa_\lambda = 10$: reducing the threshold of two of the four final state objects allows for better exploration of the lower m_{HH} tail (see Figure 5.1), which is critical for sensitivity to higher κ_λ values. At the same time, this m_{HH} region is also where the background is the largest.

Trigger	$\kappa_\lambda = 1$	$\kappa_\lambda = 10$
Low p_T τ triggers	74.4%	63.0%
Baseline Run 2 τ triggers	69.0%	55.6%
Relative Gain	+7.8%	+13.3%

Table 5.2: Improvements in the trigger efficiency of the Run 3 τ triggers, compared to Run 2 (identified selection).

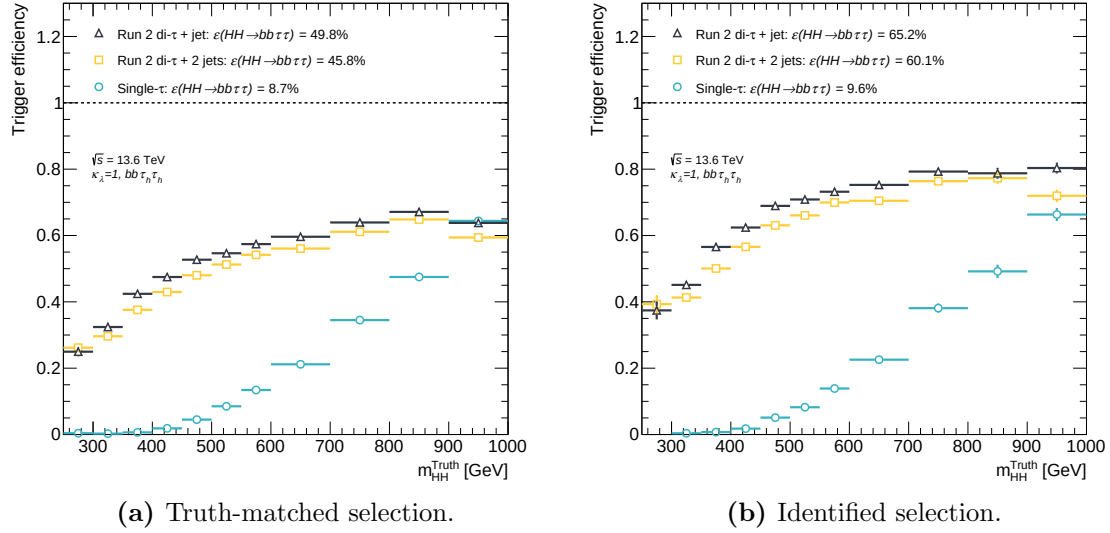


Figure 5.7: The efficiency of the individual Run 2 τ triggers plotted as a function of m_{HH}^{Truth} for $\kappa_\lambda = 1$. The triggers are studied under Run 3 conditions ($\sqrt{s} = 13.6$ TeV). Their combination is the benchmark of this chapter.

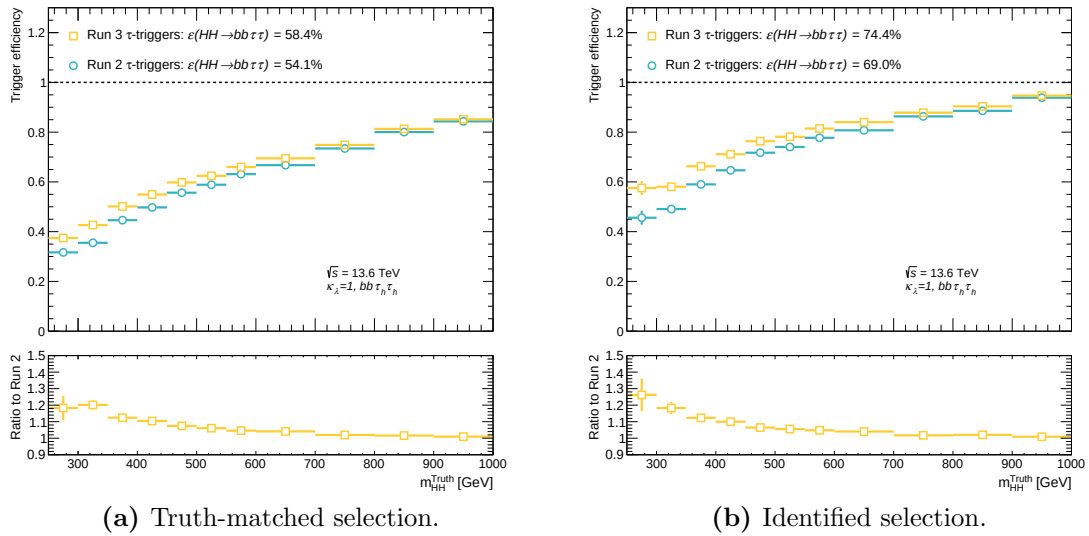


Figure 5.8: The efficiency of the Run 3 τ triggers plotted as a function of m_{HH}^{Truth} for $\kappa_\lambda = 1$. The Run 2 τ triggers serve as the benchmark, with the subplots illustrating the relative efficiency gains with respect to them.

5.3.3 $HH \rightarrow 4b$ Delayed Stream Trigger

Rather than reducing thresholds, another approach to improve signal efficiency is to exploit event kinematics. This involves combining different trigger chains to capture the final state signature more effectively. The approach is exemplified by the adoption of the $HH \rightarrow 4b$ Delayed Stream Trigger in the $HH \rightarrow bb\tau\tau$ analysis. In this chain, Level-1 requires three jets with $p_T > 15$ GeV, with at least one exceeding $p_T > 45$ GeV; the High-Level Trigger (HLT) instead requires four central jets with p_T thresholds of 75, 50, 25, and 20 GeV, among which two must be b -tagged by GN1 at the 77% efficiency operating point. No specific requirements are set for τ -leptons; however, the trigger is still relevant for $bb\tau_{\text{had}}\tau_{\text{had}}$ because hadronically decaying τ -leptons are at first reconstructed as jets (as discussed in Chapter 3). A modified version of this trigger has been active since 2022, featuring different thresholds and incorporating the DL1d tagger; in 2023 improved thresholds were implemented and the more efficient GN1 was integrated. Similar to the low- p_T τ triggers, this trigger operates within the Delayed stream as it has a relatively high rate of ~ 160 Hz.

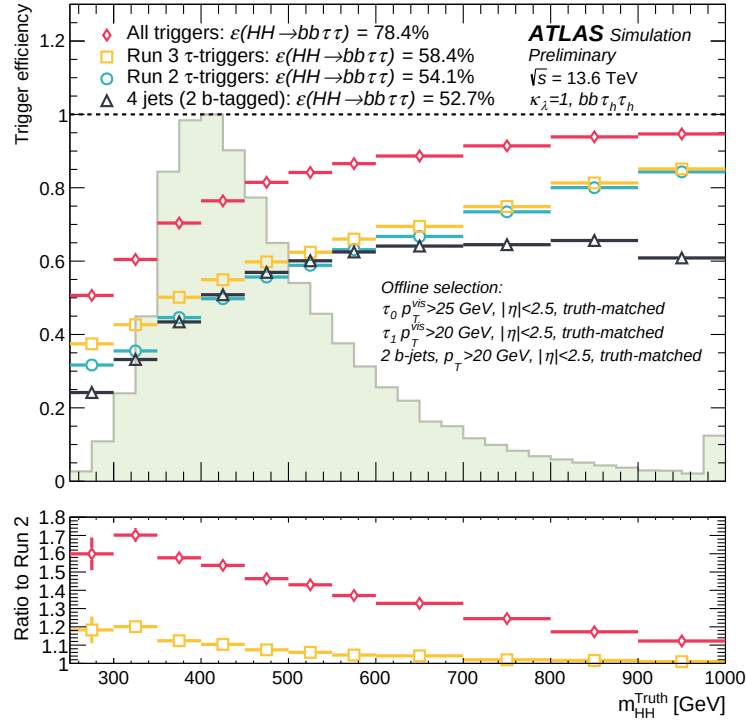
As a jet-only trigger, this chain was tailored to the $HH \rightarrow bbbb$ analysis. However, efficiency studies reveal that it can be extremely useful to the $bb\tau\tau$ analysis as well. The performance of the chain in the truth-matched selection and identified selection ($\kappa_\lambda = 1$) is plotted in Figure 5.9 as a function of m_{HH} , where it is indicated as *4 jets (2 b-tagged)*; its combination with the Run 3 τ triggers (referred to as *All triggers* in the plots) is also shown. While the chain may not exhibit exceptional performance on the $bb\tau\tau$ signal if adopted alone, a noteworthy improvement is observed when combined with the τ triggers. In particular, considering the truth-matched selection, the $HH \rightarrow 4b$ trigger alone is only 52% efficient for the $HH \rightarrow bb\tau\tau$ process, while the Run 3 τ triggers efficiency is 58%; however, combined they reach up to 78%, corresponding to a relative gain of almost 50% and 35% respectively. This value also corresponds to a relative gain of almost 45% compared to the benchmark Run 2 τ triggers.

This effect can be attributed to the complementary behavior of the jet-only trigger and the Run 3 τ triggers in tagging the final state objects of the analysis: together, they impose requirements on 2 b -jets and 2 τ -leptons, effectively encompassing all signature objects.

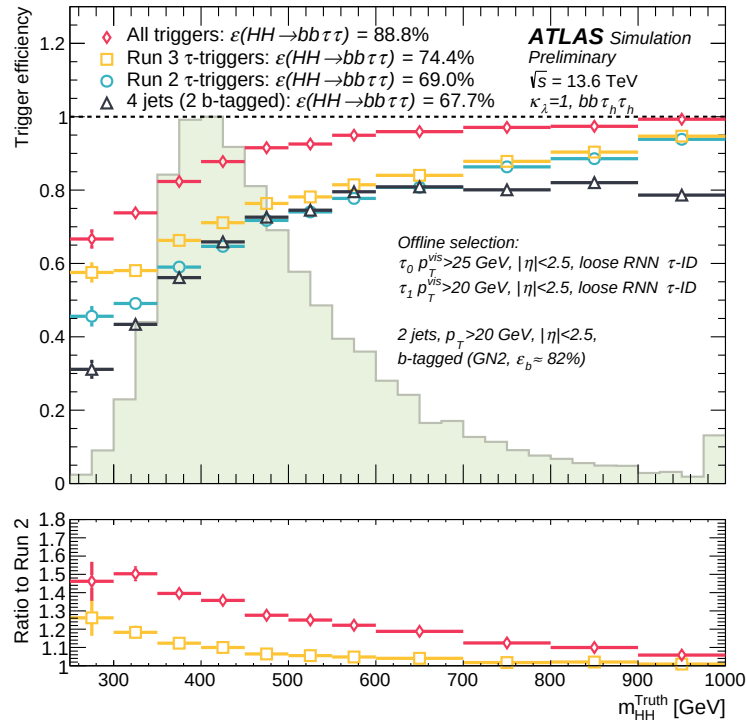
The efficiency improvements are concentrated in the low m_{HH} region close to the $2m_H$ threshold, where most of the signal goes for high κ_λ .

These plots, produced as a result of this thesis, have been made public by ATLAS as an official result of the Experiment. They can be found in Ref. [58].

The full set of plots, calculating the efficiency improvements for the three offline selections and both $\kappa_\lambda = 1$ and $\kappa_\lambda = 10$, can be found in the Appendix in Figure A.4.



(a) Truth-matched selection.



(b) Identified selection.

Figure 5.9: The plots officially published by the ATLAS Experiment, produced as an outcome of this thesis. They illustrate the efficiency of the $HH \rightarrow 4b$ Delayed stream trigger (indicated as *4 jets (2 b-tagged)*), the Run 3 τ triggers and their combination as a function of m_{HH}^{Truth} for $\kappa_\lambda = 1$. The Run 2 τ triggers serve as the benchmark, with the subplots illustrating the relative efficiency gains with respect to them. The m_{HH}^{Truth} distribution for HH production in the Standard Model is shown in the background in an arbitrary scale for the corresponding offline selection.

5.3.4 Proposed $b + \tau$ Triggers

After combining the τ triggers with the jet-only trigger, the next step is to design a unified trigger to take advantage of the event kinematics.

The analysis final state comprises two τ -leptons and two b -jets, but it suffices to tag one object of each type. To show this, assuming similar efficiencies for b -jet and τ -lepton identification ($\varepsilon_\tau \approx \varepsilon_b$, with the example value $\varepsilon = 0.7$):

- The identification efficiency for a trigger targeting 2 b -jets (like the $HH \rightarrow 4b$ chain) is $\varepsilon_{2b} \approx 0.5$, i.e. the square of the efficiency value, as both objects of the same kind must be identified for the trigger to work;
- The identification efficiency for a trigger targeting 2 τ -leptons (like the di- τ trigger) is $\varepsilon_{2\tau} \approx 0.5$, for the same reason as the 2 b -jets;
- The identification efficiency of the combination (in logic OR) of these two triggers is then given by $\varepsilon_{2b||2\tau} \approx 0.77$;
- The identification efficiency for a trigger that only needs to identify one object of each kind (one τ -lepton and one b -jet) is given by $\varepsilon_{b+\tau} \approx 0.83$.

This idea is supported by the fact that the most energetic objects of the final state, meaning the objects that are typically triggered on, are often one b -jet and one τ -lepton, as seen in Figure 5.5.

The triggers that require the identification of a single b -jet and a single τ -lepton are referred to as $b + \tau$. The design of a proposed $b + \tau$ trigger is currently in the optimisation phase: rate, CPU costs and HLT thresholds are under study.

All the $b + \tau$ designs analysed in this thesis share the same L1 requirements as the Delayed $HH \rightarrow 4b$ trigger, along with identical 4-jet p_T thresholds at HLT. As opposite to the jet-only trigger, the $b + \tau$ chains introduce a τ -lepton identified at HLT by an RNN, and require only one b -jet (instead of two) tagged by GN1 at HLT. The possible designs differ in the following HLT requirements:

- The p_T threshold of the τ -lepton, with proposed values of 20, 25, 35 GeV;
- The efficiency of the RNN identification algorithm, with the τ -lepton identified at either a loose or medium efficiency operating point;
- The efficiency of the GN1 b -tagging: proposed values of the operating point are 70%, 77% and 85% identification efficiency.

Some of the proposed $b + \tau$ chains are visualised in Figure 5.10, with the efficiencies as a function of the rates for both the truth-matched and identified selections. The spectrum of possible chains span a wide range of rates, forming an efficiency curve that overshadows all the other triggers described up to now, reaching an average of 80% for the identified selection.

For comparison with the other triggers studied, a specific $b + \tau$ trigger is chosen as an example: the chain with medium RNN, GN1 at 85% efficiency, and τ -lepton p_T threshold of 25 GeV. While having quite some high rate, this choice will allow a good flexibility for future updates of the analysis selection. In Figure 5.11, a comparison between this design and the $HH \rightarrow 4b$ trigger is shown, plotting also the combination of the two triggers. Since the combination is following closely the $b + \tau$

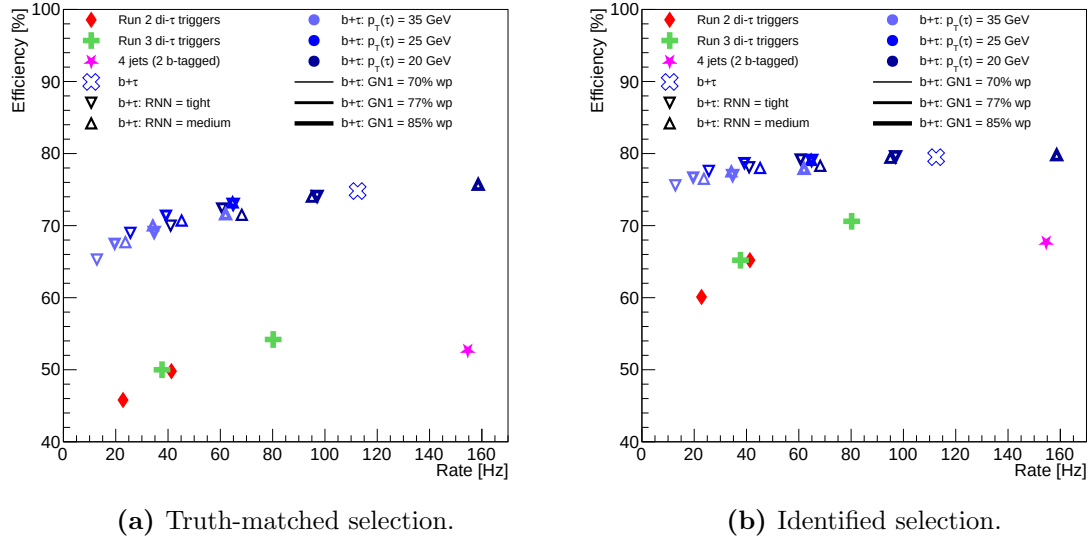


Figure 5.10: The efficiency of the triggers studied in this chapter as a function of their rate. The rates are defined with respect to a luminosity of $L = 2 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. The orientation, color and line width of the $b + \tau$ markers represent the RNN identification efficiency operating point, the τ -lepton p_T threshold and the GN1 efficiency operating point respectively. The white cross is the $b + \tau$ analysed in this chapter's plots.

curve in both selections, it is possible to infer that the $b + \tau$ contains information from the $HH \rightarrow 4b$ trigger, but not vice versa: few events are expected to be lost when substituting the jet-only trigger with the $b + \tau$ trigger.

Figure 5.12 displays the improvement that the $b + \tau$ trigger is expected to bring to the $HH \rightarrow b\bar{b}\tau\tau$ analysis: the combination with the Run 3 τ -triggers is the proposed online requirement for the upcoming analysis in substitution of the Run 2 STT and DTTs. The Run 2 triggers are also depicted for reference, and the subplots are showing the ratio compared to them. These studies shows that the trigger efficiency can be improved from 54% in Run 2 to 85% in Run 3 (truth-matched selection), corresponding to a 56% relative gain. The total efficiency spikes up to more than 90% for the identified selection. These values are expected to increase even more after the design optimisation is concluded: the $b + \tau$ trigger therefore has the potential to bring large benefits to the analysis.

The full set of plots, calculating the efficiency improvements for the three offline selections and both $\kappa_\lambda = 1$ and $\kappa_\lambda = 10$, can be found in the Appendix in Figure A.5.

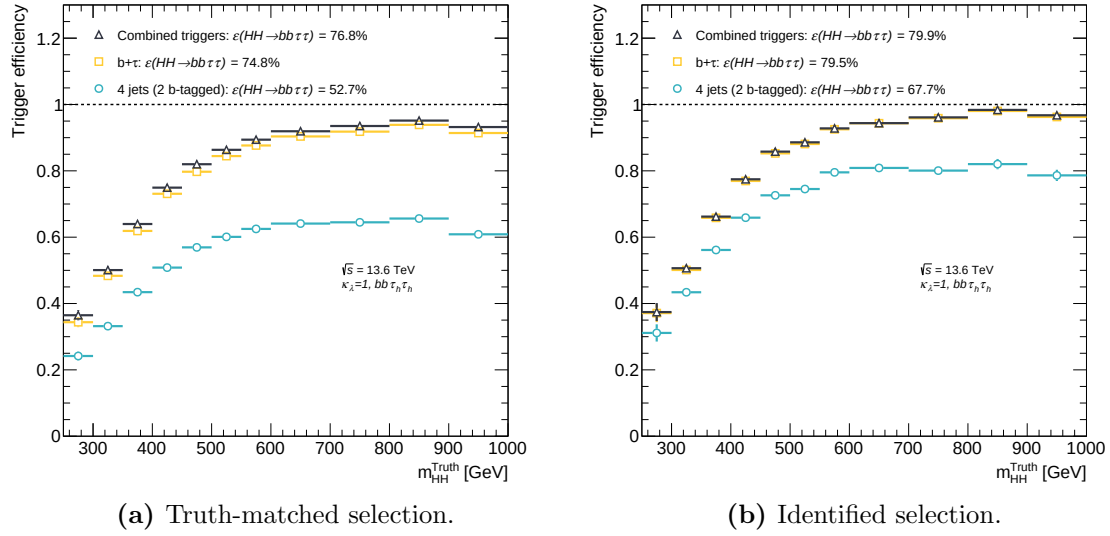


Figure 5.11: The efficiency of the $HH \rightarrow 4b$ trigger (indicated as $4 \text{ jets } (2 \text{ b-tagged})$), the $b + \tau$ and their combination plotted as a function of m_{HH}^{Truth} for $\kappa_\lambda = 1$. The closely aligned curves of the latter two imply that the events selected by the $HH \rightarrow 4b$ trigger are almost fully included within those selected by the $b + \tau$ trigger.

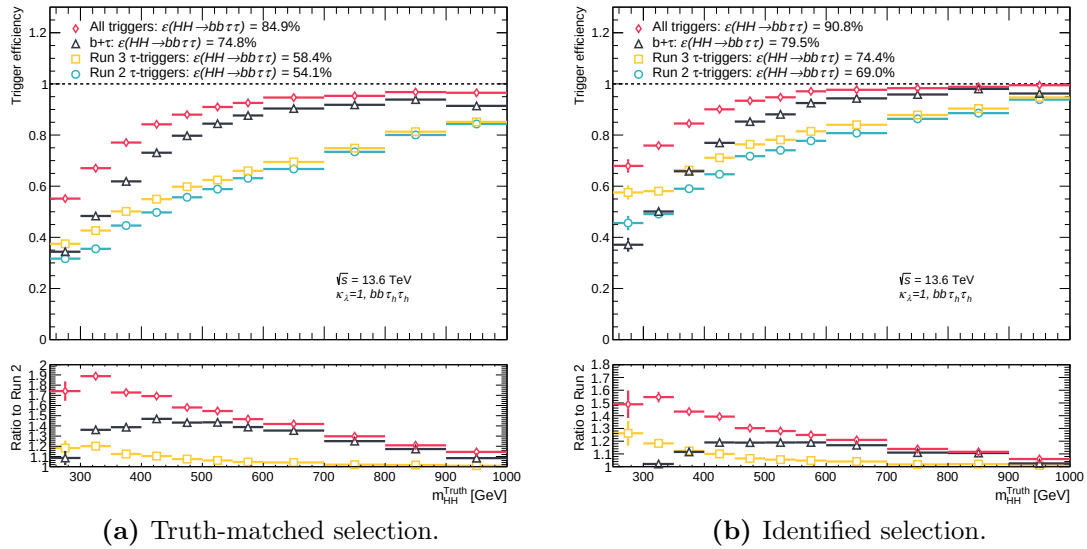


Figure 5.12: The efficiency of the $b + \tau$ trigger, the Run 3 τ triggers and their combination plotted as a function of m_{HH}^{Truth} for $\kappa_\lambda = 1$. The Run 2 τ triggers serve as the benchmark, with the subplots illustrating the relative efficiency gains with respect to them.

Chapter 6

A Transformer for $HH \rightarrow bb\tau\tau$ Event Classification

This chapter presents the work done on the $HH \rightarrow bb\tau\tau$ signal events classification through an adaptation of **GN2**, an algorithm based on a graph neural network. Since GN2 was originally designed for flavour tagging, a re-structuring is required to effectively handle event classification. Following the transformer architecture [50], GN2 is a state-of-the-art model with the potential to be employed for a diverse range of tasks. The studies of this thesis mark the first time that GN2 has been successfully adapted for a task beyond flavour tagging.

Firstly, the motivation for this project is exposed in Section 6.1; then, an overview of neural networks is given in Section 6.2; Section 6.3 discusses the experimental methodology, including the designed algorithm; finally, the results are presented in Section 6.4.

6.1 Motivation

The task of signal events classification involves distinguishing the di-Higgs events produced in the VBF mode from those produced in the ggF mode. As discussed in Chapter 4, the classification improves the κ_{2V} constraint by creating a dedicated category for VBF events. More importantly, the creation of two separate signal regions reduces the correlation between κ_λ and κ_{2V} , such that one can be constrained well without having to make assumptions about the other.

Classification tasks are often performed through machine learning (ML) algorithms. The classifier used in the Run2 $HH \rightarrow bb\tau\tau$ analysis is a BDT (introduced in Chapter 4). The aim of this project is to surpass the performance of the previous analysis with GN2, a new advanced model based on a graph neural network (GNN). For this study, GNNs offer multiple advantages compared to BDTs:

- The input variables of the BDT, displayed in Table 6.1, are all *high-level*, built from other *low-level* variables (such as the four-momentum). While the high-level variables are unique for each event, the low-level variables are associated to the individual physical objects (such as particles/jets). Contrary to the BDT, the GNN can be fed with these low-level variables, allowing for the exploration of more advanced correlations between the physical objects. This

Variable	Description
m_{HH}	Invariant mass of the HH system, reconstructed from the τ -lepton pair (using the MMC) and b -tagged jet pair
$\Delta R(\tau, \tau)$	The ΔR between the two visible τ decay products
VBF $\eta_0 \times \eta_1$	Product of the pseudorapidities of the leading and sub-leading VBF jets
$\Delta\eta_{jj}^{\text{VBF}}$	The $\Delta\eta$ between the two VBF jets
$\Delta\phi_{jj}^{\text{VBF}}$	The $\Delta\phi$ between the two VBF jets
m_{jj}^{VBF}	Invariant mass of the VBF jet system
fwm2($\tau\tau jj$)	2 nd order Fox-Wolfram moment, taking into account the τ -lepton pair and central and forward jets

Table 6.1: Kinematic variables used to train the Run 2 analysis BDT [7].

is possible with GNNs since they are inherently permutation-invariant under their inputs, meaning that the order in which the variables are fed to the model does not affect the predictions, particularly useful when the inputs are the individual physical objects of the events. BDTs, on the contrary, need to have an intrinsic ordering of their inputs: finding an optimal ordering criteria (e.g. ordering by p_T) is not a trivial task.

- These high level variables force the BDT to rely on certain assumptions: for example, the m_{jj}^{VBF} variable (invariant mass of two VBF jet candidates) requires that the event actually has the VBF jet candidates. This limitation confines the application of the BDT to VBF-candidate events (see Figure 4.1), in which the variables are defined. Since the GNN can have low-level inputs, it has a much larger flexibility, and it is possible to apply it to the whole event selection.

6.2 Neural Networks Overview

A neural network (NN) is a machine learning algorithm consisting of interconnected nodes (called *neurons*) and their links (*edges*). Information, in the form of real numbers, is passed between the neurons through the edges. The output of each neuron typically involves a linear combination of all its inputs followed by a non-linear *activation* function, as depicted in Figure 6.1. The linear combination depends on network parameters such as *weights* associated to each edge (setting the strength of the signal of any given connection) or offset values (*biases*). Common activation functions such as *ReLU*, *SiLU* and *sigmoid* are plotted in Figure 6.2; not depicted is *softmax*, a generalization of sigmoid, often used in multi-class classification tasks.

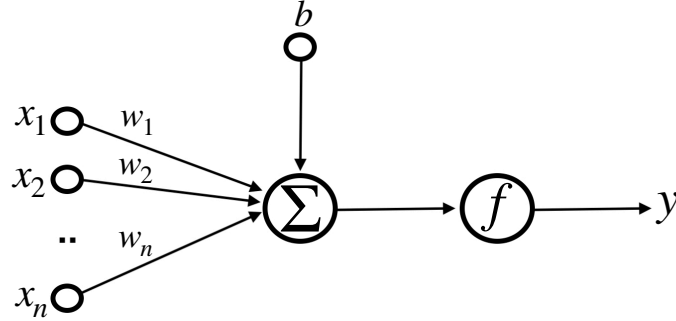


Figure 6.1: The typical computation of a neuron involves a linear combination of its inputs x_i and their associated weights w_i , plus a fixed bias term b . The sum is then passed to a non-linear activation function f to produce the neuron output y .

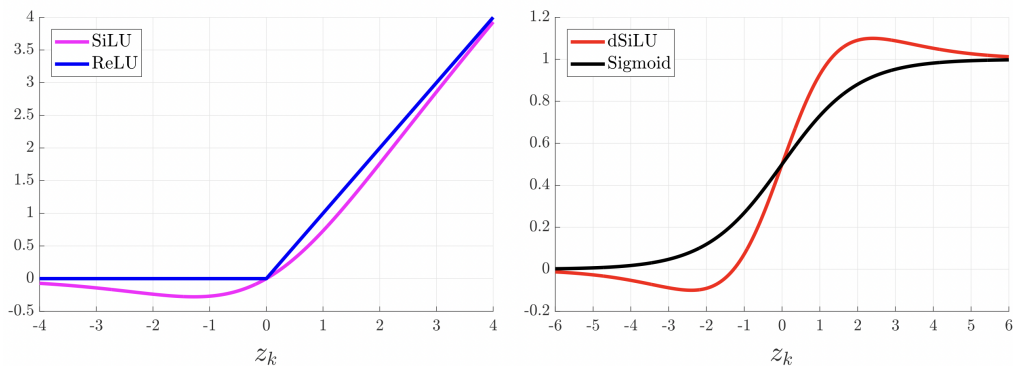


Figure 6.2: Common activation functions: on the left, *rectified linear unit* (ReLU) and *sigmoid linear unit* (SiLU); on the right the *SiLU derivative* and *sigmoid* [59].

6.2.1 Training

The performance of a model is influenced by its parameters, which are divided between *trainable* parameters (such as weights and biases) and *hyperparameters*; while the former are optimised during a preliminary training phase, the latter are not.

In this thesis the training occurs in the *supervised* way: a set of already-classified examples (samples) is passed to the network, after which the algorithm calculates a *loss function* representing the offset between the desired outputs (the actual classifications) and the NN predictions. Iteratively, different sets of samples (*batches*) are fed to the model. After each iteration, the NN updates its parameters with the aim to bring its outputs as close as possible to the desired ones, minimising the loss function (performing a gradient descent in parameters space). The step size at which the model is adjusting its parameters is called *learning rate*. After the entire sample set has been passed to the NN, the same procedure can be repeated in multiple cycles known as *epochs*. The size of the batches, the learning rate and the number of epochs are all relevant hyperparameters. The objective of the training is to learn the patterns that distinguishes the different classes, generalising beyond the provided samples; a network that performs well on the samples it was trained with but fails to generalise beyond them is said to be *overtrained*.

All input variables are commonly normalised to have zero mean and unit variance (*feature scaling*): this facilitates convergence during training and ensures consistent learning across the variables, as they can significantly vary in their ranges (e.g. $\Delta\phi$ and p_T).

Typically three independent sets are used to build a model:

- The *training* set is used for the supervised learning of the model.
- The *validation* set is employed to fine-tune the model, optimising its hyperparameters.
- The *testing* set is reserved for the final evaluation of the model.

Ensuring complete independence among the three sets guarantees a more reliable and unbiased estimation of the model's performance.

6.2.2 Architectures

Neurons are typically organised into layers, with each layer potentially executing distinct transformations of its inputs. There are many paradigms for the organisation of the layers: the *dense feedforward neural network* and the *graph neural network* are two kinds employed by GN2:

Dense Feedforward Neural Networks

In feedforward neural networks (FNNs) the layers process the inputs sequentially, simply transforming the signals with the weighted connections and the activation function; the information flows in a unidirectional manner, from an input layer to an output layer, possibly via intermediate (*hidden*) layers. Neurons in one layer are linked to neurons in the next layer and there are no connections within each layer or feedback loops. Each layer is storing a different representation of the input vector, called *feature* vector. A *dense* network has all the neurons of each layer fully

connected to the neurons of the subsequent layer.

An example of a dense FNN with 3 layers is depicted in Figure 6.3.

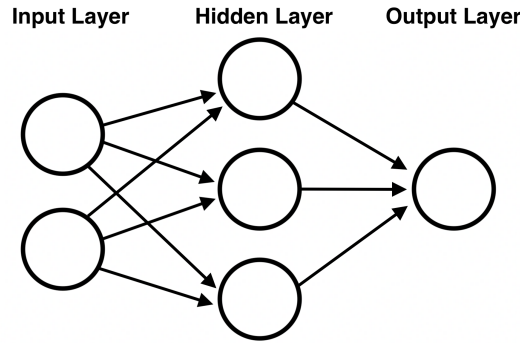


Figure 6.3: A dense feedforward neural network with 3 layers. The circles are the neurons; the connections are the weighted edges. The arrows indicate the flow of the algorithm.

Graph Neural Networks

Graph neural networks (GNNs) are suited for data that can be represented as a graph, with the edges representing (non-linear) data relationships [62]. The layers are the different stages of computation through which information is propagated and updated across **all** the nodes of the graph. For a k number of additional layers, each neuron is aggregating information from the edges and the k -adjacent nodes (nodes separated by k connections). In GNNs the feature vectors can be associated to each node; the size of the feature vectors is called the *embedded dimension*, and it is a hyperparameter of the model. An example of GNN with 3 layers is depicted in Figure 6.4. Classification tasks can be performed by classifying individual nodes, edges or entire graphs.

The state-of-the-art GNN architecture is the graph *attention* network (GAT) [61]. GATs process graph-structured data through the attention mechanism: during information aggregation, each node i dynamically weights the importance given to each neighbouring node j in parallel, calculating an attention score (α_{ij}) which is normalised (typically via the *softmax* function) and then used to calculate the final output of node i . A typical scoring function (GATv2) involves first linearly transforming the nodes i and j into *key* and *query* vectors, concatenating them and applying a *ReLU*-like activation function followed by another linear transformation [63].

A GAT can also have n -ways of independently aggregating information through the attention mechanism, called *attention heads*. More heads means that the model is examining different aspects of the neighborhood at the same time; see Figure 6.5. The recent *transformers* models can be viewed as GATs [64].

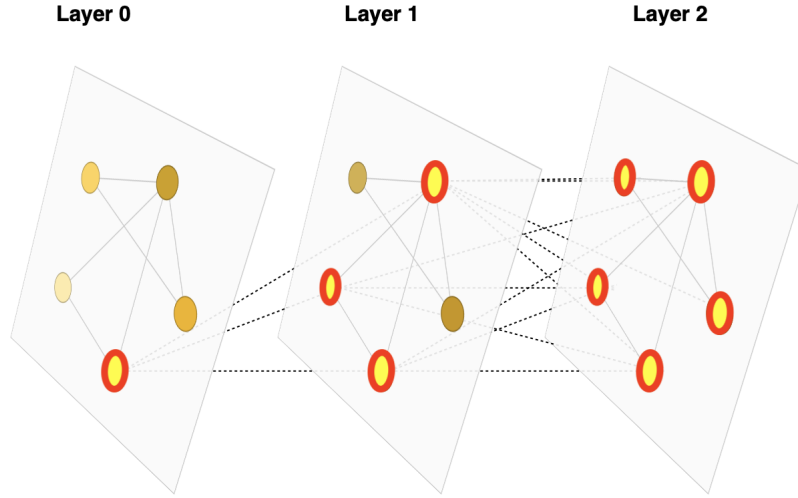


Figure 6.4: A graph neural network with 3 layers. Each layer has the same graph structure, preserving the nodes and their edges (the solid lines). For any given node, the *receptive field* is the set of all nodes that influence its output; with 3 layers, each node has a receptive field that extends up to 2 degrees of separations. In the picture, the receptive field of the bottom node is shown for each layer as the set of highlighted nodes, connected by the dashed lines between the layers. In the second layer the receptive field includes the 2 nodes directly connected via edges to the bottom node; in the third layer it already includes the entire graph. Adapted from Ref. [60].

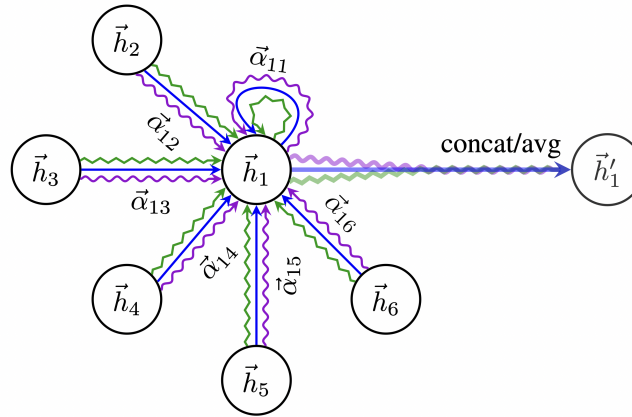


Figure 6.5: Multi-head attention visualised for 3 attention heads. The node with features $\vec{h}_1$ is attending each neighboring node ($\vec{h}_j$) in 3 independent ways (green, blue and purple lines), calculating each time an attention score α_{1j} . The aggregated features from each head are then concatenated or averaged to give the output of the node ($\vec{h}'_1$) [61].

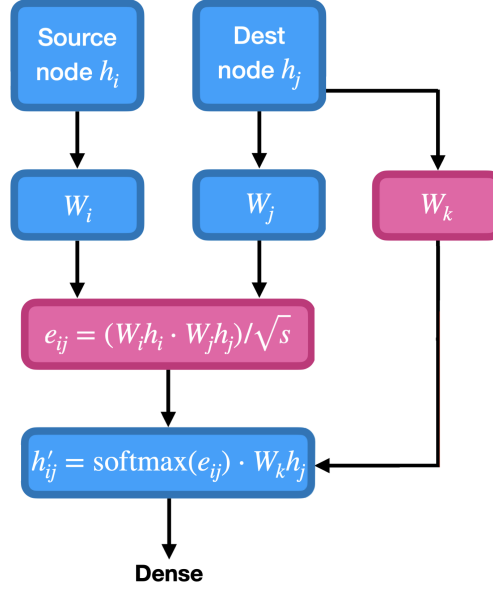


Figure 6.6: The attention mechanism employed by GN2, based on the transformer architecture. A node with features h_j assigns an attention score to the neighbouring node h_i via first linearly transforming their features into *key* and *query* vectors, and then calculating the dot-product between them. The dot-product is scaled by the dimension of the key-query vectors. After softmax normalisation, the score is multiplied by the *value* vector (another linear representation), to get the updated result [50, 65].

6.2.3 GN2

GN2 is a jet flavour tagging algorithm that marks an evolution to the previous GN1 model: it has 1.5 million trainable parameters, and it was trained with 192 million jets [65]. It takes as inputs jet- and track-level kinematic variables, and it concatenates the jet inputs with each track’s inputs. Then, the combined jet-track vectors are initialised in a dense network similar to the Deep Sets model [48]. After initialisation, the per-track outputs are then used to create a fully connected graph, in which each node correspond to a track, storing the initialised feature vector. The vectors are then changed via the multi-head attention mechanism in each layer. In each layer and for each attention head, any given node aggregates the features of the neighbouring nodes in the **scaled dot-product** way, depicted in Figure 6.6, following the transformer architecture [50]. The GNN output consists of representations of each track that are conditional on the other tracks in the jet. These correlated objects are then pooled before being inputted to 3 final dense FNNs for the classification tasks: aside from flavour tagging (achieved via graph classification), GN2 is capable of conducting auxiliary tasks such as predicting the track origins (node classification) and predicting the vertex compatibility of any pair of tracks (edge classification). The flavour classifier has 3 outputs, corresponding to the probabilities that the jet is a *b*-jet, a *c*-jet or a light-jet. The full sequence, derived from GN1, is displayed in Figure 6.7.

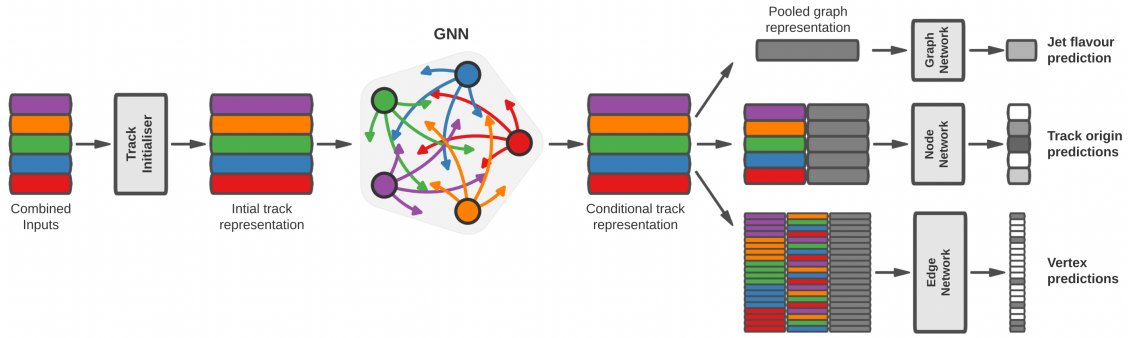


Figure 6.7: The sequence of neural network layers of GN2, derived from GN1 [49].

6.3 Methodology

The dataset of events employed for this project is the official simulated dataset with the preselection of the Run 2 analysis (see Chapter 4), which allows for a meaningful comparison with the Run 2 BDT performance [7]. Due to its simulated nature, the dataset comes pre-labeled with truth-level classes, allowing for supervised learning. To perform event classification, the following steps were taken:

- **Data preprocessing.** The dataset has to be preprocessed. Tasks include applying the event selection of the $b\bar{b}\tau\tau$ analysis and setting up the supervised learning, including splitting into training, validation, and testing sets. Furthermore, the main part of the preprocessing involves re-structuring the input: GN2 is accessed in a specific environment named **Salt**, a general-purpose framework developed by ATLAS Flavour Tagging Group for configuring and training FTAG models [66]. To generalise the purpose of Salt, a script was developed in the context of this thesis to convert event datasets into a format suitable for this framework.
- **Model design and training.** The model was built via a configuration file in Salt, following the GN2 template. The input features were selected, and the model was trained using the training set. The model hyperparameters were fine-tuned based on the validation set.

The model was evaluated with respect to the testing set. For the analysis of the results, its classification scores were compared to the BDT benchmark.

6.3.1 Data Preprocessing

To allow Salt models to perform event classification, a python toolkit has been developed as part of this thesis. It consists of a preprocessing script with configurable settings, designed to offer flexibility for different analyses, even beyond the $b\bar{b}\tau_{\text{had}}\tau_{\text{had}}$ channel.

The first purpose of this preprocessing is to convert event datasets into a format that can be processed by the Salt framework: Salt flavour tagging models take as inputs the variables of a global entity (the jet) and of its components (the tracks) to classify the global entity (tagging the jet flavour). This global/component structure

Task	Global entity	Components
Jet flavour tagging	Jet	Tracks in the jet
Event classification	Event	Physical objects

Table 6.2: Comparison between the original inputs of Salt models and the adaptation performed in this thesis.

can be leveraged to perform event classification in an analogous way (see Table 6.2), by considering the event itself to be the global entity and the physical objects of the event (such as τ -leptons and jets) to be the components; the new Salt inputs will then be the high-level variables associated to the event and the low-level variables associated to the physical objects; and the task will be classifying the event. Since the high-level variables of the kind discussed in Section 6.1 are unique for the event, they can be associated to the event itself. Examples of variables associated to the physical objects are instead the low-level variables such as the relativistic 4-vector of the objects.

Salt inputs are structured as a single file in HDF5 format [67]. It contains two datasets, in which each line is associated to a different jet:

- A dataset with the kinematic variables of each jet (global entity);
- A dataset with the kinematic variables of all the tracks (components) contained in the jet; variables belonging to the same component object are grouped in tuples.

Therefore, the output of this thesis preprocessing must be a HDF5 file with a dataset for the kinematic variables of the events (global entities) and one for the physical objects (components).

The toolkit also generates a YAML file as output, containing normalisation parameters for the kinematic variables. Salt models require this file to perform feature scaling (see Section 6.2.1).

This is the full sequence of tasks carried by the dataset preprocessing script:

1. **Dataset creation.** The script reads a *TTree* from a ROOT file [68]; a TTree is a ROOT object made of multiple containers storing information about collision events, either simulated (as in this case) or real data. In the configuration section of the script it is possible to define the global- and object- related variables to be extracted from the TTree; for supervised learning, one of them must be the class of the event. After variable extraction, the script constructs a dataset where each line corresponds to an event, and each column corresponds to a variable. At this stage, general variable re-definition can be performed, for example constructing new higher-level variables if needed.
2. **Type conversion.** The non-numerical variable types are translated to numerical ones; this is a necessary step for multiple operations, including feature scaling.
3. **Skimming.** The dataset undergoes filtering based on specified conditions. In addition to the preselection that is already applied to the TTrees, the $HH \rightarrow bb\tau_{\text{had}}\tau_{\text{had}}$ analysis requires precisely two b -tagged jets and events where the two τ -leptons have opposite charges. Given the project’s focus on classifying

production modes, only signal events from either ggF or VBF production are retained, filtering out all backgrounds.

4. **Weights normalisation.** As a simulated dataset, each event is associated to a weight, which is at first assigned by the Monte Carlo generator and then adjusted to correct for known data/Monte Carlo discrepancies estimated from calibration measurements. In this preprocessing stage, for each class, the event weights are normalised to the remaining number of events after skimming, so that the sum of weights is equal between the classes.
5. **Shuffling and splitting.** The events in the dataset are shuffled. They are then randomly divided between the training, testing and validation sets (Table 6.3), while keeping the proportions of the classes constant. Proportions are calculated considering Monte Carlo weights.

Dataset	Events
Training set	48,142
Validation set	48,142
Testing set	47,424
Total	143,708

Table 6.3: Number of events in each of the datasets produced by the preprocessing script for the $HH \rightarrow bb\tau_{had}\tau_{had}$ analysis.

6. **Reshaping.** This task implements the global/component analogy of Table 6.2. It is performed on the training, testing and validation sets independently. The task creates the two datasets required by Salt: one for the event-level variables and the other reserved for the individual physical objects variables. In both cases, the rows represent the events; the global dataset has a single column consisting of the set of event-level variables. The component dataset has instead many tuples as columns, each representing a different physical object; for $HH \rightarrow bb\tau_{had}\tau_{had}$ there are 9 tuples, i.e. 2 (hadronic) τ -leptons, 2 b -jets, the MET and up to 4 extra jets. The extra jets must be padded if absent, meaning that filler elements must be added to ensure that all sequences have the same length.

All tuples (i.e. all objects) have an identical number of entries, each corresponding to a physical low-level variable. Variables of the same kind (for example, the reconstructed mass of the leading jet and the reconstructed mass of the subleading τ -lepton, which are both variables of the TTree) must be located at the same index of each tuple. This allows to abstract a generic variable (in this case, a generic mass belonging to all objects) compatible with the permutation-invariant structure of the GNN and for the Salt input format. As an example, feature scaling is performed by Salt in a single instance across all objects, instead of individually for the different objects. If an object is not associated with the abstracted variable (for example, a τ -lepton does not have an associated output score of a b -tagging algorithm) a padding value is inserted in the corresponding entry.

The full set of abstracted variables in the component dataset is reported in Table 6.4. Among them, a new variable is defined for all objects: the *valid*

boolean, which must be added due to the padding. While not a trainable feature, its role is to signal to Salt that the padded jets are not to be considered physical objects.

7. **Extraction of normalisation parameters.** For each of the kinematic variables planned to be used as input features for the model, the mean and the standard deviation of their distributions are calculated. Masking is applied to the non-valid objects, so that their padded values do not contribute to the calculations. This task is exclusively performed on the training set.
8. **Outputs.** The training, testing and validation sets are saved in three different HDF5 files, each containing a global- and a component- dataset; the scaling dictionary is saved to a YAML file.

Variable for all objects	Construction method
Object type	Entirely new: assigns the value 0 to the b -jets, 1 to the τ -leptons, 2 to the MET, 3 to the extra jets; a padding value of -1 to the padded jets.
Transverse momentum p_T	Abstracted from the MET and from the reconstructed p_T of the b -jets, τ -leptons and extra jets.
Pseudo-rapidity η	Abstracted from the reconstructed η of the b -jets, τ -leptons and extra jets; 0 for the MET.
Azimutal angle ϕ	Abstracted from the reconstructed ϕ of the b -jets, τ -leptons, MET and extra jets.
Mass	Abstracted from the reconstructed mass of the b -jets, τ -leptons and extra jets; 0 for the MET.
b -tagging efficiency operating point	Abstracted from the quantile classes of the DL1r FTAG algorithm of both b -jets; a padding value of -1 to the other objects.
τ -lepton identification efficiency operating point	Created from the low-level τ -id variables, which consist of multiple booleans indicating if the given τ -lepton has passed the tight/medium/loose RNN efficiency operating point. From this set of booleans an integer variable is defined, with the purpose to encode the passed working point in a unique variable. A padding value of -1 is assigned to the other objects.
Prongness	Abstracted from the reconstructed prongness of the 2 τ -leptons; a padding value of -1 to the other objects.
Valid flag	Entirely new: 0 for the padded jets, 1 for all the other objects.

Table 6.4: All the variables in the dataset of the physical objects.

6.3.2 Model Design and Training

Salt allows the creation of configuration files to customise the layers of the model. The event classifier of this thesis was designed following the flavour tagging algorithm

Layer Architecture	Parameter	GN2 (FTAG)	Event Classifier
Initialiser Dense	Layer nodes (after input layer)	256	256 - 128 - 64
	Activation function	ReLU	ReLU
Transformer	Number of layers	4	6
	Attention heads	8	4
	Embedded dimension	256	64
	Dropout	No	Yes (rate: 0.1)
	Output dimension	128	128
Classifier Dense	Layer nodes	128 - 64 - 32 - 4	128 - 64 - 32 - 2
	Number of classes	4	2
	Activation function (hidden layers)	ReLU	ReLU
	Activation function (output layer)	Softmax	Softmax
	Loss function	Cross-entropy	Cross-entropy

Table 6.5: Summary of the event classifier of this thesis and comparison with GN2 of flavour tagging. As seen in Section 6.2.3, the model consists of a transformer between two dense FNNs, one for input initialisation and the other for classification. Between the transformer and the classifier dense, a pooling layer is present. The values for the GN2 parameters are updated as of February 2024 [69].

GN2, with the initialiser and classifier dense networks surrounding the transformer-based GNN. The input of the model is a set of variables belonging to an event; the outputs are two values corresponding to the probabilities that the event was produced in ggF or VBF mode¹. Since the amount of events belonging to the two classes is unbalanced (accounting also for the Monte Carlo weights), training weights are associated to the classes: 1.0 for ggF, 3.47 for VBF. This equalises the sum of weights for both samples.

The final configuration of the model is summarised in Table 6.5, with a comparison with GN2.

For the hidden layers, the chosen activation function is ReLU (rectified linear unit, see Figure 6.2) $R(z)$, defined as:

$$R(z) = \max(0, z), \quad (6.1)$$

which simply deactivates the neurons that have negative values.

For the output layer, the softmax $\sigma(z)$ function is chosen instead:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}}, \quad (6.2)$$

with i being a given class and C being the total number of classes. Since softmax ranges strictly from 0 to 1 and is normalised across all classes, it is ideal for proba-

¹Due to Salt-related technical reason, two output values were defined even if one is the complement of the other. The adopted approach with softmax is equivalent to sigmoid + 1 output value; classification is not affected.

bilities.

The chosen loss function is the cross-entropy L , defined as:

$$L(t, \sigma(z)) = - \sum_i^C t_i \ln(\sigma(z)_i), \quad (6.3)$$

with t being the true probability distributions, and $\sigma(z)$ the probabilities predicted by the model (i.e. the aforementioned softmax).

Dropouts have been inserted between the layers to prevent overtraining. A dropout randomly selects a proportion of the neurons of a layer to be deactivated, so that not all the learned information gets passed onto the following layer, adding an element of generalization to the training set. The chosen dropout rate value is 0.1.

The input features with which the model was trained are the low-level object variables (previously discussed in Table 6.4) excluding the *valid* flag. The model was trained for 30 epochs on a NVIDIA Tesla T4 GPU, with a batch size of 1000 items. The entire training runs for ~ 25 minutes, covering the 48,142 events of the training dataset.

Following GN2, the learning rate is adjusted during training in a cyclical manner by the One-Cycle learning rate scheduler (LRS [70]), with initial value $1e-7$ and maximum value $5e-4$.

The hyperparameters of the model have been manually optimised to achieve the highest performance on the validation set. The validation loss, calculated on 48,142 events, stabilises relatively quickly around the 7th epoch. The epoch with the lowest validation loss is chosen for the evaluation phase of the testing set.

The final model has a total of 113,339 trainable parameters, making it lightweight compared to the flavour tagger GN2.

The configuration file of the final model, named *Event classifier*, has been officially merged to the Salt repository and can be found in Ref. [71].

6.4 Results

The results of the model applied to the training, testing and validation sets are plotted in Figure 6.8.

The sets include both VBF candidates (the events with at least 2 extra jets) and non-VBF candidates. In the Run 2 analysis, the BDT was trained and evaluated on the former events, and was not applied to the latter.

The x-axis represents the output of VBF classification, ranging from 0 to 1; the y-axis represents the relative fraction of events that received the corresponding score. The classes have been normalised to unit. The truth-level class of each output plotted is colour-marked: as expected, the collection of VBF events (blue) is more shifted towards the value of 1, while ggF events (red) are more shifted to 0. From this plots it is possible to qualitatively see that the model has achieved a good separation power.

The three sets are nicely following each other, showing no sign of overtraining. For a quantitative check, a subplot has been included plotting the pulls of the testing and validation with respect to the training. Pulls are defined as:

$$P(x) = \frac{D_{\text{obs}}(x) - D_{\text{exp}}(x)}{\sqrt{\sigma_{\text{obs}}^2(x) + \sigma_{\text{exp}}^2(x)}}, \quad (6.4)$$

with D_{exp} being the expected distribution and D_{obs} the observed one; during checks for overtraining, the training set is the expected distribution. No significant pull has been observed.

Figure 6.9 displays the ROC (Receiver Operating Characteristic) curves. These ROCs are plotting:

- True positive rate: also known as *recall*, is shown on the x-axis. For a given signal class, it is the *efficiency* of the algorithm in detecting the signal. Considering VBF to be the signal class, it is the probability of classifying a true VBF event correctly.
- True negative rate: also known as *specificity*, it is shown on the y-axis. Considering ggF to be the background class, it is the probability of marking a ggF event as ggF. It is equal to $1 - \text{FPR}$ (false positive rate), which in turn is the probability of classifying a ggF event as VBF. It can be associated with the algorithm's ability to reject the background class.

The area under the curve (AUC) is proportional to the performance of the algorithm. The ROCs have a relatively high AUC, being 0.94 for the testing set. The three sets are again following each other tightly, not showing any sign of overtraining.

6.4.1 Comparison to the BDT

The conclusive assessment of the model's performance is the comparison to the benchmark, the Run 2 BDT. For these comparisons, the BDT was also applied to the non-VBF candidate region, which was not done in the Run 2 analysis.

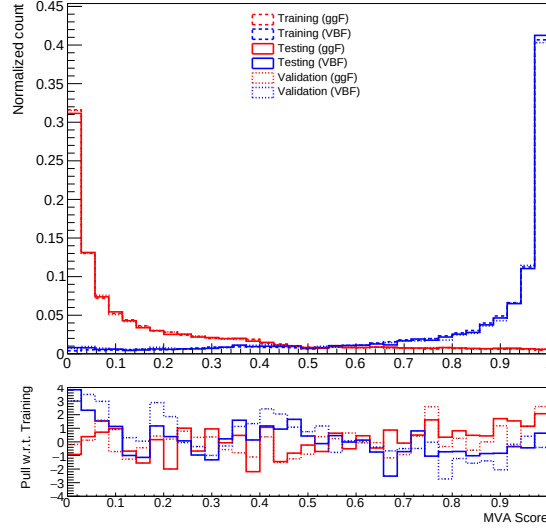
The classification output scores of both models is displayed in Figure 6.10, and the ROC curves are displayed in Figure 6.11. Each category is plotted individually.

The ROCs are also plotted with the ggF rejection defined as $\frac{1}{\text{FPR}}$ (log-scale) in Figure 6.12. A consistent improvement is visible everywhere. Compared to the working point of the Run 2 BDT in the inclusive category, marked with the blue cross, there is a $\sim 15\%$ increase of VBF efficiency for the same ggF contamination; and 100% increase of ggF rejection for the same VBF efficiency. The AUC scores of the two model are compared in Table 6.6.

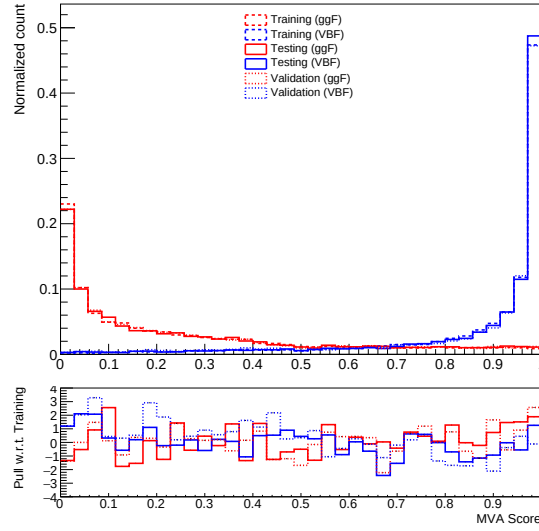
In the subplots, the ratios of the two classifiers is shown, quantifying the improvement in ggF rejection. While for the non-VBF candidate category the improvement was roughly a factor of 1.5 across the VBF efficiencies, inclusively it went up to a factor of 4, and maintained an average of 2 across all VBF efficiencies.

Model	Inclusive	VBF category	Non-VBF category
Run 2 BDT	0.911	0.917	0.894
Transformer-based	0.94	0.94	0.914

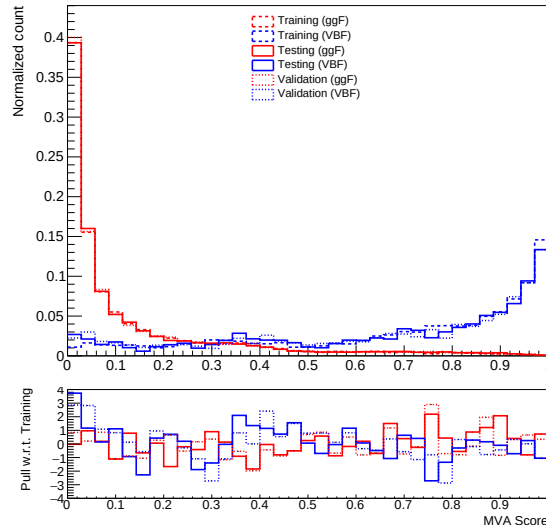
Table 6.6: AUC scores of the event classifier developed in this thesis (applied to the testing set) and of the Run 2 BDT benchmark.



(a) Inclusive

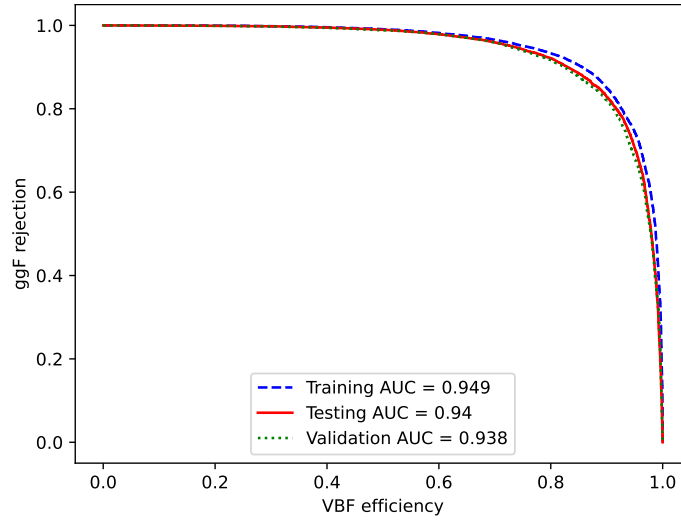


(b) VBF candidates

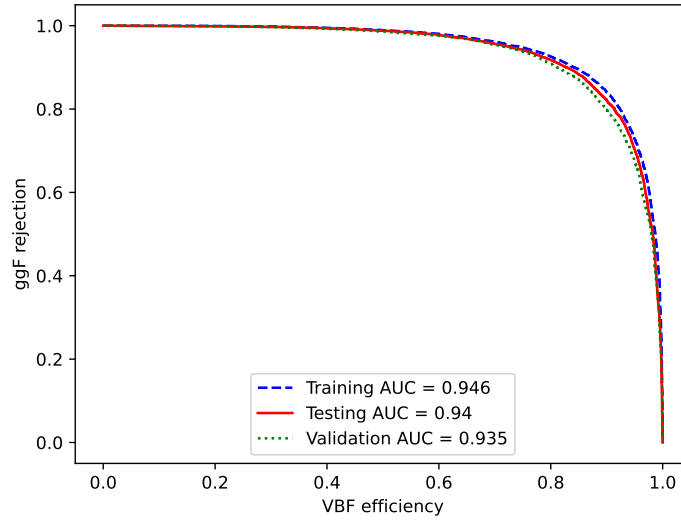


(c) Non-VBF candidates

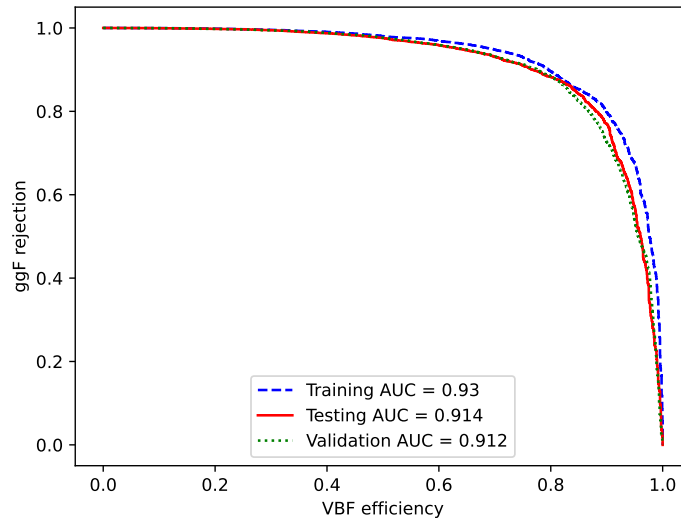
Figure 6.8: Distributions of output scores of the transformer-based event classifier for the training, testing and validation set. The pulls of the validation and testing sets with respect to the training set are displayed in the subplot.



(a) Inclusive

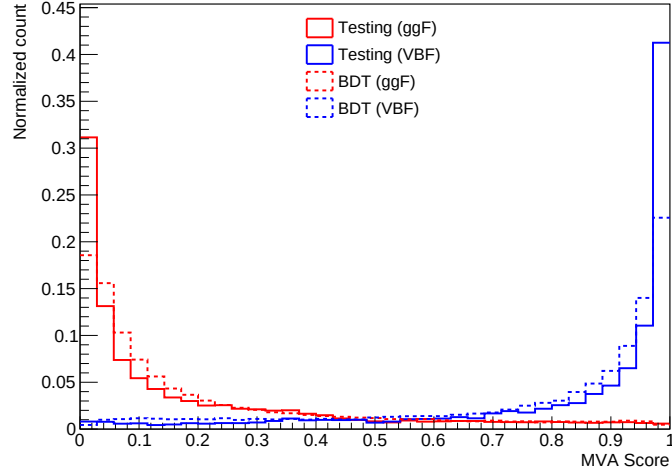


(b) VBF candidates

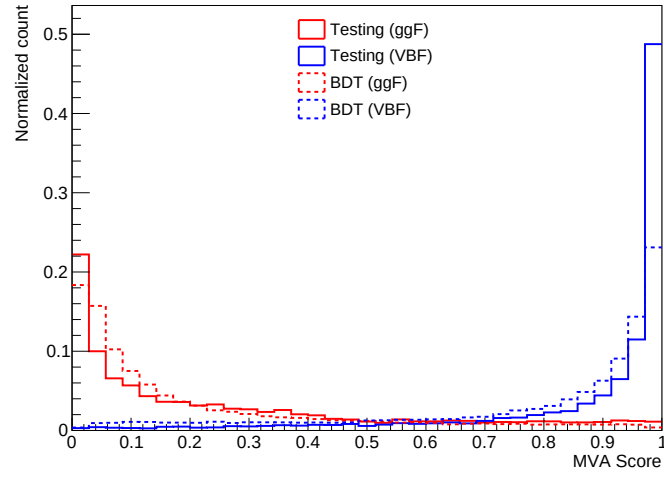


(c) Non-VBF candidates

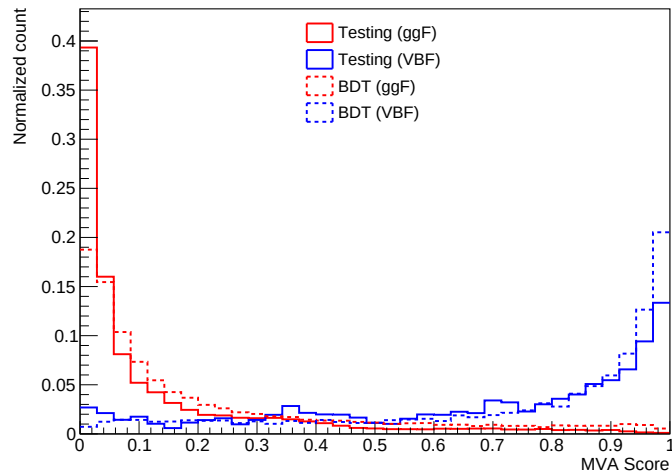
Figure 6.9: ROC curves of the transformer-based event classifier for the training, testing and validation set.



(a) Inclusive

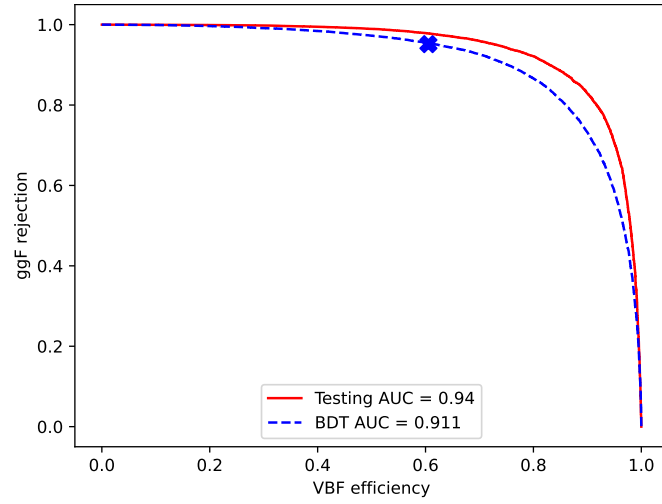


(b) VBF candidates

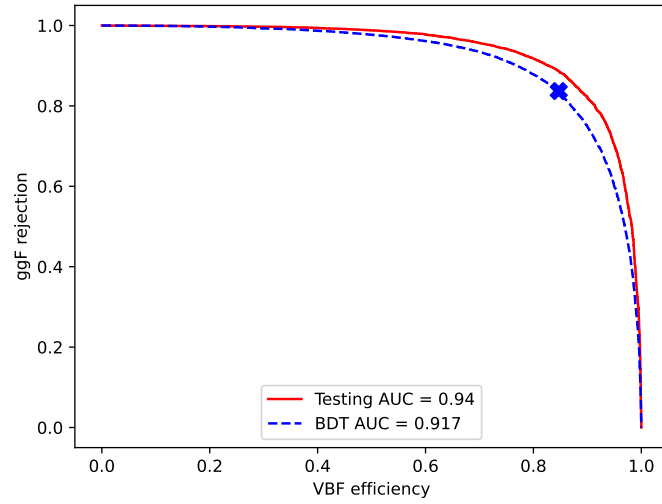


(c) Non-VBF candidates

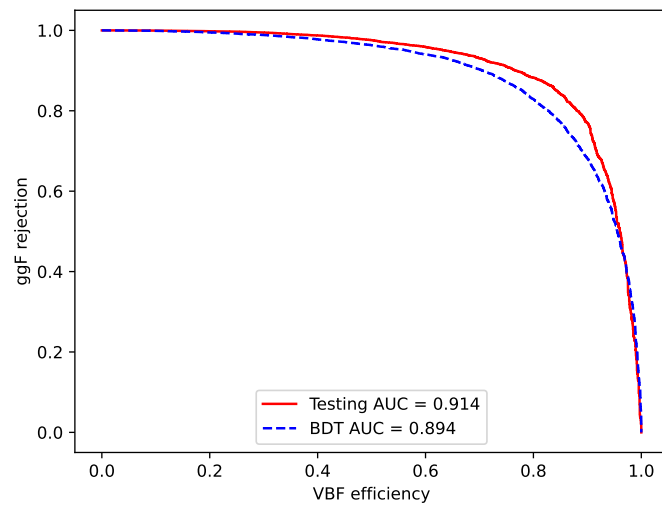
Figure 6.10: Distribution of output scores of the transformer-based event classifier (testing set) and of the Run 2 BDT.



(a) Inclusive

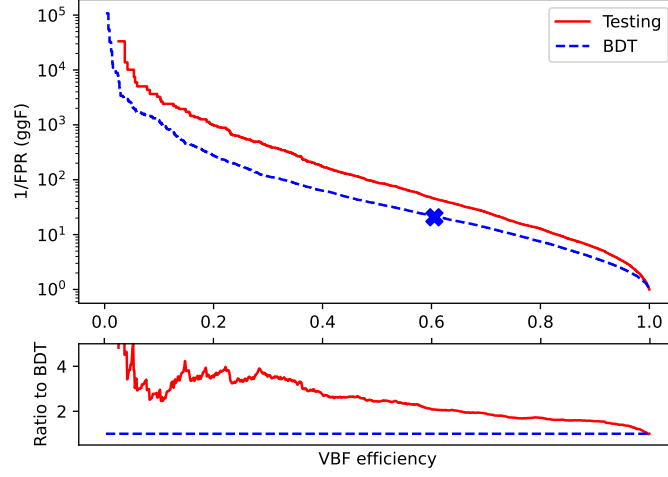


(b) VBF candidates

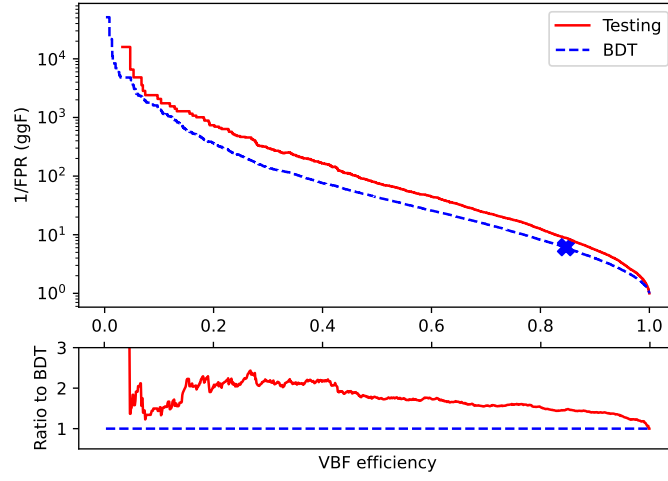


(c) Non-VBF candidates

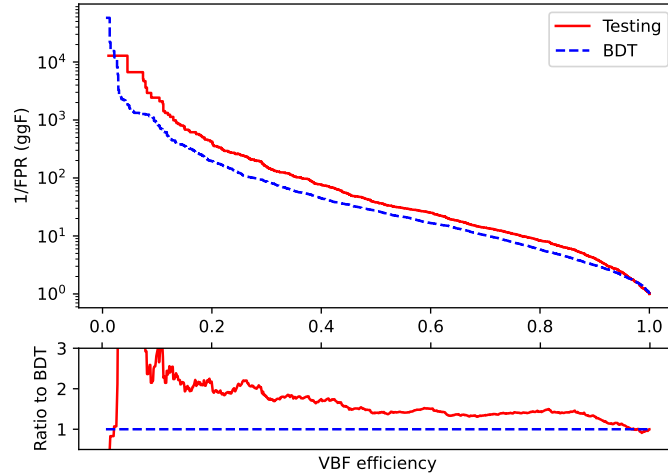
Figure 6.11: ROC curves of the transformer-based event classifier (testing set) and of the Run 2 BDT. The ggF rejection (vertical axis) is expressed as 1-FPR. The blue crosses correspond to the Run 2 analysis working point.



(a) Inclusive



(b) VBF candidates



(c) Non-VBF candidates

Figure 6.12: ROC curves of the transformer-based event classifier (testing set) and of the Run 2 BDT. The ggF rejection (vertical axis) is expressed as $1/\text{FPR}$, in log-scale. The ratio between the two is displayed in the subplot. The blue crosses correspond to the Run 2 analysis working point.

Conclusion and Outlook

The third operational run of the LHC is expected to (at least) double the integrated luminosity achieved during Run 2, which reached 140 fb^{-1} . The large luminosity, together with an increased centre-of-mass energy of 13.6 TeV, will allow ATLAS to significantly improve the precision of its searches for HH production and look for possible deviations from the SM predictions. This opportunity can be leveraged provided that the analysis strategy is further refined. Among the crucial tasks there are enhancing the trigger selection and optimising the offline event categorisation. This thesis specifically focused on tackling these two aspects for the $HH \rightarrow bb\tau_{\text{had}}\tau_{\text{had}}$ channel.

The first study involved the validation and optimisation of novel triggers to be employed by the $HH \rightarrow bb\tau\tau$ analysis during the Run 3 data-taking. A python framework was developed to calculate the efficiency gains of possible designs under different offline selections. In support of these studies, the kinematics of HH production has been investigated. The research concerned (i) validating new di- τ triggers with lower transverse momentum thresholds making use of ATLAS Delayed stream, (ii) the usage of the $HH \rightarrow 4b$ Delayed stream trigger for the $HH \rightarrow bb\tau\tau$ signal, and (iii) exploring possible $b + \tau$ triggers which take advantage of the event kinematics. All triggers have been compared to the baseline triggers used for the Run 2 analysis, taken as a benchmark. The research demonstrated a substantial improvement in total trigger efficiency, from 55% in Run 2 to 85% in Run 3 for the truth-matched selection, and exceeding 90% for the identified selection.

The second research dealt with employing a transformer, a state-of-the-art machine learning algorithm, to improve the event categorisation of the analysis. A new model based on the flavour tagging transformer GN2 was developed in the Salt framework, and adapted to handle event classification via a preprocessing toolkit developed in this thesis. The model was optimised and trained to accurately separate the VBF and ggF production modes. The classifier of Run 2, a BDT, was taken as the benchmark for the algorithm performance. This study achieved a significant enhancement compared to the BDT: considering the Run 2 analysis working point, the new model reached a 15% increase of VBF efficiency for the same ggF contamination, and a 100% increase in ggF rejection for the same VBF efficiency.

Now that the two researches of this thesis have come to a conclusion, there are several promising directions in which they can be expanded upon.

On the novel triggers side, at least three developments are poised to unfold. Firstly, the design of the $b + \tau$ chain is undergoing the final stages of refinement,

with a focus on perfecting the balance between rate, CPU costs and HLT thresholds. Furthermore, now that it is established that the novel triggers are able to increase the signal efficiency, the next step is to include them in the offline analysis and investigate in which signal region do the additional selected events end up: for example, under which score do they fall in the signal/background classification. In this thesis, the selected events have been visually studied as a function of m_{HH} ; future investigations will plot them as a function of the classifier score, to quantify how much actual sensitivity gain they will bring to the analysis.

Finally, after all the triggers are operational and included in streams, the next step will be to calibrate them, comparing the actual data they select with the Monte Carlo simulation they have been studied on up to now.

On the transformer side, the model perfected in this thesis has two major directions for further development.

Firstly, now that the region categorisation is in place, the following stage is to expand the event classification between the HH signal and its background processes. The transformer has the potential to outperform the signal/background classifier of the Run 2 analysis.

Furthermore, auxiliary tasks could be incorporated in the model. Auxiliary tasks exploit the different ways in which GNNs can be classified, via either the full graph, individual edges or individual nodes, to simultaneously perform multiple independent classifications on separate aspects of the input. To expand on the research of this thesis, these tasks will have to be adapted for event-related classifications. In the model of this thesis the full graph is used to perform the classification of the event; but the edges connecting any two nodes (two physical objects) could also be classified. An example of event-related edge classification could be object-pair origin compatibility, in analogy to the track-pair vertex compatibility of GN2. This could be particularly helpful when the model will handle the background processes, for instance in associating a b -jet and a τ -lepton that come from the same top quark, aiding the classification of the event as being not an HH signal.

Bibliography

- [1] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012). DOI: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020). arXiv: [1207.7214](https://arxiv.org/abs/1207.7214) [hep-ex].
- [2] CMS collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021). arXiv: [1207.7235](https://arxiv.org/abs/1207.7235). URL: <https://cds.cern.ch/record/1471016>.
- [3] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: [10.1103/PhysRevLett.13.585](https://doi.org/10.1103/PhysRevLett.13.585). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585>.
- [4] P.W. Higgs. “Broken symmetries, massless particles and gauge fields”. In: *Physics Letters* 12.2 (1964), pp. 132–133. ISSN: 0031-9163. DOI: [https://doi.org/10.1016/0031-9163\(64\)91136-9](https://doi.org/10.1016/0031-9163(64)91136-9). URL: <https://www.sciencedirect.com/science/article/pii/0031916364911369>.
- [5] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [6] B. Moser. Private communication.
- [7] ATLAS Collaboration. “Search for the non-resonant production of Higgs boson pairs via gluon fusion and vector-boson fusion in the $b\bar{b}\tau^+\tau^-$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. ATLAS-CONF-2023-071. Geneva, 2023. URL: <https://cds.cern.ch/record/2882132>.
- [8] M. D. Schwartz. “Quantum Field Theory and the Standard Model”. Cambridge University Press, Mar. 2014. ISBN: 978-1-107-03473-0, 978-1-107-03473-0.
- [9] M. Thomson. “Modern Particle Physics”. Cambridge University Press, 2013.
- [10] Particle Data Group. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097).
- [11] Particle Data Group. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001). URL: <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.

- [12] P. B. Palash. “An Introductory Course of Particle Physics”. Taylor & Francis, 2014. ISBN: 9781482216981. URL: <https://books.google.it/books?id=2Tf6AwAAQBAJ>.
- [13] W. Herr and B. Muratori. “Concept of luminosity”. In: (2006). DOI: [10.5170/CERN-2006-002.361](https://cds.cern.ch/record/941318). URL: <https://cds.cern.ch/record/941318>.
- [14] SLAC National Accelerator Laboratory. “Simulations”. URL: <https://theory.slac.stanford.edu/our-research/simulations>.
- [15] J. Baglio et al. “The measurement of the Higgs self-coupling at the LHC: theoretical status”. In: *JHEP* 04 (2013), p. 151. DOI: [10.1007/JHEP04\(2013\)151](https://doi.org/10.1007/JHEP04(2013)151). arXiv: [1212.5581](https://arxiv.org/abs/1212.5581) [hep-ph].
- [16] M. Carena et al. “Status of Higgs boson physics”. In: *Review of Particle Physics* (2016). URL: <http://pdg.lbl.gov/2016/reviews/rpp2016-rev-higgs-boson.pdf>.
- [17] ATLAS Collaboration. “Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector”. ATLAS-CONF-2021-030. Geneva, 2021. URL: <https://cds.cern.ch/record/2777236>.
- [18] LHC Higgs Cross Section Working Group. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. CERN Yellow Reports: Monographs. Geneva: CERN, 2017. DOI: [10.23731/CYRM-2017-002](https://cds.cern.ch/record/2227475). URL: <https://cds.cern.ch/record/2227475>.
- [19] LHC Higgs Cross Section Working Group. “LHC Cross Section Working Group Twiki”. 2021. URL: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGHH>.
- [20] L. Evans and P. Bryant. “LHC Machine”. In: *JINST* 3 (2008), S08001. DOI: [10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
- [21] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08003. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003). URL: <https://cds.cern.ch/record/1129811>.
- [22] E. Lopienska. “The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022”. In: (2022). General Photo. URL: <https://cds.cern.ch/record/2800984>.
- [23] LHCb Collaboration. “The LHCb Detector at the LHC”. In: *JINST* 3 (2008), S08005. DOI: [10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- [24] ALICE Collaboration. “The ALICE experiment at the CERN LHC”. In: *JINST* 3 (2008), S08002. DOI: [10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- [25] CMS Collaboration. “The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment”. In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004). URL: <https://cds.cern.ch/record/1129810>.
- [26] S. Fartoukh et al. “LHC Configuration and Operational Scenario for Run 3”. CERN-ACC-2021-0007. Geneva, 2021. URL: <https://cds.cern.ch/record/2790409>.

- [27] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3”. 2023. arXiv: [2305.16623 \[physics.ins-det\]](#).
- [28] ATLAS Collaboration. “The ATLAS Trigger System for LHC Run 3 and Trigger performance in 2022”. CERN-EP-2023-299. Geneva, 2024. arXiv: [2401.06630](#). URL: <https://cds.cern.ch/record/2886407>.
- [29] ATLAS Collaboration. “ATLAS Insertable B-Layer Technical Design Report”. CERN-LHCC-2010-013, ATLAS-TDR-19. 2010. URL: <https://cds.cern.ch/record/1291633>.
- [30] ATLAS Collaboration. “Production and integration of the ATLAS Insertable B-Layer”. In: *Journal of Instrumentation* 13.05 (May 2018), T05008–T05008. ISSN: 1748-0221. DOI: [10.1088/1748-0221/13/05/t05008](#). URL: <http://dx.doi.org/10.1088/1748-0221/13/05/T05008>.
- [31] ATLAS Collaboration. “ATLAS liquid-argon calorimeter: Technical Design Report”. Technical design report. ATLAS. Geneva: CERN, 1996. DOI: [10.17181/CERN.FRW.F00Q](#). URL: <https://cds.cern.ch/record/331061>.
- [32] ATLAS Collaboration. “ATLAS tile calorimeter: Technical Design Report”. Technical design report. ATLAS. Geneva: CERN, 1996. DOI: [10.17181/CERN.JRBJ.7028](#). URL: <https://cds.cern.ch/record/331062>.
- [33] ATLAS Collaboration. “New Small Wheel Technical Design Report”. CERN-LHCC-2013-006, ATLAS-TDR-020. ATLAS New Small Wheel Technical Design Report. 2013. URL: <https://cds.cern.ch/record/1552862>.
- [34] ATLAS Collaboration. “ATLAS Computing: technical design report”. Technical design report. ATLAS. Geneva: CERN, 2005. URL: <https://cds.cern.ch/record/837738>.
- [35] R. Frühwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4). URL: <https://www.sciencedirect.com/science/article/pii/0168900287908874>.
- [36] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. In: *The European Physical Journal C* 77.7 (July 2017). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-017-5004-5](#). URL: <http://dx.doi.org/10.1140/epjc/s10052-017-5004-5>.
- [37] ATLAS Collaboration. “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 79.8 (Aug. 2019). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-019-7140-6](#). URL: <http://dx.doi.org/10.1140/epjc/s10052-019-7140-6>.
- [38] ATLAS Collaboration. “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 76.5 (May 2016). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-016-4120-y](#). URL: <http://dx.doi.org/10.1140/epjc/s10052-016-4120-y>.

- [39] J. Pequeno and P. Schaffner. “How ATLAS detects particles: diagram of particle paths in the detector”. 2013. URL: <https://cds.cern.ch/record/1505342>.
- [40] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector.” In: *Eur. Phys. J. C* 77.7 (2017), p. 466. DOI: [10.1140/epjc/s10052-017-5031-2](https://doi.org/10.1140/epjc/s10052-017-5031-2). arXiv: [1703.10485](https://arxiv.org/abs/1703.10485). URL: <https://cds.cern.ch/record/2257597>.
- [41] M. Cacciari, G. P. Salam, and G. Soyez. “The anti-kt jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). URL: <http://dx.doi.org/10.1088/1126-6708/2008/04/063>.
- [42] ATLAS Collaboration. “Tagging and suppression of pileup jets with the ATLAS detector”. ATLAS-CONF-2014-018. Geneva, 2014. URL: <https://cds.cern.ch/record/1700870>.
- [43] ATLAS Collaboration. “Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run”. ATL-PHYS-PUB-2017-013. Geneva, 2017. URL: <https://cds.cern.ch/record/2273281>.
- [44] ATLAS Collaboration. “Topological b -hadron decay reconstruction and identification of b -jets with the JetFitter package in the ATLAS experiment at the LHC”. ATL-PHYS-PUB-2018-025. Geneva, 2018. URL: <https://cds.cern.ch/record/2645405>.
- [45] ATLAS Collaboration. “Secondary vertex finding for jet flavour identification with the ATLAS detector”. ATL-PHYS-PUB-2017-011. Geneva, 2017. URL: <https://cds.cern.ch/record/2270366>.
- [46] ATLAS Collaboration. “Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment”. ATL-PHYS-PUB-2017-003. Geneva, 2017. URL: <https://cds.cern.ch/record/2255226>.
- [47] ATLAS Collaboration. “Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS”. ATL-PHYS-PUB-2020-014. Geneva, 2020. URL: <https://cds.cern.ch/record/2718948>.
- [48] M. Zaheer et al. “Deep Sets”. 2018. arXiv: [1703.06114](https://arxiv.org/abs/1703.06114) [cs.LG].
- [49] ATLAS Collaboration. “Graph Neural Network Jet Flavour Tagging with the ATLAS Detector”. ATL-PHYS-PUB-2022-027. Geneva, 2022. URL: <https://cds.cern.ch/record/2811135>.
- [50] A. Vaswani et al. “Attention Is All You Need”. 2023. arXiv: [1706.03762](https://arxiv.org/abs/1706.03762) [cs.CL].
- [51] ATLAS Collaboration. “Performance of the ATLAS Secondary Vertex b -tagging Algorithm in 7 TeV Collision Data”. ATLAS-CONF-2010-042. Geneva, 2010. URL: <https://cds.cern.ch/record/1277682>.
- [52] ATLAS Collaboration. “Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC”. ATL-PHYS-PUB-2015-045. Geneva, 2015. URL: <https://cds.cern.ch/record/2064383>.

- [53] ATLAS Collaboration. “Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment”. ATL-PHYS-PUB-2019-033. Geneva, 2019. URL: <https://cds.cern.ch/record/2688062>.
- [54] A. Elagin et al. “A new mass reconstruction technique for resonances decaying to $\tau\tau$ ”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654.1 (Oct. 2011), pp. 481–489. ISSN: 0168-9002. DOI: [10.1016/j.nima.2011.07.009](https://doi.org/10.1016/j.nima.2011.07.009). URL: <http://dx.doi.org/10.1016/j.nima.2011.07.009>.
- [55] I. Neutelings. “SM Decay Pie Chart”. 2022. URL: https://github.com/IzaakWN/CodeSnippets/blob/master/LaTeX/TikZ/physics/SM_decay_piechart.pdf.
- [56] G. Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0). arXiv: [1007.1727](https://arxiv.org/abs/1007.1727) [physics.data-an].
- [57] ATLAS Collaboration. “Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data”. ATL-PHYS-PUB-2022-044. Geneva, 2022. URL: <https://cds.cern.ch/record/2827111>.
- [58] ATLAS Collaboration. “ATLAS Public Results: Tau Trigger”. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TauTriggerPublicResults>.
- [59] S. Elfving, E. Uchibe, and K. Doya. “Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning”. 2017. arXiv: [1702.03118](https://arxiv.org/abs/1702.03118) [cs.LG].
- [60] B. Sanchez-Lengeling et al. “A Gentle Introduction to Graph Neural Networks”. In: *Distill* (2021). <https://distill.pub/2021/gnn-intro>. DOI: [10.23915/distill.00033](https://doi.org/10.23915/distill.00033).
- [61] P. Veličković et al. “Graph Attention Networks”. 2018. arXiv: [1710.10903](https://arxiv.org/abs/1710.10903) [stat.ML].
- [62] F. Scarselli et al. “The Graph Neural Network Model”. In: *IEEE Transactions on Neural Networks* 20.1 (2009), pp. 61–80. DOI: [10.1109/TNN.2008.2005605](https://doi.org/10.1109/TNN.2008.2005605).
- [63] S. Brody, U. Alon, and E. Yahav. “How Attentive are Graph Attention Networks?” 2022. arXiv: [2105.14491](https://arxiv.org/abs/2105.14491) [cs.LG].
- [64] C. Joshi. “Transformers are Graph Neural Networks”. NTU Graph Deep Learning Lab. 2020.
- [65] A. Duperrin. “Flavour tagging with graph neural networks with the ATLAS detector”. 2023. arXiv: [2306.04415](https://arxiv.org/abs/2306.04415) [hep-ex].
- [66] ATLAS Flavor Tagging Group. “SALT”. 2023. URL: <https://gitlab.cern.ch/atlas-flavor-tagging-tools/algorithms/salt>.
- [67] The HDF Group. “Hierarchical Data Format, version 5”. URL: <https://github.com/HDFGroup/hdf5>.

- [68] R. Brun and F. Rademakers. “ROOT — An object oriented data analysis framework”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(97\)00048-X](https://doi.org/10.1016/S0168-9002(97)00048-X). URL: <https://www.sciencedirect.com/science/article/pii/S016890029700048X>.
- [69] ATLAS Flavour Tagging Group. “Salt - GN2 Configuration File”. <https://gitlab.cern.ch/atlas-flavor-tagging-tools/algorithms/salt/-/blob/main/salt/configs/GN2.yaml>. [2024].
- [70] L. N. Smith. “A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay”. 2018. arXiv: [1803.09820](https://arxiv.org/abs/1803.09820) [cs.LG].
- [71] ATLAS Flavour Tagging Group. “Salt - Event Classifier Configuration File”. <https://gitlab.cern.ch/atlas-flavor-tagging-tools/algorithms/salt/-/blob/main/salt/configs/GN2.yaml>. [2024].

Appendix (A)

Additional Studies on Novel Triggers for $HH \rightarrow bb\tau\tau$

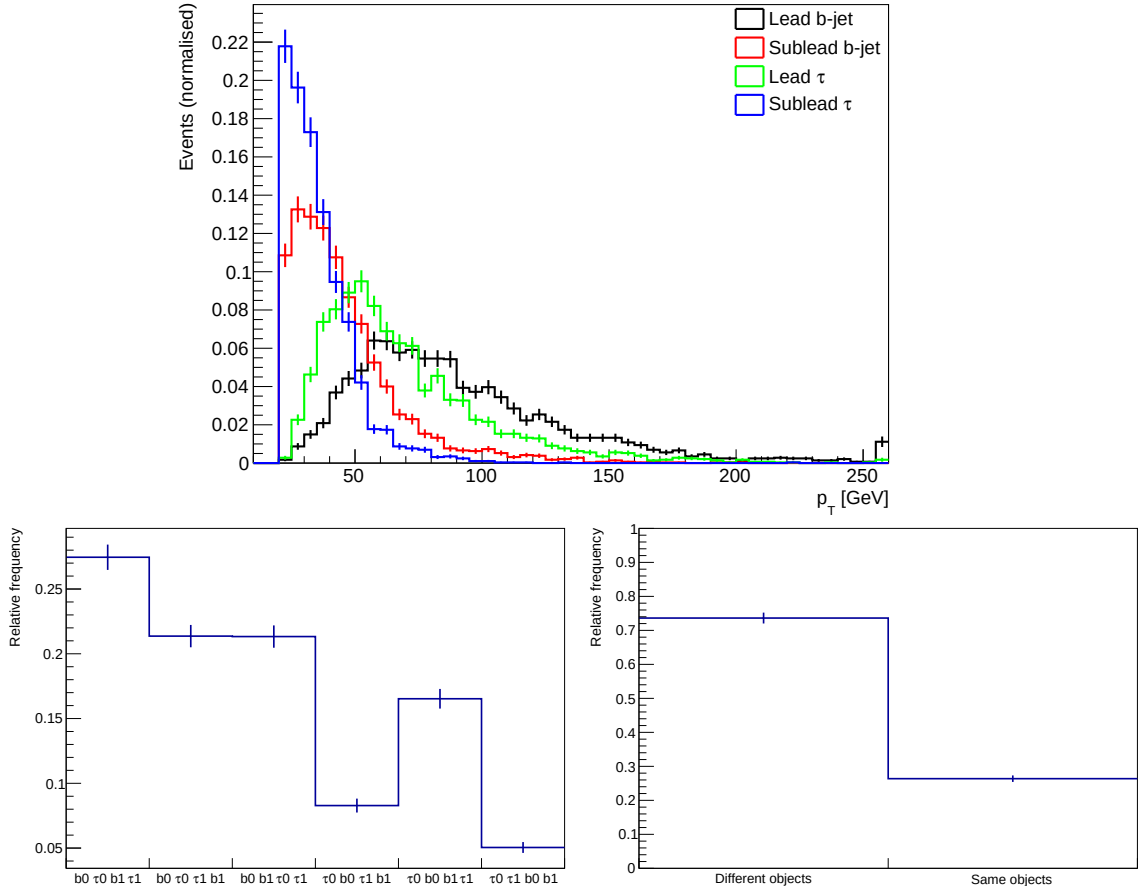


Figure A.1: Event kinematics of the $bb\tau\tau$ final state and frequency of the order of the objects p_T for $\kappa_\lambda = 10$.

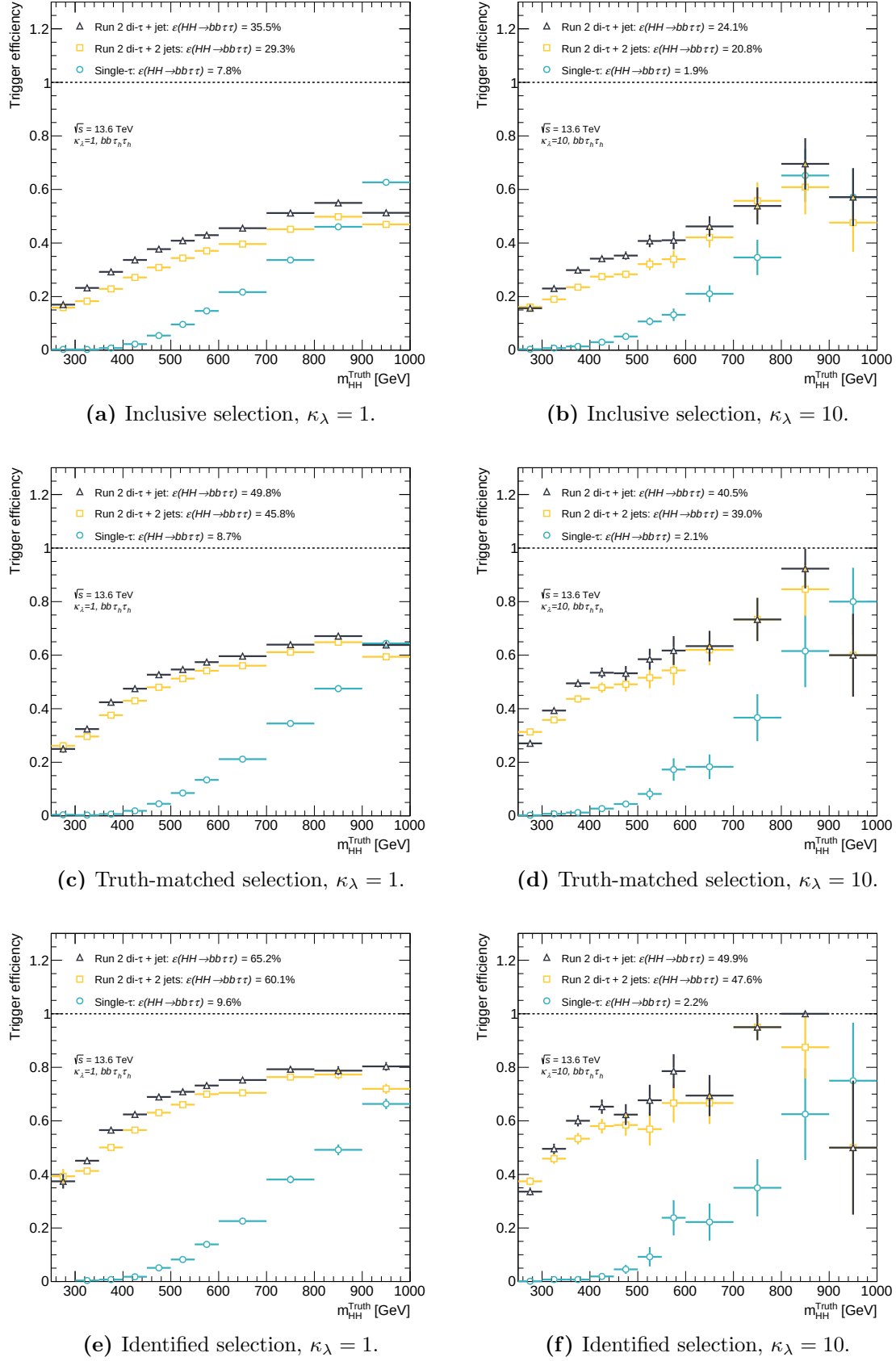


Figure A.2: The efficiency of the individual Run 2 τ triggers plotted as a function of m_{HH}^{Truth} for the three offline selections and both $\kappa_\lambda = 1$ and $\kappa_\lambda = 10$.

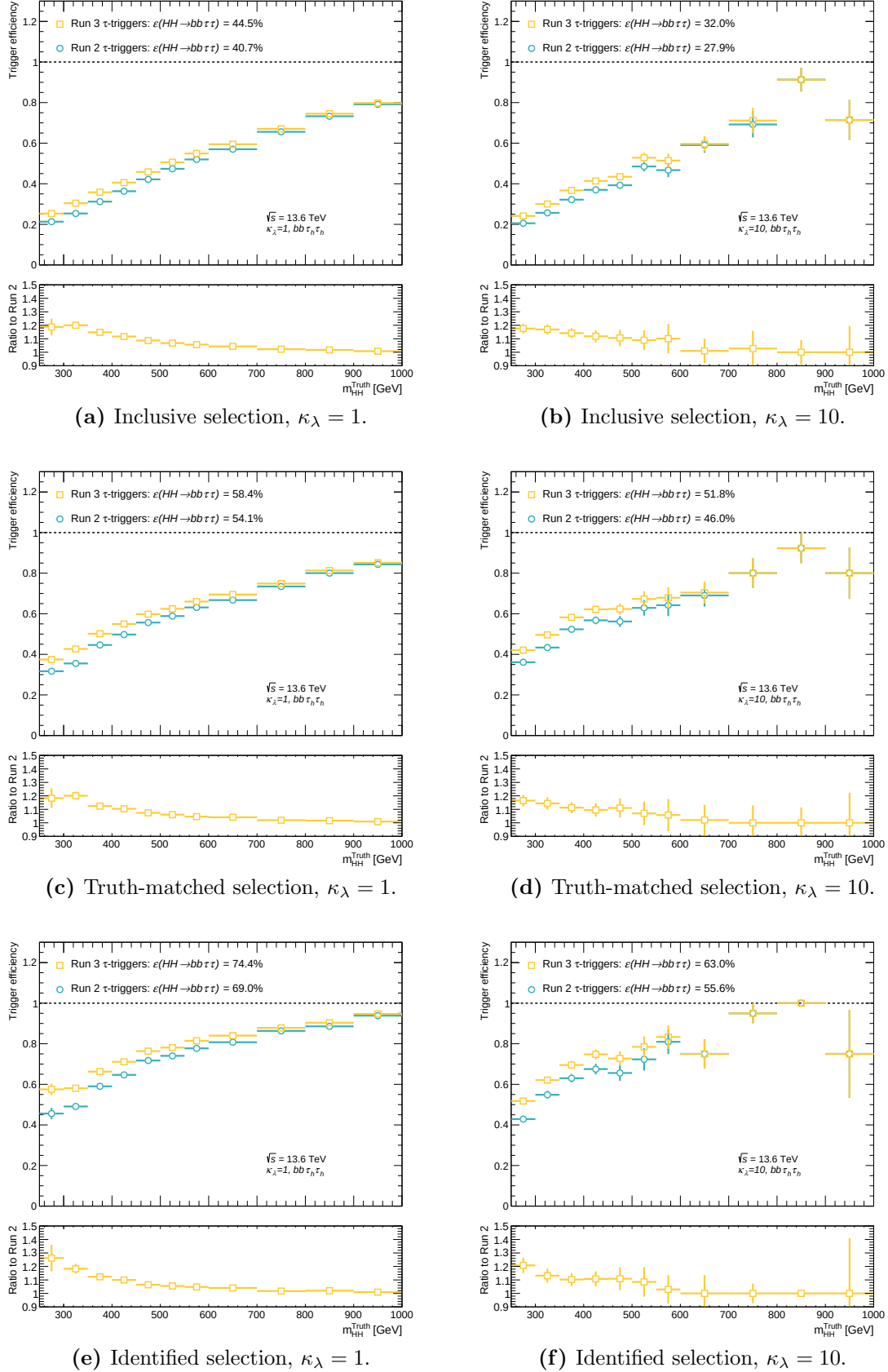


Figure A.3: The efficiency of the Run 3 τ triggers plotted as a function of m_{HH}^{Truth} for the three offline selections and both $\kappa_\lambda = 1$ and $\kappa_\lambda = 10$. The Run 2 τ triggers serve as the benchmark, with the subplots illustrating the relative efficiency gains with respect to them.

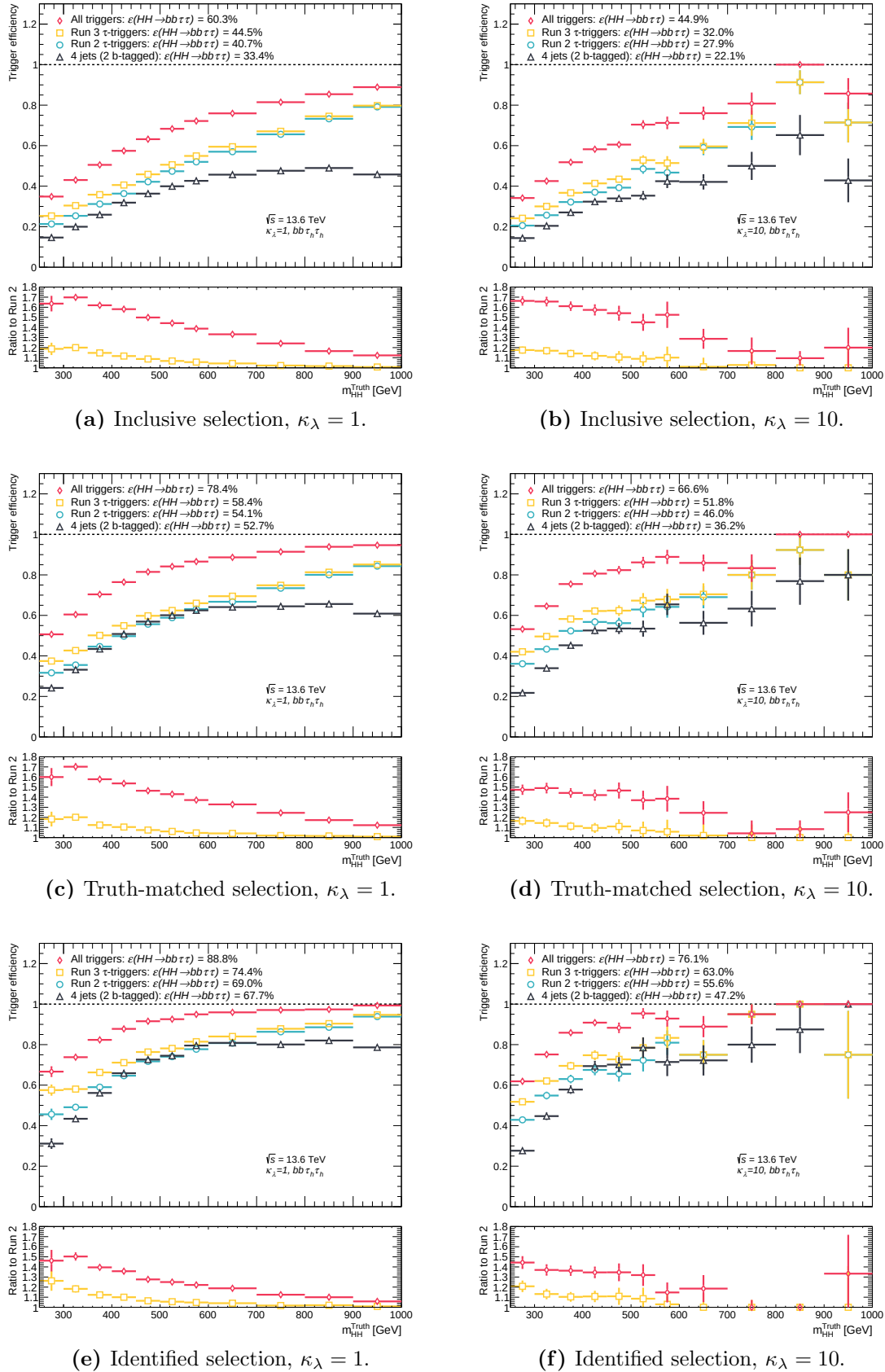


Figure A.4: The efficiency of the $HH \rightarrow 4b$ Delayed stream trigger (indicated as $4 jets$), the Run 3 τ triggers and their combination plotted as a function of m_{HH}^{Truth} for the three offline selections and both $\kappa_\lambda = 1$ and $\kappa_\lambda = 10$. The Run 2 τ triggers serve as the benchmark, with the subplots illustrating the relative efficiency gains with respect to them.

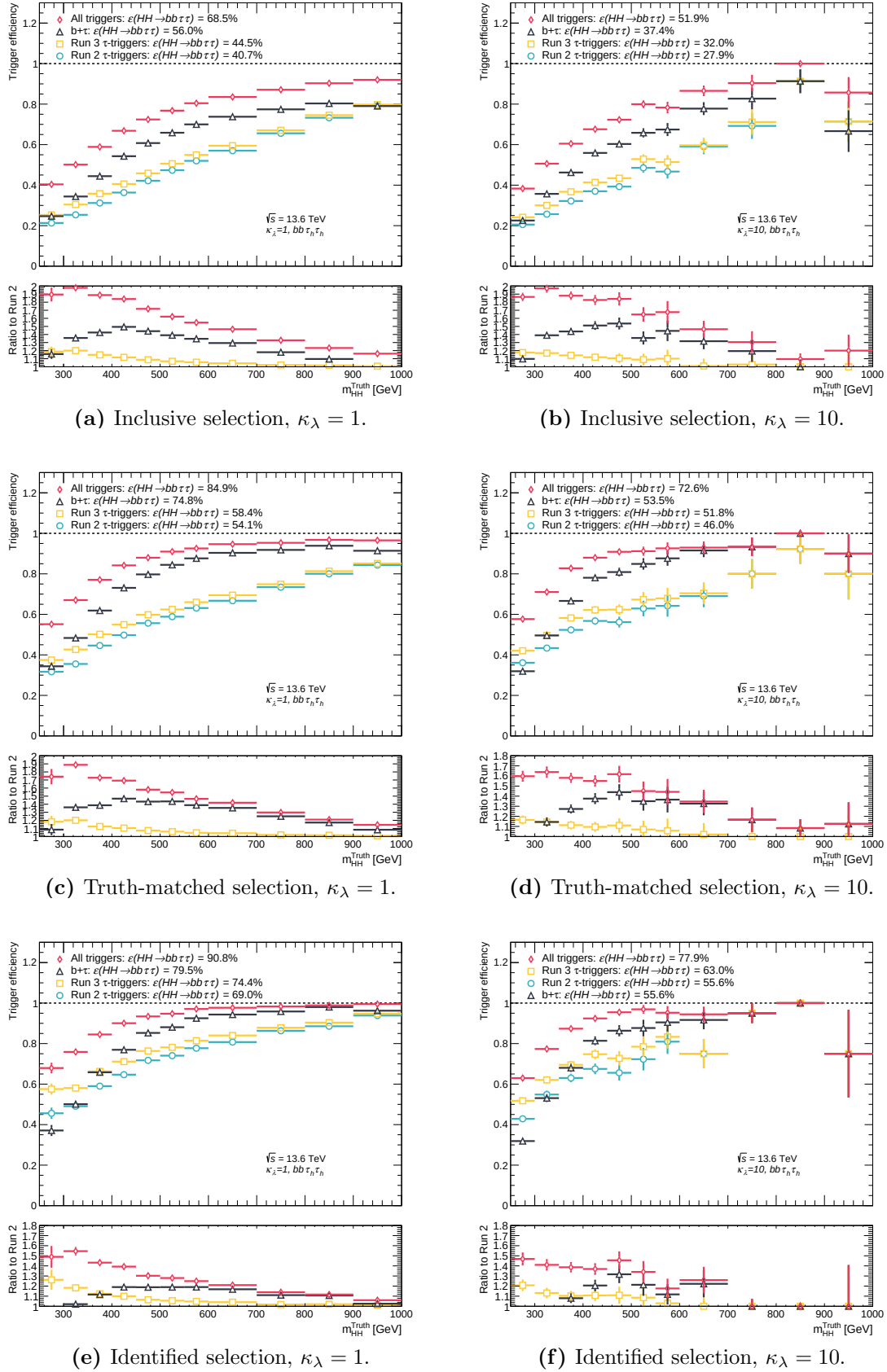


Figure A.5: The efficiency of the $b + \tau$ trigger, the Run 3 τ triggers and their combination plotted as a function of m_{HH}^{Truth} for the three offline selections and both $\kappa_\lambda = 1$ and $\kappa_\lambda = 10$. The Run 2 τ triggers serve as the benchmark, with the subplots illustrating the relative efficiency gains with respect to them.

Acknowledgements

It is always a gamble whether words can truly do justice to meaning: in this case, the *meaning* is the extent of how I have felt in these months at CERN with Dr. Brian Moser and Prof. Valerio Dao. So I will go straight to the point. My bosses have been the best I could have asked for. The synergy I felt with them not only made me absorb the constant learning environment they gifted me, but also made my experience unforgettable. Every morning I woke up with a sense of delight for the work that awaited me with them. Most importantly, the idea of making them proud acted as a fuel for me. On one hand I realised that I was perceiving Valerio as being more than a boss; he has been a *mentor*, in the most noble sense of the term. At the same time, Brian has always been, and continues to be, a constant source of inspiration. These months with them have been enriching on both the research experience and the human side. I hope my contributions matched the bond I felt towards both of you.

A big thank you to my supervisor Prof. Maximiliano Sioli and my Alma Mater, Università di Bologna, for having allowed me to embark in this experience and having supported me.

About the research, I thank the whole trigger meeting for the nice atmosphere and for all the opportunities they provided me, including allowing me to present at the ATLAS Week poster session and to publish plots with ATLAS. A heartfelt thank you in particular to Prof. Chris Pollard and Prof. Katharine Leney. On the transformer side, I am grateful to Dr. Sam Van Stroud for his kind availability, which was as precious as his research; I gained valuable knowledge by exploring his work.