

School of Physics and Astronomy **בית הספר לפיזיקה ולאסטרונומיה**
The Raymond and Beverly Sackler **הפקולטה למדעים מדויקים**
Faculty of Exact Sciences **ע"ש ריימונד ובברלי סאקלר**
Tel Aviv University **אוניברסיטת תל אביב**

Boosted Decision Trees for the ATLAS Level-1 Calorimeter Tau Trigger

Uriel Barron

Research thesis for the degree of Master of Science

Carried out under the supervision of:

Dr. Liron Barak

July 2022



Abstract

The Large Hadron Collider at CERN generates 10^{34} proton-proton collisions per square centimeter per second in the ATLAS detector. Unfortunately, it is not feasible to save to disk such enormously large volumes of data - necessitating the concept of the trigger. The ATLAS Trigger System decides which collision events contain interesting physics and which do not. The interesting events are saved to disk and the rest are discarded. This thesis specifically discusses the Tau Trigger System which attempts to detect in real-time if a given collision contains a signal compatible with a tau lepton. Until the writing of this thesis, a simple energy cutoff algorithm was used which was moderately successful at classifying true tau events. Since the Trigger System suffers from harsh latency constraints as well as being based on hardware with limited resources, it was challenging to improve upon these results. This thesis proposes a novel, hardware specific, machine learning based approach which improves the performance of the ATLAS Tau Trigger. The use of this new method will result in the recording of more tau lepton collision data which is critical to many ongoing analyses.

CONTENTS

1	Introduction	5
2	Theoretical Background	7
2.1	The Standard Model	7
2.2	Tau Leptons	9
2.3	Taus in Analyses	10
3	The LHC and the ATLAS Experiment	11
3.1	The Large Hadron Collider	11
3.2	The ATLAS Detector	13
3.3	ATLAS Coordinate System	13
3.4	The Calorimeters	16
3.5	ATLAS Trigger System	18
4	The Tau Trigger	19
4.1	Introduction	19
4.2	LAr Phase-I Upgrade	20
4.3	Seeding and Clustering	20
4.4	Trigger Object (TOB)	22
4.5	Isolation	22

4.6	Hardware Requirements	23
4.7	Background Rate Requirement	24
4.8	The Need for a New Tau Trigger Algorithm	25
5	Gradient Boosted Decision Trees (GBDT)	26
5.1	Hardware Constraints	26
5.2	GBDT	28
5.3	Hyperparameters	29
6	GBDT Based Tau Trigger	31
6.1	Training Data	31
6.2	High Level Features	33
6.3	Training and Evaluation	41
6.4	Turn-on Curve	42
6.5	Model Optimization	44
7	GBDT on FPGAs	46
7.1	Conifer	46
7.2	FPGA Resources and Performance	47
7.3	Optimization for FPGA	47
8	Summary and Outlook	50

LIST OF FIGURES

2.1	The Standard Model of elementary particles with their respective mass, electric and color charges, and spin [1].	9
2.2	Feynman diagram of a tau particle decay by emission of an off-shell W boson.	10
3.1	The CERN accelerator complex [2]	12
3.2	The LHC baseline plan for the next decade and beyond, showing the energy of the collisions and luminosity [3].	13
3.3	Cut-away view of the ATLAS detector.	14
3.4	The ATLAS Cartesian coordinates.	14
3.5	Pseudorapidity values shown on a polar plot.	15
3.6	The ATLAS Calorimeter System [4].	16
3.7	Schematic of the Liquid Argon Calorimeter [5].	17
3.8	Diagram of the ATLAS Data Acquisition System (DAQ) and Trigger System in Run 2 and Run 3. The reductions in data rates can be seen for each stage in the trigger [6].	18
4.1	An electron (with 70 GeV of transverse energy) as seen by the existing Level-1 Calorimeter trigger electronics (a) and by the proposed upgraded trigger electronics (b) [7].	21
4.2	The granularity increase. A single cell in Run 2 (the ‘Trigger tower’) became ten separate cells (‘Supercells’) in Run 3.	22

4.3	ATLAS Level-1 Calorimeter Trigger Phase-I Upgrade Electronics [8].	23
5.1	Schematic view of the trigger tower used as input to the L1Calo Run 2 tau trigger algorithm [9].	27
5.2	Example of a decision tree used to classify a TOB.	29
6.1	Feynman diagrams of the signal and background samples used for training the model.	32
6.2	Transverse momentum distribution of the visible component of the decay products from tau leptons originating from Z boson decays.	32
6.3	The energy deposits in each layer of the calorimeter of a randomly selected TOB. . .	33
6.4	It is required to travel six cells in L_1 distance to arrive from one maximum to the other. Thus, in this example maxdist=6.	36
6.5	Signal and background distributions for all 24 high level features.	38
6.5	Signal and background distributions for all 24 high level features.	39
6.5	Signal and background distributions for all 24 high level features.	40
6.6	Feature correlation heatmap.	41
6.7	ROC curve for the GBDT model with an area under curve of 0.963.	42
6.8	Turn on curve for the GBDT and Run 2 showing the signal efficiency for each energy bin. Both models have the the same FPR which is the Run 2 background rate R_{bg} . The GBDT achieves a higher efficiency than Run 2 across the entire range of energies.	44
6.9	Two dimensional histogram describing the distribution of TOBs with respective truth pT and reco pT. It can be seen that although the energies are correlated, the reconstructed energy is systematically overestimated.	45
7.1	AUC and LUT usage for different models showing the trade-off between better predictive ability and higher resource usage.	48
7.2	Best possible AUC given the available percentage of the FPGA's total LUT slices. The models improve as the resources available grow. Run 2 is given for reference. . .	49

CHAPTER

1

INTRODUCTION

The idea that matter is composed of fundamental indivisible components dates back to at least the 5th century BC when Democritus proposed his atomic theory. This theory was refined when in the 19th century, the atoms, forming the basic building blocks of each element, were discovered and subsequently catalogued in Mendeleev's periodic table. In the early 20th century, research in the fields of nuclear physics and quantum mechanics showed that these atoms were not in fact fundamental and that an improved theory was required. In the two decades following 1950, confirmed by the experimental discovery of quarks in the 1970s, the Standard Model of elementary particles was formulated. It is to date the most successful physical theory of elementary particles receiving further confirmation from the discoveries of the top quark (1995), the tau neutrino (2000), and the Higgs boson (2012).

The refinement of the Standard Model parameters to ever higher degrees of precision as well as searches for physics beyond the Standard Model requires vast experimental data. This data is mostly collected from particle accelerator experiments such as the Large Hadron Collider. The enormous effort put into the creation and improvement of the experimental apparatuses and the ensuing data analysis goes hand in hand with the advancement of our theoretical understanding. This thesis continues this Herculean effort by improving the ATLAS Level-1 Tau Trigger, the mechanism responsible for identifying and recording tau lepton data.

This improvement in tau identification is brought about by the integration of a machine learning model, namely gradient boosted decision trees. The steady increase in computing power in recent years has facilitated a boom in machine learning and enabled its application in circumstances where it was previously impossible. This had a profound effect on the field of high energy physics where these methods are utilized to benefit processes across the many stages of the experiments. In this work, these methods are applied to improve tau lepton identification. The specifics of the implementation on ATLAS hardware are detailed.

This work is organized as follows. Chapter 2 provides the underlying theoretical framework required as well as the motivation for this work. Chapter 3 describes in detail the Large Hadron Collider and the ATLAS experiment at CERN. Chapter 4 delves into the necessity of the tau trigger and its operation. Chapter 5 provides a rudimentary description of gradient boosted decision trees (GBDTs) and argues that they are apt candidates to replace the current tau trigger algorithm. Chapter 6 details the specifics of the GBDT based tau trigger and presents the improvement achieved with this method. Chapter 7 explains the process of implementing the new tau trigger algorithm on the existing hardware in ATLAS. Finally, chapter 8 summarizes the work and discusses possible avenues for future improvement.

CHAPTER

2

THEORETICAL BACKGROUND

2.1 The Standard Model

The Standard Model (SM) of particle physics is a theory, developed in the 1960's and 1970's, that describes the electromagnetic, weak, and strong interactions and classifies all known elementary particles [10][11][12][13]. The SM has been hugely successful, providing many verified experimental predictions covering a wide range of energies and processes, over the last 50 years. It includes two classes of elementary particles, denominated fermions and bosons.

The fermions constitute all everyday matter around us, have half-integer spin, and can be further separated into two more classes, the leptons and the quarks. Each of these come in three generations of pairs, i.e., there are six types of leptons and six types of quarks. The charged leptons consist of three particles: the electron (e), the muon (μ), and the tau (τ), with each one being more massive than the former. They all have identical electric charge of $-e$ - defined to be the charge of the electron. Each of these is accompanied by a corresponding neutrino, the electron-neutrino (ν_e), the muon-neutrino (ν_μ), and the tau-neutrino (ν_τ), which have no electric charge and according to the SM are all massless. The neutrinos can only interact via the weak interaction while the charged leptons can also interact electromagnetically.

The quarks come in three pairs, each consisting of an up-type quark with electric charge $+\frac{2}{3}e$

and a down-type quark with electric charge $-\frac{1}{3}e$. The first pair are called the up and down quarks, the second pair the strange and charm quarks, and the third the top and bottom quarks. The quarks are all massive with the masses increasing with each generation of pairs. Unlike leptons, the quarks also carry color charge and can thus interact not only via the electromagnetic and weak interactions but also through the strong interaction. The color charge comes in three different colors - red, green, and blue, with corresponding anti-colors. Colored objects exhibit a phenomenon called ‘confinement’ which forbids the existence of free particles with color. Thus every manifestation of quarks in nature must be of a composite structure with total color zero. These colorless particles are called hadrons. One of the most common of such manifestations is the meson, a quark and antiquark pair with opposite colors. Another kind of hadron is the baryon, three quarks, one red, one green, one blue, which together are colorless. While confinement is not mathematically proven, isolated colored particles have never been experimentally observed. Because the strong interaction exhibits asymptotic freedom, i.e., the energy of the interaction increases with distance, at a certain distance it becomes energetically favorable to create a new quark-antiquark pair rather than to separate further. This process is called hadronization and when repeated many times in a process creates a ‘jet’ - a collimated group of colorless particles. In summary, counting the leptons and the quarks we arrive at a total of twelve fermions. Since in addition each fermion has a corresponding antiparticle there are in total twenty four fermions in the Standard Model.

The bosons have integer spin and mediate the interactions between the SM particles. First, there is the photon, the mediator of the electromagnetic interaction. In addition, there are the neutral Z particle and two charged W particles - one with electric charge $+e$ and the other with $-e$ - that mediate the weak interaction. Finally, there are eight gluons that mediate the strong interaction. Though electrically neutral, the gluons carry a color charge. Thus, unlike the photon which only mediates the electromagnetic interaction the gluons both mediate the strong interaction and also participate in it allowing gluon-gluon interactions. The last boson in the SM is the Higgs boson, which through the Higgs mechanism allows some particles to acquire mass. The Higgs boson is massive and has no electric or color charge. All the elementary particles in the SM are summarized in Fig. 2.1.

Despite the unprecedented success of the SM, it faces several challenges. First of all, experiments have shown that at least some of the neutrinos are massive. This directly contradicts the SM which predicts that neutrinos must be massless particles. In addition, the Higgs mechanism gives rise to the hierarchy problem [14] if some new physics is present at high energy scales. These issues, the ad hoc nature of the 19 numerical constants required to describe the SM, as well as others pose challenges for the SM. Attempts to expand on or otherwise adjust the SM to address these issues are termed Beyond the Standard Model (BSM) physics.

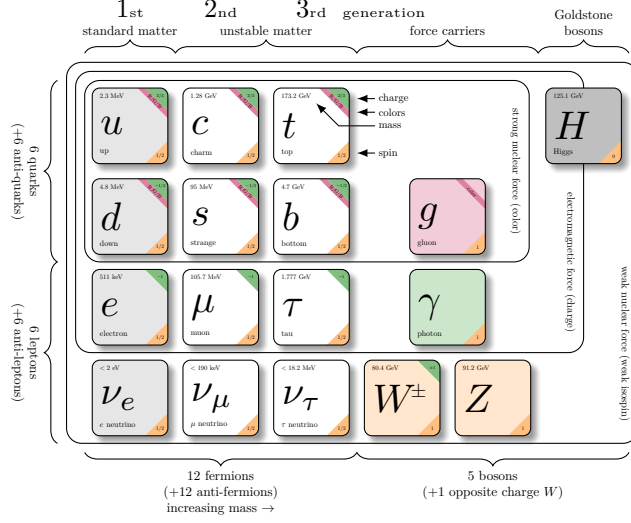


Figure 2.1: The Standard Model of elementary particles with their respective mass, electric and color charges, and spin [1].

2.2 Tau Leptons

The tau lepton, denoted τ , is the heaviest of the leptons with a mass of 1776.86 ± 0.12 MeV [15]. Similar to both the electron and the muon it has negative electric charge and spin half. It also has an antiparticle $\bar{\tau}$ with positive charge. It has a proper lifetime of 2.9×10^{-13} s and a proper decay length of $c\tau = 87.03 \mu\text{m}$ [15].

The tau is the only lepton that can decay into hadrons since both the muon and the electron do not have sufficient mass for such decays to be kinematically allowed. The tau can decay both leptonically ($\tau \rightarrow \ell \nu_\ell \nu_\tau$, $\ell = e, \mu$) and hadronically ($\tau \rightarrow \text{hadrons } \nu_\tau$). The decay is mediated through the weak interaction charged current and can be visualized by its Feynman diagram in Fig. 2.2.

Tau leptons that decay leptonically, i.e. into either an electron or a muon, have a branching fraction of 35%. The other 65% of decays occur hadronically with the most common hadronic decay products being pions. Decay modes with n charged particles in the final state are referred to as n -prong decays. Of the hadronic decays, 77% decay into the 1-prong channel and 22% into the 3-prong channel. The 5-prong channels and above occur very rarely. Table 2.1 lists the most common tau lepton decay modes amounting to about 90% of decays. The remaining 10% correspond to another 38 decay modes involving either kaons or rarer decays involving pions, omega, or eta mesons [1].

With such a short decay length the tau leptons typically cannot be detected directly and only their decay products can be observed. Since lepton number is conserved in weak decays one of the

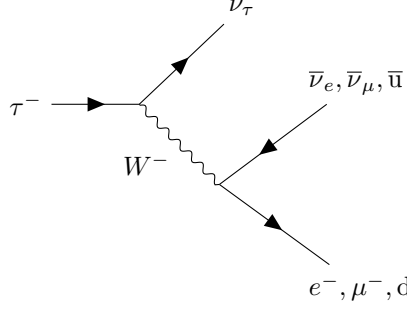


Figure 2.2: Feynman diagram of a tau particle decay by emission of an off-shell W boson.

decay mode	branching fraction (%)
$\mu^- \bar{\nu}_\mu \nu_\tau$	17.39 ± 0.04
$e^- \bar{\nu}_e \nu_\tau$	17.82 ± 0.04
$\pi^- \nu_\tau$	10.82 ± 0.05
$\pi^- \pi^0 \nu_\tau$	25.49 ± 0.09
$\pi^- \pi^0 \pi^0 \nu_\tau$	9.26 ± 0.10
$\pi^- \pi^0 \pi^0 \pi^0 \nu_\tau$	1.04 ± 0.07
$\pi^- \pi^- \pi^+ \nu_\tau$	8.99 ± 0.05
$\pi^- \pi^- \pi^+ \pi^0 \nu_\tau$	2.74 ± 0.07

Table 2.1: Most common tau branching fractions[1].

tau's decay products will always be a neutrino tau.

2.3 Taus in Analyses

Triggering efficiently on tau leptons is essential since they play a vital role in the ATLAS physics program. The unique position of the tau as the heaviest lepton ensures that it has a crucial role in Higgs phenomenology as it provides the strongest signal for measurements of the SM Higgs boson coupling to fermions [16]. While the top does have a stronger coupling to the Higgs boson, it is too massive to constitute a Higgs decay product. Moreover, since the tau's hadronic decay channel has the highest branching ratio, triggering on them significantly increases the available tau lepton data in ATLAS. In addition, in many BSM scenarios, final states containing tau leptons are also often favoured by heavier Higgs bosons or other new resonances [17, 18]. Most of these analyses rely on triggering exclusively on hadronically decaying tau leptons.

CHAPTER

3

THE LHC AND THE ATLAS EXPERIMENT

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest and highest-energy particle collider in the world. It was constructed by the European Organization for Nuclear Research (CERN) between 1998 and 2008 inside the tunnel originally used by LEP (Large Electron–Positron Collider), a previous collider built in 1989. It is a 26.7 km long circular accelerator located under the French-Swiss border near Geneva. The LHC primarily collides two proton beams leading to proton-proton (pp) collisions with energies of up to 13 TeV [19]. The data collected from the pp collisions, at energies never observed before in a laboratory, allows the confirmation or rejection of proposed physical theories.

The LHC is in fact the last stage of a series of accelerators belonging to the CERN accelerator complex [20]. The protons originate from a hydrogen source and are accelerated to 50 MeV by the linear accelerator LINAC2. They are then injected into the Proton Synchrotron Booster (PSB) and accelerated to 1.4 GeV. They are then further accelerated by the Proton Synchrotron (PS) to 25 GeV and by the Super Proton Synchrotron (SPS) to 450 GeV. After that they are fed into the LHC where they are split into two opposing proton beams, one going clockwise in the tunnel and

the other counter-clockwise. Each beam is then accelerated to a final 6.5 TeV such that at the center of mass there is a total collision energy of 13 TeV. In addition to the requirement for high energy collisions there is also a requirement for a high rate of collisions. The accumulation of vast amounts of collision data allows analyses to be performed at high levels of statistical precision as well as the creation of sufficient data to probe rare processes. The collision rate can be quantified by the instantaneous luminosity parameter L which is defined as the number of collision events per second. The LHC instantaneous luminosity is equal to $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. Integrating this value over the entire run-time of the detector until the year 2018 yields the integrated luminosity of 138 fb^{-1}

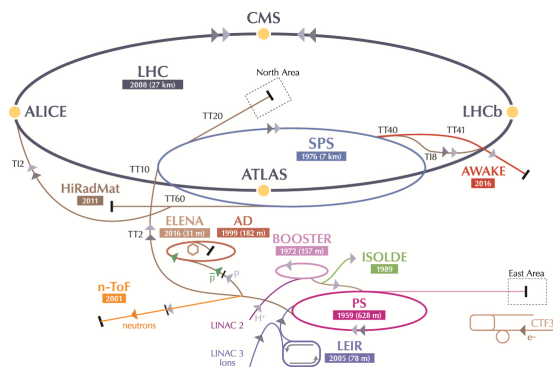


Figure 3.1: The CERN accelerator complex [2]

There are four beam collision points in the LHC corresponding to the four main LHC detectors ATLAS [21], CMS [22], LHCb [23] and ALICE [24]. Each of these detectors is placed at one of the collision points to record the high energy collisions at that location. ATLAS and CMS are general purpose detectors and cover a wide range of phenomena whereas LHCb focuses primarily on b-quark physics and CP-violation and ALICE on heavy ion collisions.

The first operational run of the LHC began in 2009 and ended in 2012 obtaining a maximal collision energy of 7 TeV. This period is called Run 1. This was followed by Long Shutdown 1 in which the LHC was shut down for upgrades and maintenance. In 2015 Run 2 began and continued until 2018 in which a maximal collision energy of 13 TeV was achieved [25]. This was followed by Long Shutdown 2 which, after some delay, is expected to end in 2022. This in turn will be followed by Run 3, a.k.a. “Phase-I”, and will continue until Long Shutdown 3 begins in 2025. Between then and 2027 the High Luminosity LHC (HL-LHC) will be installed after which the HL-LHC will begin operations marking the beginning of “Phase-II” (Fig. 3.2).

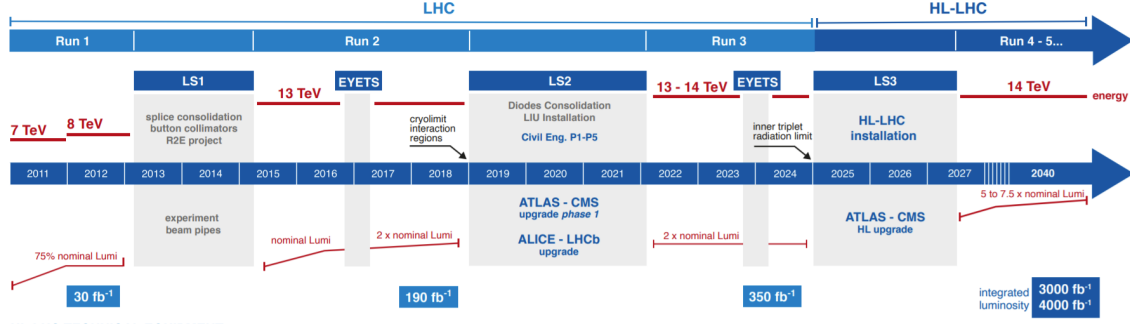


Figure 3.2: The LHC baseline plan for the next decade and beyond, showing the energy of the collisions and luminosity [3].

3.2 The ATLAS Detector

ATLAS (**A Toroidal LHC ApparatuS**) is the largest, general-purpose particle detector experiment being 46 meters long, 25 meters in diameter, and weighing about 7,000 tonnes. The ATLAS detector consists of a series of ever-larger concentric cylinders around the interaction point, i.e. the point where the proton beams from the LHC collide. The detector is divided into four major systems:

1. The Inner Detector
2. The Calorimeters
3. The Muon Spectrometer
4. The Magnet

Each of the three detectors is in turn made of multiple layers (see Fig. 3.3). These detectors work together to allow the separation and identification of different particles. The Inner Detector (ID) precisely tracks electrically charged particles, the calorimeters measure the energy loss of particles traversing the material, and the Muon Spectrometer (MS) detects muons since they are not stopped by the calorimeters. The magnets bend the paths of charge particles via the Lorentz force in both the Inner Detector and the Muon Spectrometer. The only known particle contained in the SM that does not leave a signal in the ATLAS detector is the neutrino, although it can be indirectly inferred by the measurement of a momentum imbalance along the transverse direction.

3.3 ATLAS Coordinate System

In collider physics, and therefore also in ATLAS in particular, a more convenient non-Cartesian coordinate system is used. The ATLAS Cartesian coordinate system is defined as follows: the z-axis

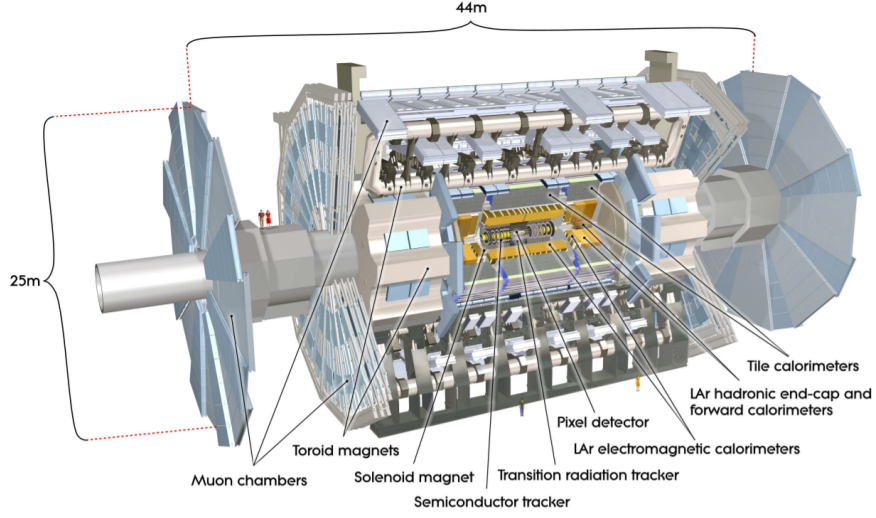


Figure 3.3: Cut-away view of the ATLAS detector.

is defined along the beam pipe in the counter-clockwise direction, the x-axis pointing towards the center of the LHC ring, and the y-axis pointing upwards to the zenith. This coordinate system is depicted in Fig. 3.4.

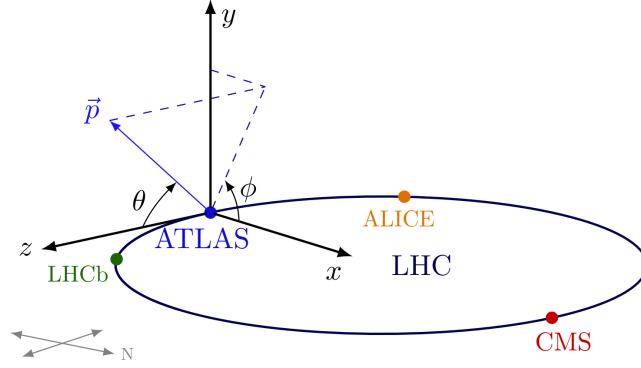


Figure 3.4: The ATLAS Cartesian coordinates.

Since the x-y plane is transverse to the beam axis it is denominated the transverse plane. For a given vector originating from the interaction point, the polar angle θ is defined as the angle between it and the z-axis and the azimuthal angle ϕ is defined as the counter-clockwise angle from the x-axis when the vector is projected onto the transverse plane. Thus, any vector can be decomposed into its longitudinal, i.e., along the z-axis, component and its transverse component. For example, for a

given momentum vector P , the transverse component will be $P_T = P \cdot \sin(\theta)$.

It is useful instead of θ to define the more commonly used pseudorapidity η :

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

This is useful because for massless particles the difference in pseudorapidity $\Delta\eta$ is Lorentz invariant under longitudinal boosts. A zero η value is perpendicular to the beam axis whereas as η approaches infinity it converges with the beam axis. Thus, for high η values, particles will escape the detector by exiting via the uncovered areas near the beam axis. This can be seen visually in Fig. 3.5.

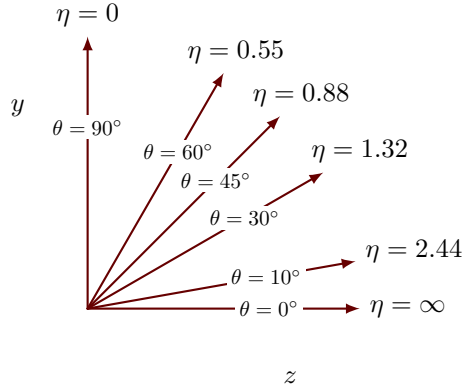


Figure 3.5: Pseudorapidity values shown on a polar plot.

Momenta in the transverse plane are also invariant under longitudinal Lorentz transformations. The standard definition of rapidity y used in experimental particle physics is defined as:

$$y \equiv \frac{1}{2} \ln \left(\frac{E + P_L}{E - P_L} \right)$$

In the high energy limit of $|P| \gg m$ the definitions of the pseudorapidity η and the rapidity y converge. This can be used to rewrite the standard definition of angular separation ΔR as a function of η instead of y :

$$\Delta R \equiv \sqrt{(\Delta y)^2 + (\Delta\phi)^2} \simeq \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

which is Lorentz invariant under a boost along the longitudinal direction.

3.4 The Calorimeters

The ATLAS Calorimeter System forms a concentric cylinder around the Inner Detector and is thus the second detector system reached by particles originating from the IP. The calorimeter is designed to stop and measure the energy that particles lose as they travel through the material. It is created from a high-density material that forces to a stop almost all particles other than muons and neutrinos, absorbing the particles' energy in the process. The cylinder is comprised of a barrel and two end-caps. The total thickness of the EM calorimeter is > 22 radiation lengths (X_0 , the characteristic amount of matter traversed in the material measured in $\text{g} \cdot \text{cm}^{-2}$) in the barrel and $> 24 X_0$ in the end-caps.

The Calorimeter System is divided into two parts (see Fig. 3.6): the Liquid Argon (LAr) Calorimeter and the hadronic Tile Calorimeter. The LAr Calorimeter is an electromagnetic calorimeter designed to measure the energy of particles that participate in electromagnetic interactions such as electrons and photons. The hadronic Tile Calorimeter samples the energy of hadrons as they interact with atomic nuclei.

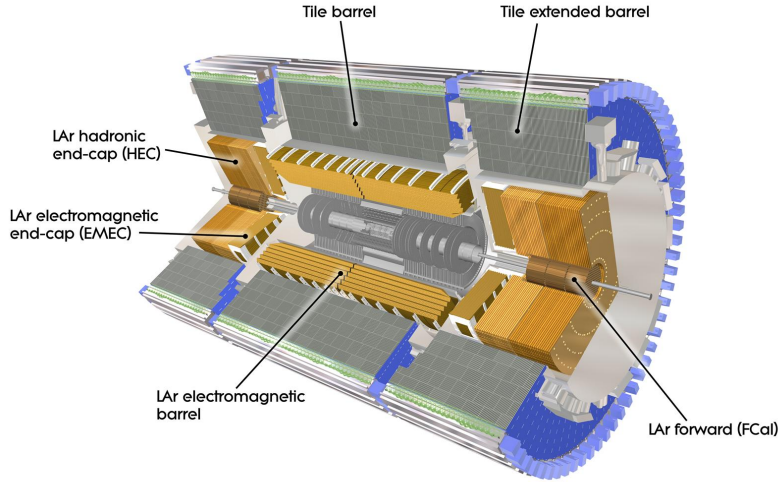


Figure 3.6: The ATLAS Calorimeter System [4].

The LAr Calorimeter is composed of lead plates with liquid argon (LAr) between them. Electrons and positrons passing through the LAr Calorimeter radiate photons via bremsstrahlung, while photons decay into an e^+e^- pair. Both processes thus produce showers in the lead plates and the charged particles from the showers ionize the liquid argon yielding a measurable signal proportional to the energy of the original absorbed particle.

The LAr Calorimeter is divided into a barrel part ($|\eta| < 1.475$) and two end-cap components

($1.375 < |\eta| < 3.2$) and is composed of multiple layers. The first layer is the presampler (PS) which extends up to $|\eta| < 1.8$ with granularity $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. The purpose of the PS is to correct for the energy lost by electrons and photons upstream of the calorimeter. This is followed by three more layers denominated in the following EM1, EM2, and EM3 with the first two with granularity $\Delta\eta \times \Delta\phi = 0.025 \times 0.1$ and EM3 with $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ within the range of $|\eta| < 2.5$. In the range $2.5 < |\eta| < 3.2$ the granularity is degraded. A segment from the barrel can be seen with respective measurements in Fig. 3.7

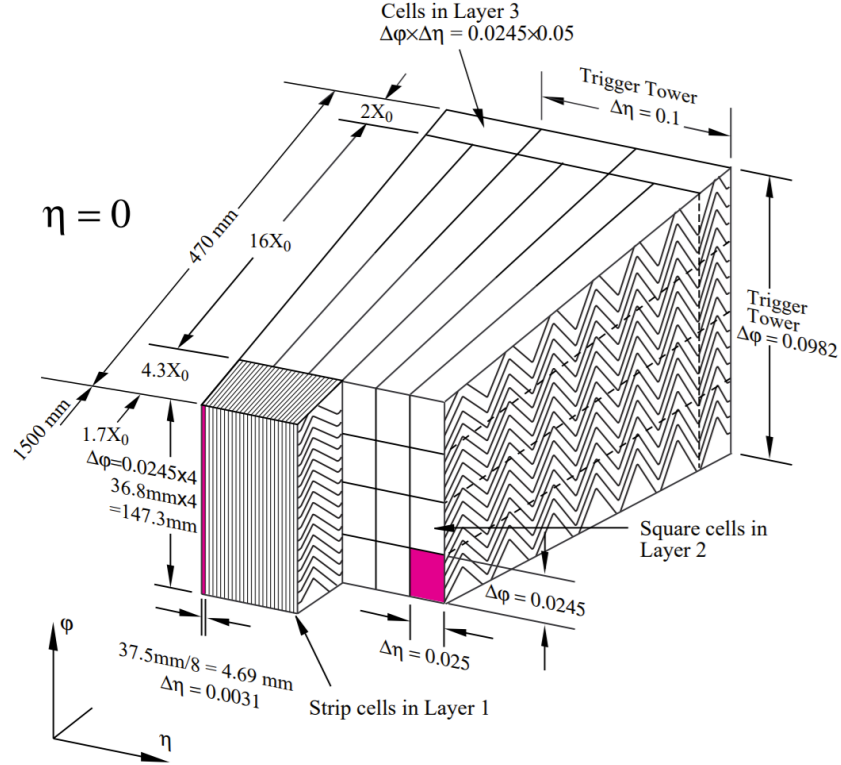


Figure 3.7: Schematic of the Liquid Argon Calorimeter [5].

The second part, the Hadronic Calorimeter, is placed directly outside the LAr Calorimeter envelope and is also composed of two parts: the Tile Hadronic Calorimeter and the LAr Hadronic End-cap Calorimeter (HEC). The Tile Hadronic Calorimeter is made of one central barrel and two extended barrels covering the range $|\eta| < 1.7$. The HEC slightly overlaps with the Tile Calorimeter and extends up to $|\eta| < 3.1$. Both have a cell granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$.

3.5 ATLAS Trigger System

Every 25 ns there is a proton beam crossing, a.k.a. a ‘bunch crossing’, in the center of the ATLAS detector. This amounts to 40 million bunch crossings per second [9]. Each collision, called an event, takes up approximately 1.6 MB [26]. Thus, to save all events, ATLAS would require a bandwidth of $1.6 \times 40 \cdot 10^6 \approx 60 \text{ TB/s}$ - an enormous and unfeasible rate. To solve this problem, the trigger system identifies in real time the most interesting events which are then recorded for later detailed analysis. The rest of the events are not kept.

The triggering is accomplished in two steps which each further reduce the event rate. The first step is the Level-1 (L1) trigger which uses reduced granularity from the calorimeters and the Muon Spectrometer and reduces the rate to 100 kHz. This is followed by the High Level Trigger (HLT) which saves triggered events to the disk at a final rate of 1.5 kHz. While the HLT is software based, the L1 trigger is entirely hardware based. This is necessary for it to cope with the very high collision rate in the detector. A diagram of the trigger system can be seen in Fig. 3.8

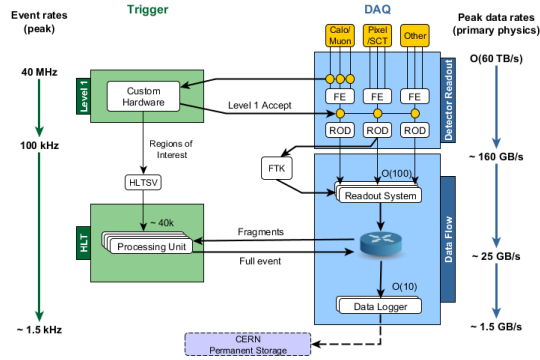


Figure 3.8: Diagram of the ATLAS Data Acquisition System (DAQ) and Trigger System in Run 2 and Run 3. The reductions in data rates can be seen for each stage in the trigger [6].

CHAPTER

4

THE TAU TRIGGER

4.1 Introduction

Since tau leptons play a key role in the ATLAS physics program, events containing tau decays are valuable events. Ideally, all ATLAS collision events would be saved to the disk including all the events containing tau decays. Sadly, as mentioned in [section 3.5](#), because of the high rates it is not technologically feasible to save all the collision events. Thus, it is the purpose of the ATLAS Trigger System to choose which events contain the most interesting physics and save only them. The trigger is divided into many subsystems each trying to detect events containing specific particles (or sets of particles) with certain characteristics. For example, the e/γ algorithm's purpose is to distinguish electrons and photons from the dominant jet background. The purpose of the tau algorithm is similarly to distinguish taus from the dominant jet background.

Taus can decay either leptonically or hadronically (see [Table 2.1](#)). Unfortunately, leptonically decaying taus cannot be separated at trigger level from prompt leptons, i.e., those originating from the main collision. A tau decaying to an electron for example cannot be distinguished at trigger level from an electron originating from the interaction point. Because of this, the purpose of the tau trigger is only to detect hadronically decaying taus.

4.2 LAr Phase-I Upgrade

Part of the Phase-I upgrade of the LAr Calorimeter is the design and installation of new trigger readout electronics. While the physical cell resolution is higher, in Run 2 the trigger only had sufficient computational resources to evaluate the detector readout at a coarse-grained (downsampled) resolution. The Phase-I upgrade will improve the cell readout resolution allowing a much more fine reading [7, 27]. The cell readout resolution is also known as granularity and is generally given in $\Delta\eta \times \Delta\phi$ dimensions. In Run 2 the granularity was $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ with a single value for all the energy deposited in the electromagnetic layer. Each cell was called a Trigger Tower. After the Phase-I upgrade, what was once registered in the trigger as a single electromagnetic layer will be divided into four - EM0 (the PS), EM1, EM2, and EM3. The two main layers are EM1 and EM2 and they capture the majority of the electromagnetic energy deposits. They are also by far the thicker layers (see Fig. 3.7) with EM2 being the thickest. In addition, the granularity in these layers is improved in the Phase-I upgrade to $\Delta\eta \times \Delta\phi = 0.025 \times 0.1$. These new Phase-I cells with improved granularity are called Supercells. Table 4.1 sums up the original physical granularity of the cells (which can be accessed offline), the granularity during Run 2, and the improved granularity after the upgrade. An example of the improvement in granularity can be seen in Fig. 4.1.

Layer	Elementary Cell ($\Delta\eta \times \Delta\phi$)	Trigger Tower ($\Delta\eta \times \Delta\phi$)	Supercell ($\Delta\eta \times \Delta\phi$)
PS	0.025×0.1	0.1×0.1	0.1×0.1
EM1	0.003125×0.1		0.025×0.1
EM2	0.025×0.1		0.025×0.1
EM3	0.05×0.1		0.1×0.1

Table 4.1: Cell granularity of the four LAr Calorimeter layers as seen offline, during Run 2, and after the Phase-I upgrade.

4.3 Seeding and Clustering

The calorimeter cell matrix extends in the aforementioned granularity in an η range of $-2.5 < \eta < 2.5$ and in ϕ the full azimuthal coverage of $-\pi < \phi < \pi$. Each beam collision, which occurs every 25 ns, creates many particles and jets which deposit their energy in different areas of the detector. This leads to many deposits in widely varying η and ϕ coordinates. The final product of the collision, called an event, is the full matrix of the detector energy deposits along the entire range of η and ϕ for each layer of the detector. Since events contain large amounts of information and are generated every 25 ns only the most interesting events can be kept. Specifically, events containing taus are

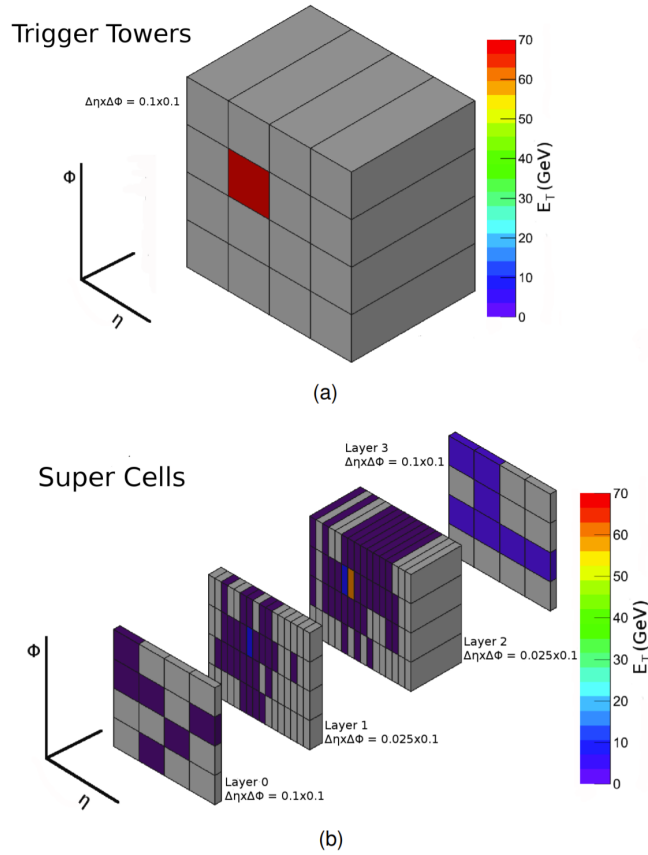


Figure 4.1: An electron (with 70 GeV of transverse energy) as seen by the existing Level-1 Calorimeter trigger electronics (a) and by the proposed upgraded trigger electronics (b) [7].

deemed interesting for the ATLAS physics program. It is the purpose of the tau trigger to detect these events and notify the appropriate system to save them to the disk.

Given this very large, three dimensional matrix of energy deposits called an event, the tau trigger needs to determine whether within this matrix any of the energy deposits correspond to the decay products of a tau particle. The first step applied is called seeding and refers to finding specific cells in the event which are likely significant. In Phase-I the seeding is based on significant energy deposits relative to the expected background in EM2 supercells that have energy deposits greater than all their neighbours. These “seeds” provide starting points for the trigger algorithm.

The next standard step in triggering is called clustering. Clustering involves reading the information from the area around the seed. Choosing the optimal cluster size is important as too small a cluster will not provide sufficient information but too large a cluster will likely contain both more

noise and irrelevant information from other nearby decay products. For the tau trigger a cluster size of $\Delta\eta \times \Delta\phi = 0.3 \times 0.3$ was chosen since it is the minimal size that the decay products of a tau are guaranteed to be contained in it.

4.4 Trigger Object (TOB)

The final result of the clustering is the energy deposits in a $\Delta\eta \times \Delta\phi = 0.3 \times 0.3$ area of the detector. Every $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ is composed of eleven cells, ten Supercells from the improved granularity in the LAr Calorimeter and a single cell of dimensions $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ from the Hadronic Calorimeter (see Fig. 4.2). Since each cluster has an area of $\Delta\eta \times \Delta\phi = 0.3 \times 0.3$ it is in fact a 3×3 matrix of these resulting in a total of $3 \cdot 3 \cdot 11 = 99$ cells per cluster. These 99 cells along with their relevant η and ϕ coordinates are saved into what is called a trigger object (TOB). Each TOB is evaluated by the tau trigger algorithm and is selected as a tau candidate or not. Since each event usually contains many TOBs, the tau trigger algorithm runs individually on each TOB. If even one of the TOBs is identified as a tau then the entire event is saved. This is critical since the triggering is done on a per-event basis and not individually for each TOB.

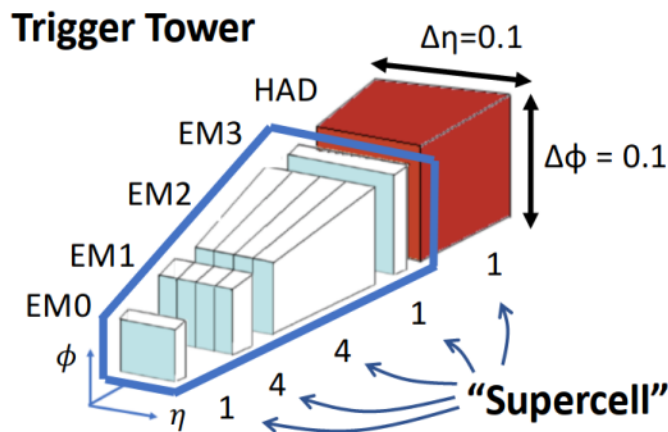


Figure 4.2: The granularity increase. A single cell in Run 2 (the ‘Trigger tower’) became ten separate cells (‘Supercells’) in Run 3.

4.5 Isolation

An additional step, not mentioned in the previous section, is called isolation. The purpose of isolation is to try to discriminate between the prompt particles that are of interest and between those that aren't prompt, e.g., jets resulting from a hadronic decay. Since jets typically have a larger

spread this reduces the background fed as input to the trigger algorithm easing later classification. In Run 2 the isolation was calculated by summing the $\Delta\eta \times \Delta\phi = 0.4 \times 0.4$ ring around the Run 2 TOB of dimensions $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$. If this sum was greater than a certain threshold the TOB was considered not well isolated. This calculation was carried out both for the sum of all cell energies in the electromagnetic layer and in the hadronic layer. Any TOB that is not well isolated is not worth keeping and as such was classified as background.

4.6 Hardware Requirements

In order to facilitate the trigger's increase in granularity, the Calorimeter Level-1 trigger (L1Calo) electronics are undergoing a major upgrade for Phase-I. Although the existing legacy hardware will still run alongside the upgraded electronics, they will eventually be completely replaced by them. The upgrade consists of a new system of feature extractor (FEX) modules which will process the higher granularity information from the calorimeters and execute more sophisticated algorithms for improved triggering [28]. A schematic of the Phase-I upgrade electronics can be seen in Fig. 4.3.

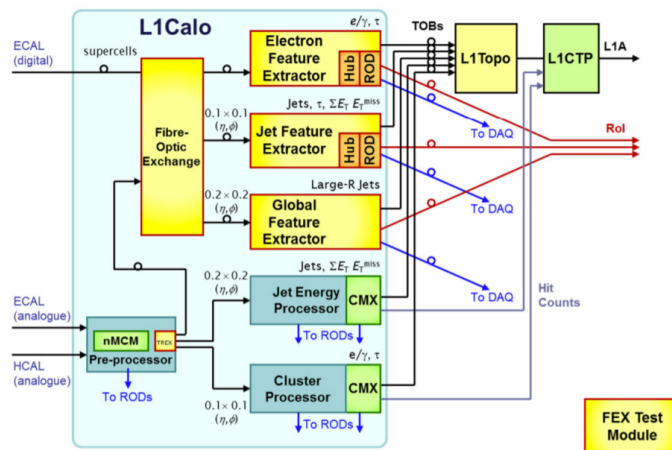


Figure 4.3: ATLAS Level-1 Calorimeter Trigger Phase-I Upgrade Electronics [8].

The FEX modules consist of the electron, global, and jet Feature Extractors, or eFEX, gFEX, and jFEX respectively. The eFEX system consists of 24 modules with 4 Xilinx Virtex 7 processor FPGAs (Field Programmable Gate Arrays - programmable integrated circuits) and 1 Xilinx Virtex 7 control FPGA on each module. The purpose of the eFEX is to identify electron, photon, and tau objects by utilizing the now increased granularity of the Supercells. This is where the tau trigger algorithm resides. The gFEX system consists of 3 Xilinx Virtex Ultrascale+ FPGAs and 1 Zynq Ultrascale+ System on Chip on 1 ATCA module. The entire calorimeter is read out on one board,

which identifies large radius jets, E_T^{miss} and $\sum E_T$. The jFEX system is made up of 6 ATCA modules, each with 4 Xilinx Virtex Ultrascale+ FPGAs and 1 Zynq Ultrascale+ control FPGA. The jFEX identifies large and small radius jets, taus, E_T^{miss} and $\sum E_T$.

Since the tau trigger algorithm must be fully contained in the eFEX it is pertinent to understand the resources available to it. The Virtex 7 FPGAs used in the eFEX[29] are of device version XC7VX550T and contain 86,600 slices[30]. Each FPGA slice contains four LUTs (Look Up Tables) and eight flip-flops. In addition, since each FPGA is responsible for processing multiple TOBs at once there need to be eight algorithm cores per FPGA to cover the full detector. As the data is received in the 40 MHz domain and is processed in the 200MHz domain there are 5 clock cycles to process each event allowing a total of $8 \cdot 5 = 40$ locations to be processed at once. Since eight separate instances of the algorithm must run concurrently on the FPGA each instance can only use a fraction of the resources, i.e., LUTs and flip-flops, available. In addition, since tau triggering is only one of the purposes of the eFEX, only some of the FPGA resources can be used for this purpose in the first place. Compounding this with the throughput and latency specifications, the requirements placed upon the tau triggering algorithm are very limiting.

4.7 Background Rate Requirement

Although the Phase-I LAr Calorimeter upgrade will bring an improvement in cell granularity, the ATLAS front-end detector electronics will remain largely unchanged. Therefore, the total Level-1 trigger rate will still be limited by the Run 2 readout bandwidth of the front-end electronics of 100 KHz or less [31]. This limitation provides an upper bound to the rate of events that can be saved by the trigger. Since the vast majority of events are background events (events containing only QCD jets), it is sufficient to place the rate limitation on the total number of accepted background events and treat the number of signal events (events containing taus) as negligible. This leads to the definition of the background rate R_{bg} : the maximal fraction of misidentified, and therefore triggered upon, background events that does not overwhelm the electronics. R_{bg} must by definition always be between zero and one. Since this limit hasn't changed, R_{bg} in Run 3 is the same as in Run 2.

The background rate R_{bg} is directly related to another concept called the false positive ratio or FPR (also known as fall-out or false alarm ratio). It is defined as:

$$FPR = \frac{FP}{FP + TN}$$

where FP (false positives) is the total number of background events that were falsely classified as signal events and TN (true negatives) is the total number of background events that were classified correctly. The background rate requirement necessitates that the tau trigger's FPR will always be smaller or equal to R_{bg} . Else, the electronics will fail to process all the events within the bandwidth

constraints. Thus, any tau trigger algorithm must always fulfil the criterion:

$$FPR \leq R_{bg}$$

4.8 The Need for a New Tau Trigger Algorithm

In Run 3, the LHC instantaneous luminosity will double to $\sim 2.5 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$ [28]. This will greatly increase the pileup, i.e., the number of additional proton-proton collisions that take place simultaneously with the collision of interest. Since an increase in pileup directly implies an increase in TOBs per event, the tau trigger will be required to process a much larger number of TOBs per event. This has two effects. First of all, this increases the probability of an event containing a tau, resulting on average in more signal events. Since the upper limit on the rate hasn't changed since Run 2, the number of events that can be triggered on remains the same. From this it follows that in Run 3, without an improvement to the trigger algorithm, there will be a degradation in percentage of signal events triggered on. Second, the background events (those without taus) are now more difficult to classify as such since the number of TOBs that can incorrectly be detected as taus increases. Overall, this means that the Run 3 environment is more challenging than Run 2. On the other hand, the extra information gained by the improvement in cell granularity allows for more powerful and inventive algorithm design.

CHAPTER

5

GRADIENT BOOSTED DECISION TREES (GBDT)

5.1 Hardware Constraints

As mentioned in the previous chapter, Run 3 brings with it both the challenge of increased instantaneous luminosity and the benefit of increased granularity. In Run 2 the tau trigger algorithm was designed as follows: a $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ local maximum was found, this constituted the Run 2 TOB. Isolation was applied using a $\Delta\eta \times \Delta\phi = 0.4 \times 0.4$ ring around the TOB. Then, the reconstructed energy of the TOB was calculated as the sum of the maximum of all possible pairs of two adjacent cells in the electromagnetic calorimeter and the sum of all four cells of the hadronic calorimeter. Finally, a transverse energy threshold of 12.5 GeV was required for a TOB to be classified as an L1-Calo tau candidate. If even a single TOB in an event was classified as a tau candidate then the entire event passed the first level of the trigger. If it also passed the HLT it was then saved to disk. Otherwise, it was classified as a background event and discarded. The Run 2 algorithm is graphically shown in Fig. 5.1.

Although the Run 2 tau trigger algorithm can similarly be used in Run 3 as well, since it doesn't

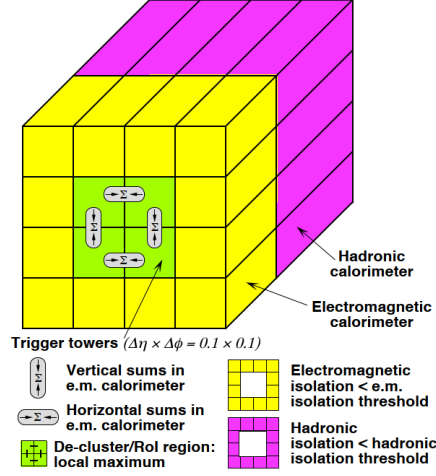


Figure 5.1: Schematic view of the trigger tower used as input to the L1Calo Run 2 tau trigger algorithm [9].

benefit from the improved cell granularity it is likely no longer the optimal algorithm. Multiple attempts have been made at replacing the Run 2 algorithm. One of these used varied geometrical patterns within the area of the $\Delta\eta \times \Delta\phi = 0.3 \times 0.3$ cluster that best reconstructed the TOB's energy. Similar to Run 2, a threshold was applied to the reconstructed TOB's energy classifying it as either signal or background. However, this algorithm as well as other attempts have not yielded a significant improvement over the Run 2 algorithm.

The problem that the tau trigger algorithm faces is a binary classification problem - the TOB is one of two categories, either signal or background. In addition, TOB data can be created artificially through simulations. As such, being a binary classification problem with a wealth of labeled data available, it would seem obvious that the problem is well suited for machine learning methods. The obvious method of choice would be to implement a neural network - a machine learning based algorithm that mimics the way the human brain operates. With some upsampling, the TOB's cells can be converted to a $3 \times 12 \times 5$ matrix. This can be treated as a 3×12 RGB (red, green, blue) image with five color channels instead of three. Since TOB's internal structure is equivalent to those of images, a Convolutional Neural Network (CNN) [32] or ResNet [33] can now be trained to be subsequently implemented as the tau trigger algorithm. Another option is to train a Graph Neural Network (GNN) such as a Deep Set [34] using each cell as a node in the graph. Since most cells are empty this method will likely improve upon the results of the CNN. Either way, the optimal method can likely yield a powerful discriminator between signal and background TOBs.

Unfortunately, as the tau trigger algorithm is hardware based, any such trained model must be then translated into VHDL code to be implemented on the Xilinx FPGA. Although there exist some

libraries that support specific machine learning models this condition severely limits the variety of available models. This problem is exacerbated by the minimal resources available on the FPGA for the task. Since only part of each FPGA is allocated for the tau trigger and on that part eight instances of the tau trigger must run simultaneously, the final model after training must be sufficiently small to fulfill these constraints. This further limits the space of model choices that can be used. A million parameter ResNet, for example, vastly exceeds the maximal resources available. Even a minimal CNN, with limited channels for each feature map is extremely challenging and borderline impractical.

Given these limitations, the notion of applying image based machine learning methods becomes quite ambitious. This thesis suggests an alternative approach, specifically, the use of gradient boosted decision trees. This method is traditionally reserved for tabular data, i.e., data arranged in a table where each column has significance. Although the TOB’s data isn’t initially tabular, a physical understanding of the differences between the shapes of the electromagnetic and hadronic showers developed in the calorimeters by tau decay products vs. those developed by QCD jets can be utilized to design high level features which can then be used to train the model.

5.2 GBDT

A decision tree is a tree shaped flowchart which leads to a decision based on the attributes of the test sample. A decision tree is composed of a root node, branches and leaves (end nodes). At each node a conditional test is performed on one of the attributes and the branch that fulfills the condition is traversed. This is done recursively until a leaf node, which represents a final decision, is reached. The imposed condition at the node can be either categorical or a threshold on a numeric value of one of the attributes.

Since the decision making process is discrete, i.e. the number of leaves is a small integer with each of them leading to a specific value, decision trees can form a classifier. In binary classification the decision tree assigns one of two classes to each sample. In addition, the tree will split into exactly two possibilities at each node until the leaf nodes are reached. An example of a binary tree that classifies a TOB as either a tau particle or a jet can be seen in Fig. 5.2.

Given two classes “signal” and “background”, the decision tree is usually trained by finding the optimal thresholds on the attributes that maximize the classification accuracy.

While decision trees can be used for binary classification, in practice they tend to be weak learners and do not usually classify very well. They also usually overfit to the training set. To rectify this, instead of using a single tree, an “ensemble” of decision trees can be used to improve the classification results. By combining the predictions of all the trained trees the accuracy achieved will significantly improve upon that of any single tree. There exist two primary methods to train an ensemble of decision trees.

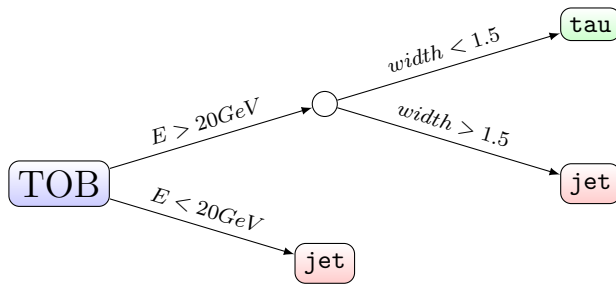


Figure 5.2: Example of a decision tree used to classify a TOB.

The first method, called bootstrap aggregating (bagging), randomly generates a new dataset to train each tree. Generally, the majority of the new dataset is randomly taken from the original samples and the rest are duplicates. Thus, each tree is trained on a similar but different dataset generating many trees that together form the ensemble. Random Forest [35] is a method that uses bagging to generate its trees and then uses a majority voting system to choose the winning class.

The second method, called boosting, trains the trees sequentially instead of in parallel. Thus, each tree is created with the specific purpose of fixing the misclassifications of all the previous trees. For example, AdaBoost [36], short for Adaptive Boosting, up-weights observations that were previously misclassified thus giving future trees impetus to classify them correctly. It uses stumps, small trees with only one condition and two leaves, to quickly generate the ensemble. Gradient Boosting [37] improves upon AdaBoost by generalizing the loss function and using gradient descent for convergence. This is done by calculating the ‘pseudo-residuals’, i.e., the differences for each sample between the observed value and the current prediction and then by training a new tree to minimize these residuals. This is done continuously until the residuals are very small or the maximal number of trees is reached. This choice of using gradient boosting with trees as the weak learner is known as GBDT and within the CERN community it is generally known as BDT. Finally, it is worth mentioning XGBoost (eXtreme Gradient Boosting) [38] which offers a further improvement over standard GBDT.

5.3 Hyperparameters

The training of the GBDT model is dependent on several external parameters, denominated hyperparameters, that define the model and must be chosen before the training begins. Different choices of GBDT hyperparameters can significantly affect the quality as well as the size of the model. A non-exhaustive list of GBDT hyperparameters is presented here:

- **minimum_samples_split** - Defines the minimum number of samples (or observations) which are required in a node to be considered for splitting. Useful for not overfitting.

- **minimum_samples_leaf** - Defines the minimum samples (or observations) required in a terminal node or leaf. Used to control over-fitting similar to `min_samples_split`.
- **maximum_depth** - The maximum depth of trees.
- **maximum_features** - The number of features to consider while searching for a best split. These will be randomly selected. The default is the square root of the total number of features.
- **learning_rate** - The relative weights given to each tree in the ensemble. A lower learning rate generally requires a higher number of trees.
- **number_of_estimators** - The number of trees trained in the ensemble.
- **subsample** - The fraction of observations to be selected randomly for each tree. A value smaller than one generally makes the model more robust.

These hyperparameters are tuned in order to optimize the final model for the tau trigger.

CHAPTER

6

GBDT BASED TAU TRIGGER

6.1 Training Data

As with any learning based algorithm, training data is required to teach the model. In this case, tagged tau trigger data taken directly from the ATLAS detector best represents the test case and serves as the ideal training dataset. Unfortunately, this data is in short supply and insufficient for training. Therefore, an alternative data source was used - Monte Carlo (MC) generated samples. These are accurate enough to properly model the detector and their use results in a model that generalizes well for true ATLAS data.

For the signal, $5 \cdot 10^5$ $Z \rightarrow \tau\bar{\tau}$ decay events were generated using PowHeg+Pythia8. This decay is sufficiently representative to serve as a training dataset for all possible SM tau events in the detector. This is because any other decay (e.g. W or Z boson jet events) can only be both more energetic and more collimated than the decay chosen for training and thus easier to identify. The only exception to this are theoretical BSM low mass particles which are not accounted for. Since these would be lost to energy cuts later in any case there is no downside to only using these samples and they are thus sufficient.

For background samples, $2 \cdot 10^5$ di-jet QCD events were generated using Sherpa. Since these have by far the highest cross-section, they represent the vast majority of background in ATLAS

and are the main source of the background TOBs. Since there are no other events that can both be confused with tau decays and common enough to overwhelm the L1 trigger it is sufficient to consider only these background events. Once generated, the events were processed through the ATLAS detector simulation [39] based on Geant4 [40]. The Feynman diagrams of the generated samples for signal and background can be seen in Fig. 6.1.



Figure 6.1: Feynman diagrams of the signal and background samples used for training the model.

Each signal sample consists of the ‘truth_pt’ - the true visible energy of the tau, the η and ϕ information, and the TOB including all 99 energy deposits. The true tau energy distribution of all the samples is shown in Fig. 6.2. Each background sample consists of the η and ϕ information, and the TOB including all 99 energy deposits. An example sample TOB is shown in Fig. 6.3.

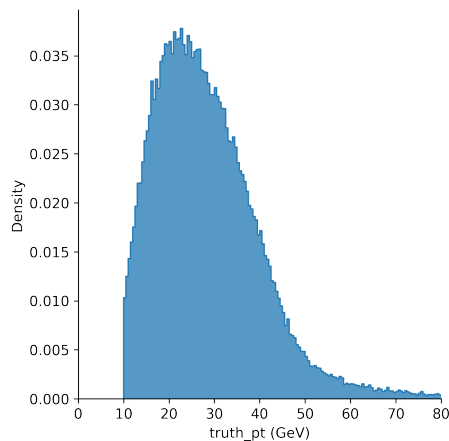


Figure 6.2: Transverse momentum distribution of the visible component of the decay products from tau leptons originating from Z boson decays.

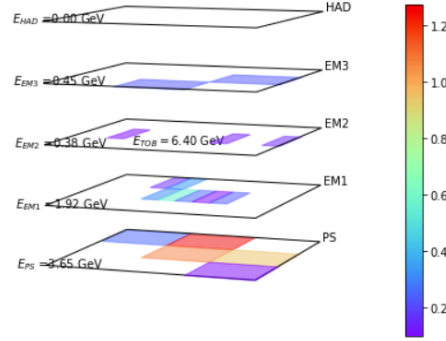


Figure 6.3: The energy deposits in each layer of the calorimeter of a randomly selected TOB.

6.2 High Level Features

All classifiers, and in particular GBDTs, assign a category to each observation in the data. For the case of the tau trigger algorithm the observation is the TOB and the output is whether or not it is classified as a tau candidate. In the training process GBDTs create an ensemble of trees where each tree consists of a series of decision thresholds on the TOB's features. These thresholds on the features are optimized through gradient boosting to yield the best possible classification of the observations. Naively, the raw energy deposits of the TOB could have been used as the features for the GBDT. Unfortunately, as the specific energy deposits in the cells are not particularly meaningful, the resultant thresholds on the cell energies fail to successfully discriminate between taus and QCD jets producing a model of limited predictive power. Thus, more meaningful features, otherwise known as high level features, must be constructed.

In total, 24 features were designed for this purpose utilizing the lateral radial symmetry of the TOB (up to cell discretization) and the relative narrowness of tau decay products relative to QCD jets. These features are defined using the notation described in the following. $E_{h,w}^l$ is the energy of the cell of layer l with height index h and width index w , where l ranges from zero to four with the five indices corresponding to the layers: PS, EM1, EM2, EM3 and HAD, h ranges from one to three since TOBs consist of a 3×3 grid of Supercells, and w ranges from one to three for the PS, EM3 and HAD layers and one to twelve for EM1 and EM2 due to the improved granularity. With this notation, the TOB energy of a given layer E^l can be written as:

$$E^l = \begin{cases} \sum_{h=1,w=1}^{3,3} E_{h,w}^l & l \in \{0, 3, 4\} \\ \sum_{h=1,w=1}^{3,12} E_{h,w}^l & l \in \{1, 2\} \end{cases}$$

The energy of an entire TOB can then be expressed as:

$$E^{TOB} = \sum_{l=0}^4 E^l$$

The 24 high level features used for training the GBDT are:

- **pt** - The total energy deposited in the TOB, given by, E^{TOB} .
- **eta** - The absolute eta value $|\eta^{TOB}|$ of the central Supercell of the TOB.
- **width** - This feature quantifies how focused the TOB's energy is in its center. The normalized η and ϕ values, $\Delta\eta_{h,w}^l$ and $\Delta\phi_{h,w}^l$, are defined as the distance in η and ϕ units of cell h, w of layer l from the center of the TOB. These are defined as follows:

$$\Delta\eta_{h,w}^l = \eta_{h,w}^l - \eta^{TOB}$$

$$\Delta\phi_{h,w}^l = \phi_{h,w}^l - \phi^{TOB}$$

Where η^{TOB} and ϕ^{TOB} are the η and ϕ values of the center of the TOB. $\Delta\eta_{h,w}^l$ and $\Delta\phi_{h,w}^l$ can have values in the range $[-0.15, 0.15]$. Thus, 'width' is defined by the average of the energy fractions of the cells multiplied by their radial distance from the center of the TOB:

$$\frac{\sum_{l,h,w} \left(E_{h,w}^l \cdot \sqrt{(\Delta\eta_{h,w})^2 + (\Delta\phi_{h,w})^2} \right)}{E^{TOB}}$$

- **depth** - The average penetration depth of the energy deposits in the TOB along the calorimeter depth axis:

$$\frac{\sum E^l \cdot D^l}{E^{TOB}}$$

Where D^l is the depth of the center of layer l in cm when measured with respect to the first layer of the calorimeter and is given by $D = [5, 55, 268, 499, 800]$.

- **density** - the ratio of the cell energies squared to their respective volumes:

$$\ln \left(\frac{\sum_{l,h,w} \left(\left(10^3 \cdot E_{h,w}^l \right)^2 / V^l \right)}{E^{TOB}} \right)$$

Where the V^l is the volume in cm^3 of a single cell in layer l . The height of all layers is 147cm, the width also 147cm except for EM1 and EM2 which are a quarter of that, and the depth

varying for each layer. Thus the vector V equals:

$$V = \begin{bmatrix} 147 \cdot 147 \cdot 10 \\ 147 \cdot 36.8 \cdot 90 \\ 147 \cdot 36.8 \cdot 336 \\ 147 \cdot 147 \cdot 42 \\ 147 \cdot 147 \cdot 42 \end{bmatrix}$$

- **frac0** - the fraction of total energy in the PS:

$$\frac{E^0}{E^{TOB}}$$

- **frac1** - the fraction of total energy in EM1:

$$\frac{E^1}{E^{TOB}}$$

- **frac2** - the fraction of total energy in EM2:

$$\frac{E^2}{E^{TOB}}$$

- **frac3** - the fraction of total energy in EM3:

$$\frac{E^3}{E^{TOB}}$$

- **frach** - the fraction of total energy in the HAD layer:

$$\frac{E^4}{E^{TOB}}$$

- **max1** - the energy of the cell with the highest energy value in the second layer:

$$\max_{h \in \{1,2,3\}, w \in \{1, \dots, 12\}} \{E_{h,w}^2\}$$

- **max2** - the energy of the cell with the highest energy in the second layer, now excluding the cell with maximum energy (found in the previous feature) and its adjacent cells:

$$\max_{h \in \{1,2,3\}, w \in \{1, \dots, 12\}} \{E_{h,w}^2\}$$

- **maxratio** - the ratio between the two maxima:

$$\frac{\max_2 \{E_{h,w}^2\}}{\max \{E_{h,w}^2\}}$$

- **maxdist** - the Manhattan distance (or L_1 norm) between the two maxima cells:

$$\left\| \underset{h,w}{\operatorname{argmax}} \{E_{h,w}^2\} - \underset{h,w}{\operatorname{argmax}_2} \{E_{h,w}^2\} \right\|_{L_1}$$

If the indices of the maximum cell in layer two are given by h_1, w_1 and of the second maximum by h_2, w_2 , then maxdist can be simplified as:

$$|h_2 - h_1| + |w_2 - w_1|$$

For example, in Fig. 6.4 the distance between the maxima is equal to six.

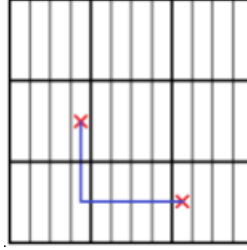


Figure 6.4: It is required to travel six cells in L_1 distance to arrive from one maximum to the other. Thus, in this example maxdist=6.

- **maxf1** - the ratio between the maximum and the total energy of the second layer:

$$\frac{\max \{E_{h,w}^2\}}{E^2}$$

- **maxf2** - the ratio between the second greatest cell energy and the total energy of the second layer:

$$\frac{\max_2 \{E_{h,w}^2\}}{E^2}$$

- **maxf3** - the ratio between the sum of both maxima and the total energy of the second layer:

$$\frac{\max \{E_{h,w}^2\} + \max 2 \{E_{h,w}^2\}}{E^2}$$

- **emhad** - the ratio between the electromagnetic layers and the hadronic layer:

$$\frac{\sum_{l=0}^3 E^l}{E^4}$$

- **had3** - the ratio between the first three layers and the last two:

$$\frac{E^0 + E^1 + E^2}{E^3 + E^4}$$

- **rmoment2_<1>** - this in fact defines five separate features, one for each layer l . For each layer l , this defines the energy weighted second moment of the radial (i.e. longitudinal) distances from the ‘center of mass’ of the TOB. The center of mass is the energy weighted average η and ϕ coordinates of the TOB. The center of mass coordinates are defined:

$$\eta_{CoM}^l = \sum_{h,w} \frac{E_{h,w}^l \cdot \eta_{h,w}}{E^l}$$

$$\phi_{CoM}^l = \sum_{h,w} \frac{E_{h,w}^l \cdot \phi_{h,w}}{E^l}$$

Although these η and ϕ coordinates will still typically point to somewhere in the central Supercell of the TOB, it represents a more accurate center than the middle of the central cell. Utilizing these definitions, **rmoment2_<1>** can be defined for layer l as:

$$\sum_{h,w} \frac{E_{h,w}^l \cdot \left[(\eta_{h,w} - \eta_{CoM}^l)^2 + (\phi_{h,w} - \phi_{CoM}^l)^2 \right]}{E^l}$$

- **dmoment2** - similar to the previous feature, this feature gives the second moment of the energy weighted distances along the calorimeter depth (i.e. transverse) axis:

$$\sum_l \frac{E^l \cdot (D^l - r_{\text{depth}})^2}{E^{\text{TOB}}}$$

Where r_{depth} is the depth feature and D is once again given by $D = [5, 55, 268, 499, 800]$.

While specific physical intuition is not given for each individual feature, the leading motivation was the statistically wider energetic signature of QCD jets over taus and potential statistical differences in average depositions in the layers. The features were all designed to be symmetric in η and ϕ as there is no physically preferred orientation for the deposits. The feature distributions can be seen in Fig. 6.5 and the feature correlations in Fig. 6.6.

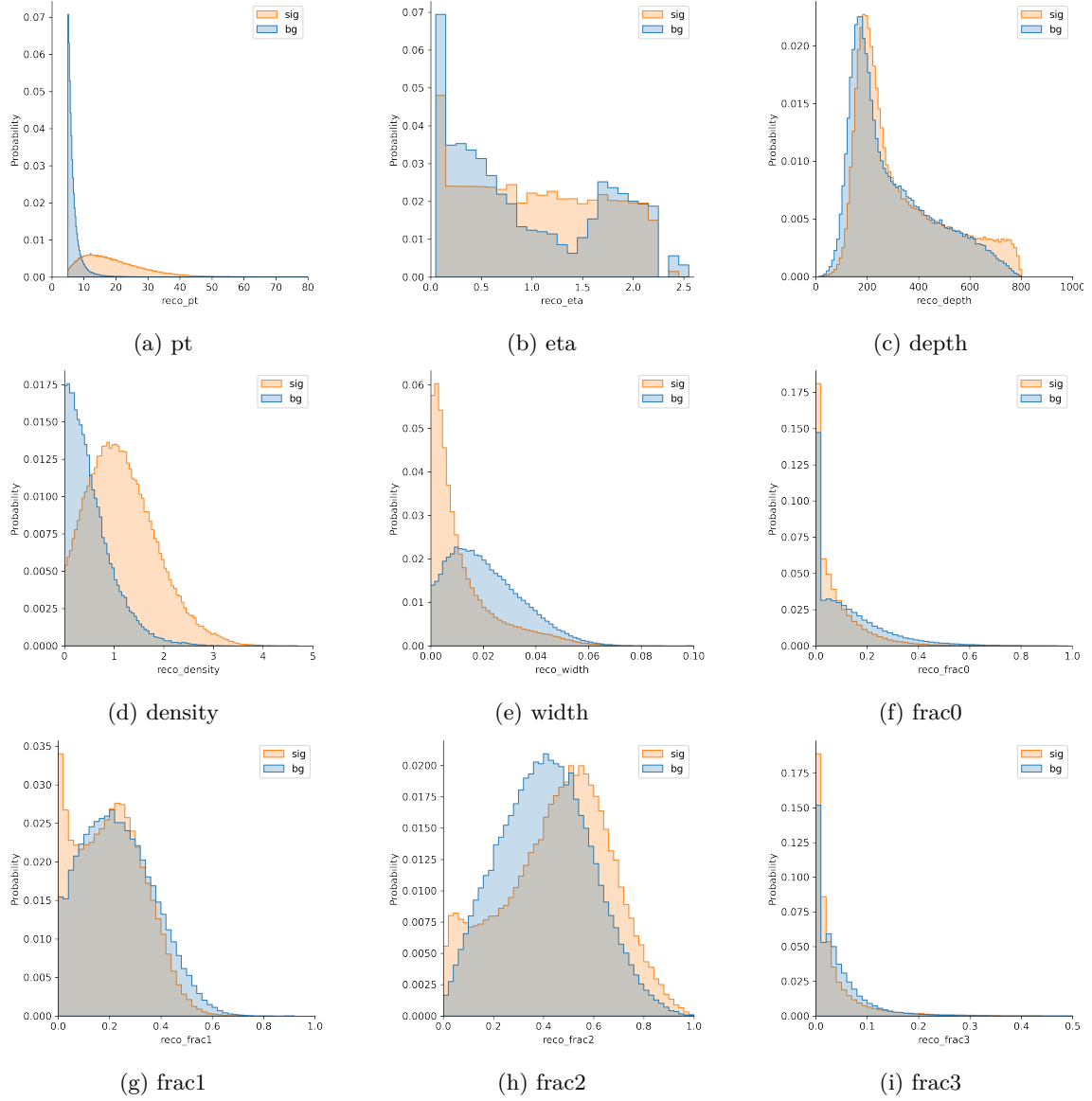


Figure 6.5: Signal and background distributions for all 24 high level features.

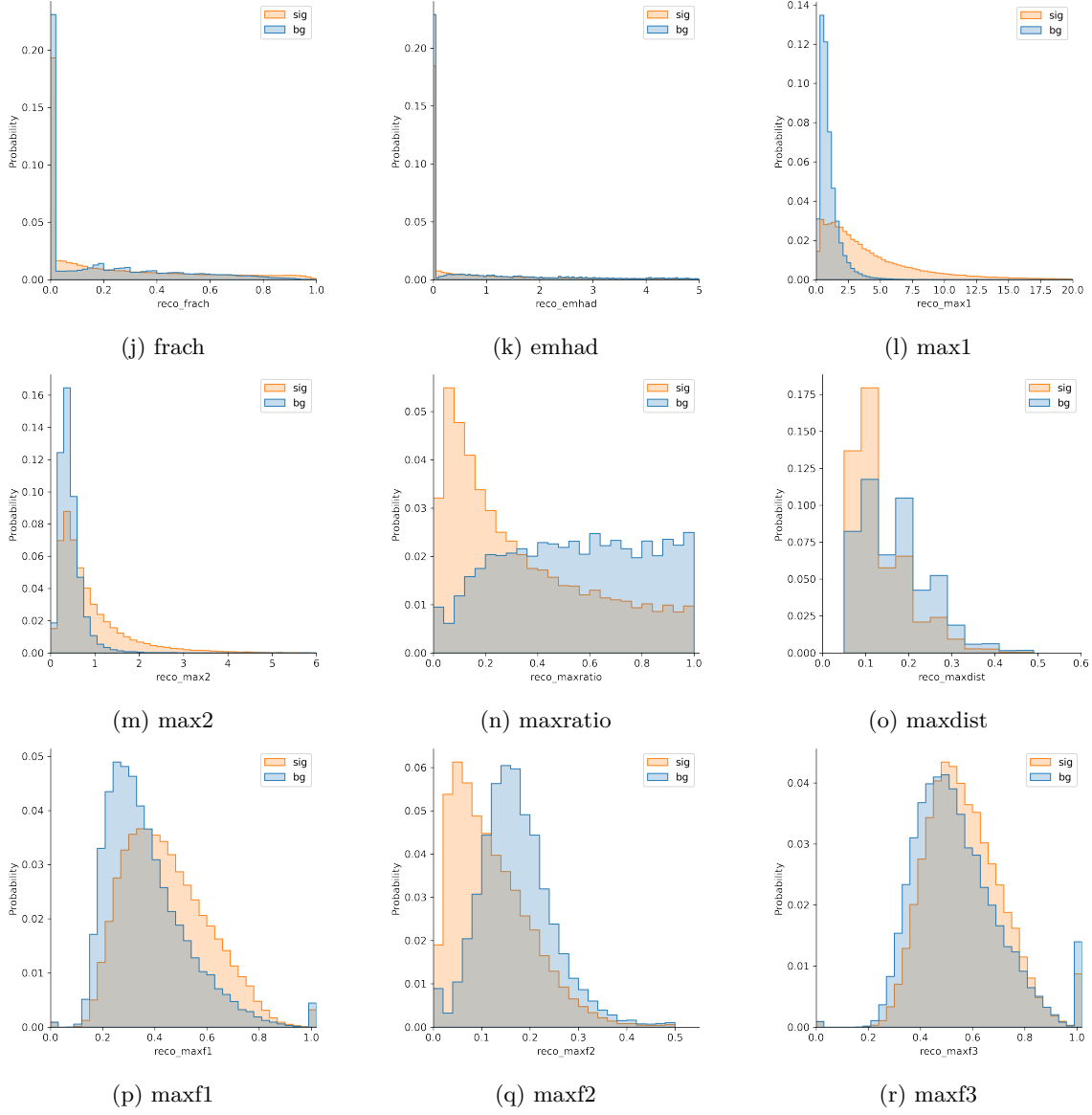


Figure 6.5: Signal and background distributions for all 24 high level features.

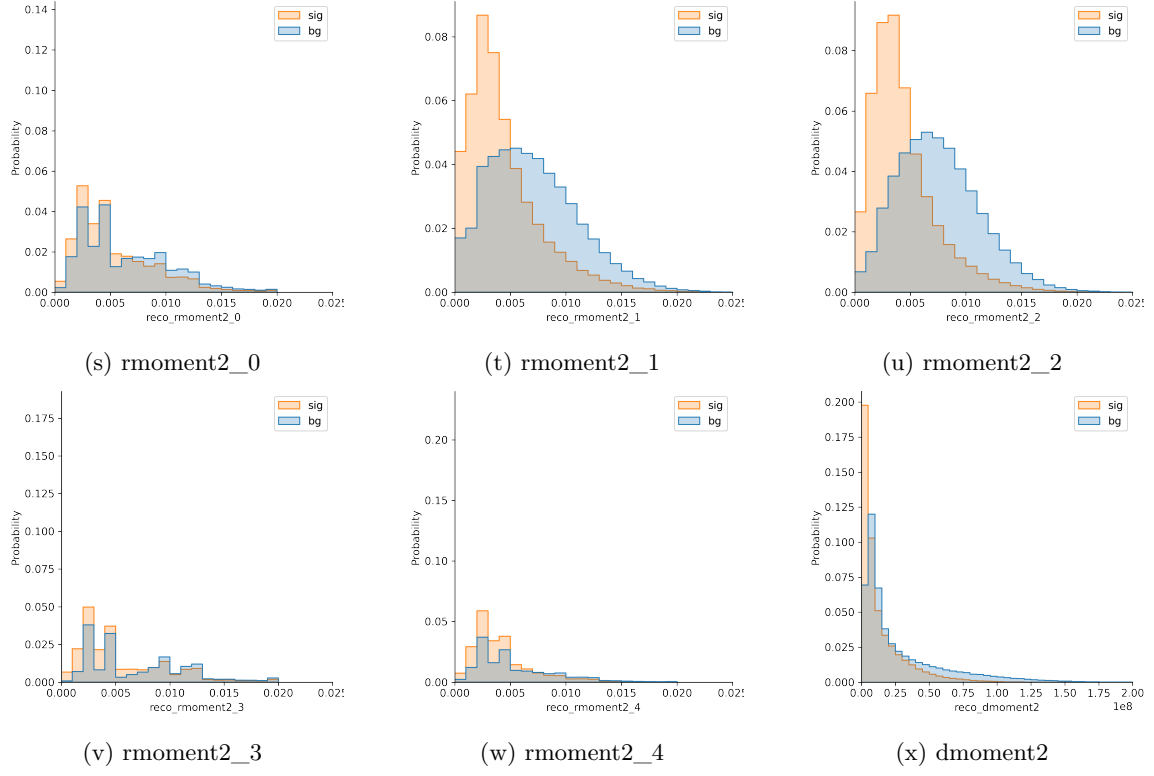


Figure 6.5: Signal and background distributions for all 24 high level features.

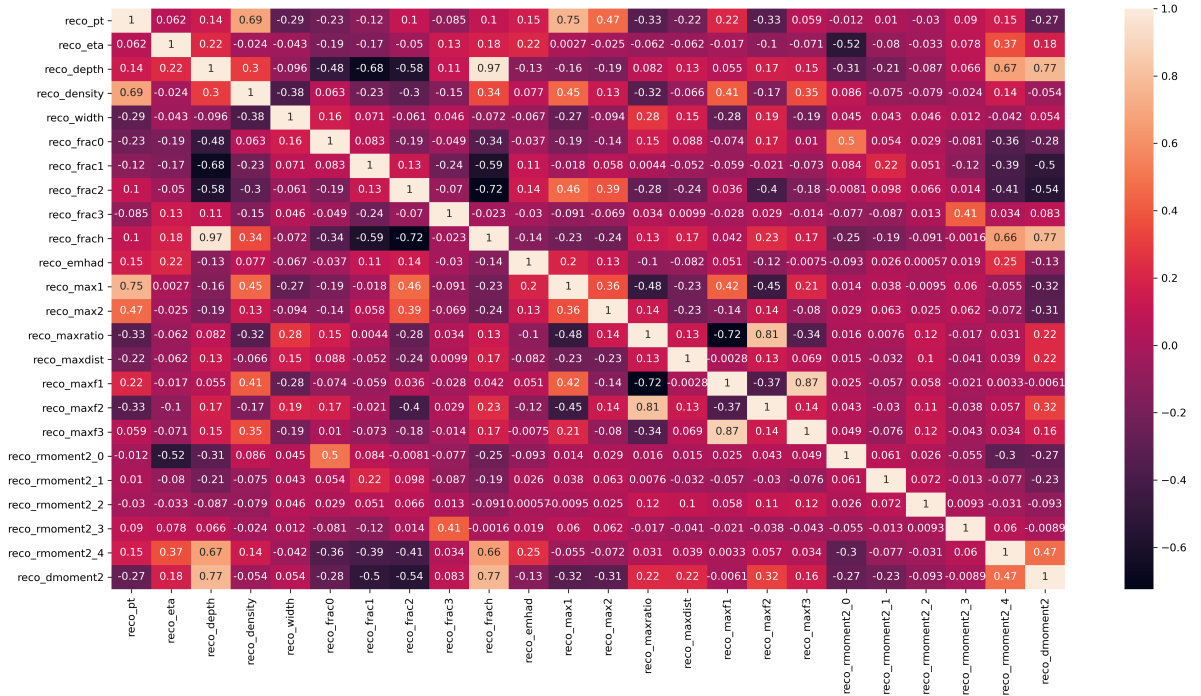


Figure 6.6: Feature correlation heatmap.

6.3 Training and Evaluation

As can be seen in the distributions, many of the features separate fairly well between signal and background while not being too correlated. This supports the notion that training a tree based model combining all these features will produce a powerful discriminator. Motivated by this, a GBDT model was trained producing the ROC (Receiver Operating Characteristic) curve that can be seen in Fig. 6.7. The model was trained using the Scikit-learn [41] framework with the hyperparameters: `min_samples_split = 12000`, `min_samples_leaf = 50`, `max_features = 'sqrt'`, `max_depth=4`, `learning_rate=0.1`, `n_estimators=100`.

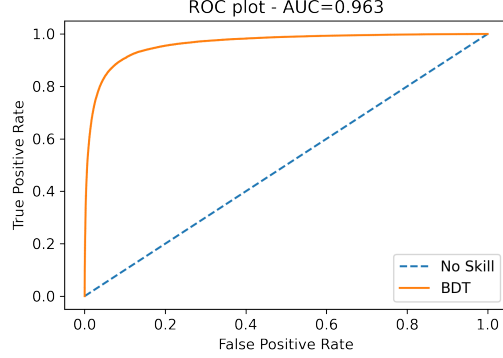


Figure 6.7: ROC curve for the GBDT model with an area under curve of 0.963.

6.4 Turn-on Curve

The ROC curve or sometimes the Precision-Recall curves are used to visually represent the quality of a model and its performance for different cutoffs. For triggering purposes a different representation comes in handy. Assuming a function $f(x)$, a threshold T and an observation x_i , the binary classifier $C(f(x), x_i, T)$ can be defined:

$$C(f(x), x_i, T) = \mathcal{H}(f(x_i) - T)$$

where $\mathcal{H}$ is the Heaviside function. Since $C(f(x), x_i, T)$ is dependant on a threshold T it should by definition fulfill the following condition:

$$\forall T_1, T_2 \in \mathbb{R}, T_1 < T_2 : C(f(x), x_i, T_1) \geq C(f(x), x_i, T_2)$$

An obvious example for $f(x)$ is a GBDT model that being applied to observation x_i returns a probability $0 \leq p \leq 1$. This probability is then compared via the classifier $C(f(x), x_i, T)$ to a threshold $0 \leq T_{gbdt} \leq 1$ producing a predicted class (1 or 0) for observation x_i . Another simple example for such a model is a model that uses an energy cutoff. $f(x)$ returns the total energy of x which can be expressed as $E_x \geq 0$. If the energy is greater than some threshold $T_E \geq 0$, i.e., $E_x \geq T_E$, than x is classified via $C(f(x), x_i, T_E)$ as signal ('1') and otherwise as background ('0'). All such classifiers C can be represented on a turn-on curve.

Since an event must be either saved in its entirety or alternatively entirely discarded, triggering must be applied at the event level. For event i containing many TOBs (observations) $x_{i,j}$, the event will be classified as signal if even a single TOB is classified as such. Thus, event i is classified as

follows:

$$\bigvee_j C(f(x), x_{i,j}, T)$$

Where $\bigvee$ is the disjunction (logical OR) symbol and f and T a chosen model and threshold. Since $C(f(x), x_i, T)$ always returns either 1 or 0 this expression too will return either 1 or 0 predicting the event class. Assuming N_{bg} background events in a dataset, the false positive ratio for that dataset is defined as:

$$FPR = \frac{\sum_{i=1}^{N_{bg}} \bigvee_j C(f(x), x_{i,j}, T)}{N_{bg}}$$

I.e. the fraction of background events that were incorrectly classified as signal.

Since the trigger electronics were not upgraded for Phase-I, any tau trigger algorithm must have the same background rate as Run 2 (see 4.7). Since the expression for the FPR is a monotonically decreasing function in T , $f(x)$ is bounded, and $-\infty \leq T \leq \infty$, there will always exist a specific $T = T_{bg}$ which, up to a discretization error, fulfills:

$$FPR(f) = R_{bg}$$

Where R_{bg} is the false positive ratio of Run 2.

Finally, the turn-on curve is defined as a histogram where each bin displays the efficiency, or true positive ratio (TPR) for its respective energy range. The total efficiency or TPR can be defined in a very similar manner to the FPR:

$$TPR = \frac{\sum_{i=1}^{N_{sig}} \bigvee_j C(f(x), x_{i,j}, T)}{N_{sig}}$$

Where f and T are a chosen model and threshold and N_{sig} the number of signal events in the dataset. After dividing into histogram bins for the turn-on curve, the efficiency for bin $[E_K, E_{K+1}]$:

$$TPR = \frac{\sum_{i \in N_K} \bigvee_j C(f(x), x_{i,j}, T)}{N_{sig}}$$

Where N_K is the set of signal events that fulfill $E_K \leq E_{N_K} < E_{K+1}$. For the GBDT model trained in the previous section the threshold T_{gbdt} that fulfills the condition $FPR = R_{bg}$ was found. Using this value, a turn-on curve with the GBDT model and the Run 2 model for reference can be seen in Fig. 6.8.

The results show that despite the higher pileup of Run 3, the GBDT outperforms the original Run 2 algorithm across the entire energy range. In addition, as the Run 3 implementation of the Run 2 algorithm is inferior than its predecessor, the improvement in performance offered by the GBDT is greater than shown. For each algorithm the area under curve (AUC) is shown which in

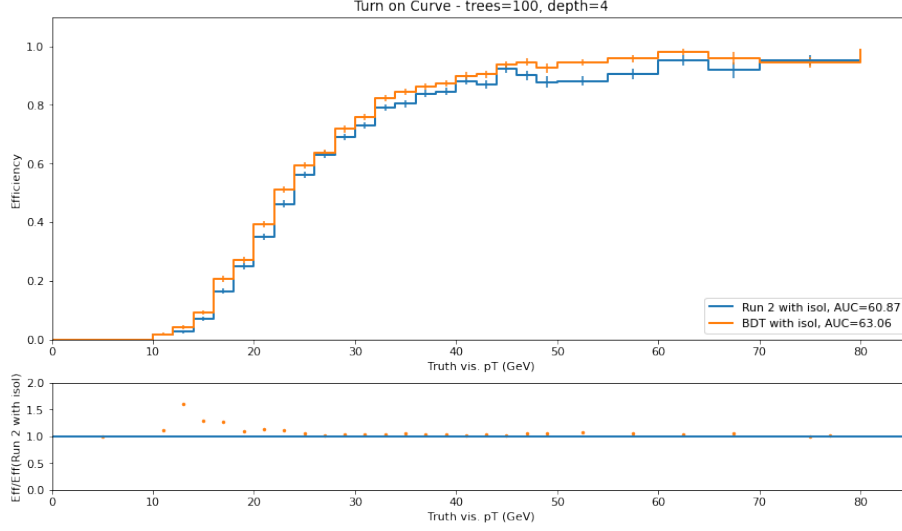


Figure 6.8: Turn on curve for the GBDT and Run 2 showing the signal efficiency for each energy bin. Both models have the the same FPR which is the Run 2 background rate R_{bg} . The GBDT achieves a higher efficiency than Run 2 across the entire range of energies.

this case is the simple integral of the area under the efficiency curve.

6.5 Model Optimization

Typically, one of many standard metrics is chosen and subsequently used to evaluate the success of the machine learning model. This metric can be accuracy, precision, recall, F1-score, or ROC-AUC score, among others. This metric is usually chosen based on both the model requirements as well as the distribution of the data (e.g. for an imbalanced dataset accuracy is an inappropriate metric). Since this chosen metric defines the success of the trained model, optimization schemes such as grid search and Bayesian optimization [42] are designed to find the hyperparameters that maximize this metric. Although these standard ‘out of the box’ metrics are usually sufficient, they do not perform well when the success of the model is determined by the turn-on curve. The reason for this is that while each bin in the turn-on curve has equal importance, the number of samples (TOBs) in each bin is vastly different. Since all the aforementioned metrics maximize global values of the model and don’t maximize each energy range individually, they fail to successfully optimize the turn-on curve. Specifically, since there are many more samples in the lower energies as can be seen in Fig. 6.2, these tend to be negligibly improved in the turn-on curve at a much heavier cost to the higher energy bins. Naively, the samples could be reweighted by their energies to allow a one-to-one mapping between one of the standard metrics and turn-on curve optimization. Unfortunately, as

can be seen in Fig. 6.9, the true energy of the taus which is used in the turn-on curve cannot be easily reconstructed from the total TOB energy. As such, the chosen metric for the Bayesian Optimization was the AUC parameter of the model in the turn on curve i.e. the integral of the area under the efficiency curve. This definition differs from the standard definition of ROC AUC which must obtain a value in the range $[0, 1]$. Thus, using this turn-on curve specific definition of AUC, the GBDT model's hyperparameters were optimized using Bayesian optimization with this metric.

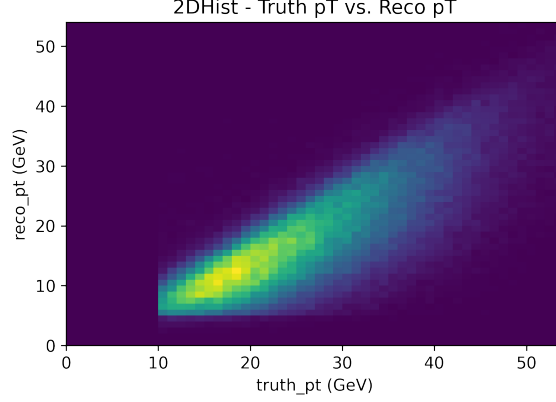


Figure 6.9: Two dimensional histogram describing the distribution of TOBs with respective truth pT and reco pT. It can be seen that although the energies are correlated, the reconstructed energy is systematically overestimated.

CHAPTER

7

GBDT ON FPGAS

7.1 Conifer

As mentioned in 5.1 the tau trigger algorithm must run on the eFEX FPGAs. Since this is the case, the GBDT model trained and optimized in 6.3 must be translated from Python to VHDL. While it is theoretically possible for a hardware engineer to manually translate the model to VHDL, this would take a lot of time and resources. Fortunately, there exist Python packages that already do the translation. One of these, developed specifically for high energy physics, is the HLS4ML package [43]. Building upon HLS4ML the Conifer package was developed offering support for TMVA (Toolkit for MultiVariate data Analysis) [44], Scikit-learn[45], and XGBoost[46]. Although the model for this thesis was originally developed in TMVA, at the time Conifer did not offer support for TMVA and therefore the model development framework was switched to Scikit-learn.

Conifer offers two methods of model translation from Python to hardware language. The first is a direct translation to VHDL. The second is done in two stages. It first translates the model to HLS (High Level Synthesis) code. It can then be converted and programmed to the FPGA with tools such as Vivado HLS or its newer version Vitis HLS. In this thesis the direct approach is chosen as it yields more resource efficient models.

7.2 FPGA Resources and Performance

In addition to providing good discrimination between signal and background, the GBDT model must also obey the requirement to fit within the resource constraints of the eFEX FPGAs. While for models involving multiplication the most critical resource is the number of DSPs, for GBDT models the most critical resources are the number of LUTs and flip flops. Of the two, it turns out that the LUTs tend to be the bottleneck and it is that number that must be reasonable when converting a model from Python. Currently, the eFEX FPGAs use about 55% of their total LUTs [47] leaving a maximum of 45% for the tau trigger algorithm. Since eight instances of the algorithm must run in parallel, then at least naively, each instance can use up to about 5%.

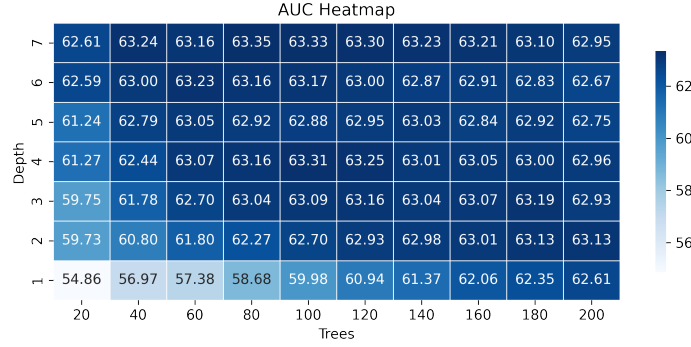
An additional choice that must be made when running Conifer is the number of bits allotted to each variable. Firstly, all variables were centered around zero and normalized to consume a minimal number of bits. While this is standard procedure for neural networks, it is not usually necessary for GBDT models run on the software level. In this case though, it was necessary to minimize FPGA resource usage. After normalizing the GBDT variables, different numbers of bits were chosen to represent the variables. After trial and error it was found that 18 bits offers optimal performance and the VHDL version of the GBDT achieves the same AUC as the Scikit-learn version. Under 15 bits the loss in resolution of the variables is already noticeable and a severe performance degradation can be seen.

7.3 Optimization for FPGA

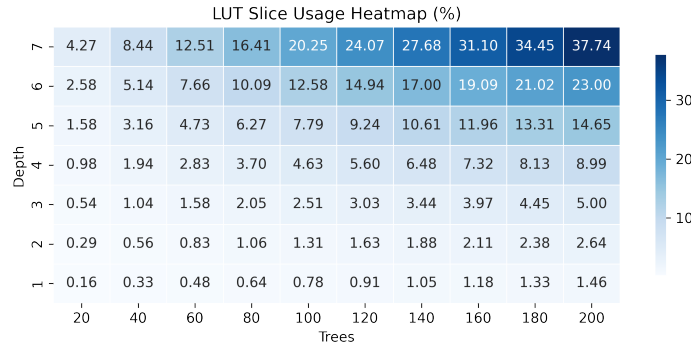
The model trained in 6.3 was optimized to maximize the AUC without taking the size of the model into account. Naturally, once the model is also optimized for its FPGA resource requirements as well as AUC it becomes a trade off between the two. As explained in the HLS4ML paper, the two GBDT hyperparameters that affect the LUT usage of the model are ‘max_depth’ and ‘n_estimators’ - the maximal depth of each tree as well as the total number of trees.

To this end, a grid search was performed over the space of depth and trees. The depth parameter was run from 1-7 and the number of trees from 20-200 in jumps of 20. For each permutation of these parameters a model was trained using them. That model was then translated to VHDL using Conifer and then the LUT usage was extracted from Vivado HLS. The results of the grid search can be seen in Fig. 7.1a and Fig. 7.1b showing for different depth and tree combinations both the resulting AUC of the model and the LUT usage in percent of the FPGA’s total LUTs. As expected, as the number of depth and trees increases so does the total LUT usage. Although the model AUC initially increases with the increase in depth and trees, it eventually enters the overfitting regime with negligible to no improvement in AUC despite incurring increasing costs in terms of LUT usage.

To sum up the two plots in a simple manner, Fig. 7.2 was plotted to show the best possible AUC



(a) Area Under Curve score of the resulting turn-on curve for different tree and depth combinations.



(b) LUT slice usage of the model (in percent) for different tree and depth combinations.

Figure 7.1: AUC and LUT usage for different models showing the trade-off between better predictive ability and higher resource usage.

achievable given a percent of the FPGA's total LUTs. As can be seen, models using under 0.6% offer inferior or equivalent performance to the Run 2 algorithm. Significant improvement is achieved at the 1.6% mark. Although there is further improvement in model AUC beyond the 3% shown in the graph, at this point the improvement is limited and probably not worth the cost incurred in FPGA resources. This allows for the best possible model to be chosen while still maintaining the resource requirements of the FPGA.

Finally, it is worth noting that attempts to reduce the total number of variables in the GBDT only had limited impact on the LUT usage. Reduction from 24 variables to five reduced the LUT usage by a factor of about 10%. Because of this, these attempts were abandoned in favor of tweaking the model hyperparameters.

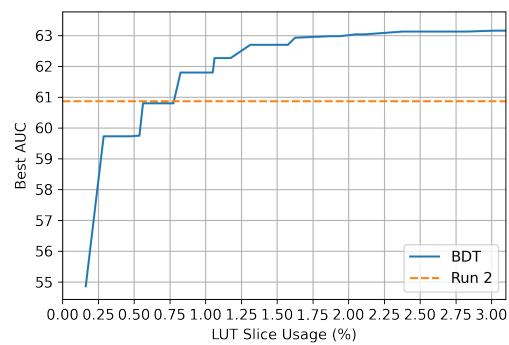


Figure 7.2: Best possible AUC given the available percentage of the FPGA’s total LUT slices. The models improve as the resources available grow. Run 2 is given for reference.

CHAPTER

8

SUMMARY AND OUTLOOK

In this thesis, a new tau trigger algorithm was developed using GBDTs which achieved a significantly better turn-on curve than its default Run 3 counterpart. This improvement over the much simpler method is a result of GBDTs' ability to learn and generalize from the physically meaningful high level features generated from the simulated data. These features were crafted to achieve maximal separation between tau lepton signatures and those of the dominant background.

Although the GBDT outperforms the Run 3 algorithm at the software level, it was necessary to show that it is also feasible to be implemented under the harsh timing and resource constraints imposed by both the trigger hardware and the 25 ns bunch spacing of the LHC. As was shown in Chapter 7, the GBDT model resource requirements can be heavily reduced with minimal impact on performance. This was done by optimizing the hyperparameters over the search space of models with differing resource allocations. The timing constraint on the other hand, is as of writing, not yet met. It will hopefully be fulfilled in the coming months by the engineering team. If this effort succeeds, the algorithm will be integrated into the ATLAS Trigger System by the end of the year thus improving the identification of tau leptons at trigger level.

In addition to the benefits provided by the immediate improvement of trigger level tau lepton identification, this work also serves as an excellent foundation for the integration of machine learning models into future ATLAS tau triggers as well as to other triggers at the LHC. Although in Run 3

the limited computing power offered by the hardware necessitated the use of simpler models such as GBDT, the coming upgrades will not suffer from these limitations and the use of more powerful, deep learning based methods will become increasingly viable.

An additional issue that plagued this project was the lack of available libraries to compile models into VHDL in order to run on the trigger hardware. With the upsurge in the use of deep learning models across varied hardware platforms, generating these models to run on hardware is becoming more widely available and more standardized. This, along with the increase in trigger compute power and the loosening of latency requirements, will open new avenues for the tau trigger algorithm design that are not currently feasible.

Naively, without these constraints, CNN based models such as ResNet can likely outperform the generated GBDT model. Modeling the three-dimensional structure of the calorimeter hits as a graph, with the nodes representing the cell hits and the vertices the distances between them, can be used to train a Deep Set model that should perform better than a CNN based model. Finally, since the point in space of the tau decay forms a secondary vertex from which the resulting tau jet originates, it is possible to extract the respective shower vertex as well as its center of mass via principal component analysis [48]. Since the fundamental jet structure remains the same irrespective of the location of the secondary vertex and of rotations around the shower axis, a representation of the jet that captures this three-dimensional invariance simplifies the subsequent classification stage. SO(3) equivariant networks [49] do just this by mapping points to a latent space where the general orientation is not represented. This method can be used to then create a classifier that learns from the three-dimensional structure of the jet and as such has the potential to further outperform the above methods.

BIBLIOGRAPHY

- [1] Zyla, P A, et al., “Review of particle physics, 2020-2021. rpp,” *PTEP*, vol. 2020, no. 8, p. 2093, 2020. [Online]. Available: <https://cds.cern.ch/record/2729066>
- [2] E. Lopienska, “The cern accelerator complex, layout in 2022,” 2022, [Online; accessed May 20, 2022]. [Online]. Available: <https://cds.cern.ch/images/CERN-GRAPHICS-2022-001-1>
- [3] HiLumi Collaboration, CERN, “Lhc/ hl-lhc plan,” 2022, [Online; accessed May 20, 2022]. [Online]. Available: https://hilumilhc.web.cern.ch/sites/default/files/HL-LHC_Janvier2022.pdf
- [4] J. Pequenaio, “Computer Generated image of the ATLAS calorimeter,” Mar 2008, [Online; accessed May 20, 2022]. [Online]. Available: <https://cds.cern.ch/record/1095927>
- [5] N. Nikiforou, “Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade,” CERN, Tech. Rep., Jun 2013. [Online]. Available: <http://cds.cern.ch/record/1558820>
- [6] The ATLAS Experiment - Public Results, “Functional diagram of the atlas trigger and data acquisition system in run 2 showing expected peak rates and bandwidths through each component.” 2017, [Online; accessed May 20, 2022]. [Online]. Available: <https://twiki.cern.ch/twiki/pub/AtlasPublic/ApprovedPlotsDAQ/tdaqFullNew2017.png>

- [7] M. C. Aleksa, W. P. Cleland, Y. T. Enari, M. V. Fincke-Keeler, L. C. Hervás, F. B. Lanni, S. O. Majewski, C. V. Marino, and I. L. Wingerter-Seez, “ATLAS Liquid Argon Calorimeter Phase-I Upgrade: Technical Design Report,” CERN, Tech. Rep., Sep 2013, final version presented to December 2013 LHCC. [Online]. Available: <https://cds.cern.ch/record/1602230>
- [8] R. Schwienhorst, “The phase-1 upgrade of the ATLAS first level calorimeter trigger,” *Journal of Instrumentation*, vol. 11, no. 01, pp. C01018–C01018, jan 2016. [Online]. Available: <https://doi.org/10.1088/1748-0221/11/01/c01018>
- [9] The ATLAS Collaboration, “Performance of the atlas trigger system in 2015,” *The European Physical Journal C*, vol. 77, no. 317, may 2017. [Online]. Available: <https://doi.org/10.1140/epjc/s10052-017-4852-3>
- [10] S. L. Glashow, “Partial-symmetries of weak interactions,” *Nuclear Physics*, vol. 22, no. 4, pp. 579–588, 1961. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0029558261904692>
- [11] S. Weinberg, “A model of leptons,” *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>
- [12] A. Salam, “Weak and Electromagnetic Interactions,” *Conf. Proc. C*, vol. 680519, pp. 367–377, 1968.
- [13] M. Gell-Mann, “A Schematic Model of Baryons and Mesons,” *Phys. Lett.*, vol. 8, pp. 214–215, 1964.
- [14] M. Mulders, Ed., *Beyond the Standard Model Lectures at the 2013 European School of High Energy Physics*, CERN. Geneva: CERN, Sep 2013. [Online]. Available: <https://cds.cern.ch/record/1522578>
- [15] K. Olive, “Review of particle physics,” *Chinese Physics C*, vol. 40, no. 10, p. 100001, oct 2016. [Online]. Available: <https://doi.org/10.1088/1674-1137/40/10/100001>
- [16] The ATLAS Collaboration, “Evidence for the higgs-boson yukawa coupling to tau leptons with the ATLAS detector,” *Journal of High Energy Physics*, vol. 2015, no. 4, apr 2015. [Online]. Available: <https://doi.org/10.1007%2Fjhep04%282015%29117>
- [17] The ATLAS Collaboration, “Search for charged higgs bosons produced in association with a top quark and decaying via $h^\pm \rightarrow \tau\nu$ using pp collision data recorded at $\sqrt{s} = 13$ tev by the ATLAS detector,” *Physics Letters B*, vol. 759, pp. 555–574, aug 2016. [Online]. Available: <https://doi.org/10.1016/j.physletb.2016.06.017>

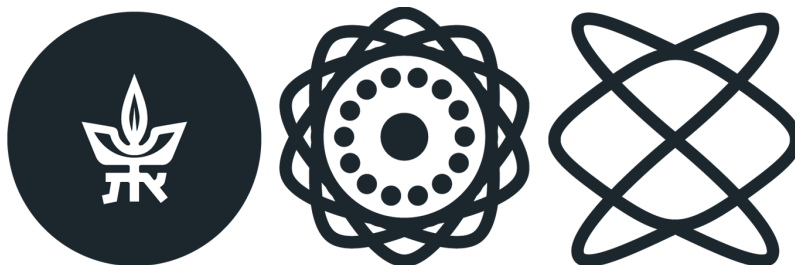
- [18] The ATLAS Collaboration, “Search for squarks and gluinos in events with hadronically decaying tau leptons, jets and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ tev recorded with the ATLAS detector,” *The European Physical Journal C*, vol. 76, no. 12, dec 2016. [Online]. Available: <https://doi.org/10.1140/epjc/s10052-016-4481-2>
- [19] L. Evans and P. Bryant, “LHC machine,” *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 001–S08 001, aug 2008. [Online]. Available: <https://doi.org/10.1088/1748-0221/3/08/s08001>
- [20] C. De Melis, “The CERN accelerator complex. Complexe des accélérateurs du CERN,” Jan 2016. [Online]. Available: <https://cds.cern.ch/record/2119882>
- [21] The ATLAS Collaboration, “The ATLAS experiment at the CERN large hadron collider,” *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 003–S08 003, aug 2008. [Online]. Available: <https://doi.org/10.1088/1748-0221/3/08/s08003>
- [22] The CMS Collaboration, “The CMS experiment at the CERN LHC,” *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 004–S08 004, aug 2008. [Online]. Available: <https://doi.org/10.1088/1748-0221/3/08/s08004>
- [23] The LHCb Collaboration, “The LHCb detector at the LHC,” *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 005–S08 005, aug 2008. [Online]. Available: <https://doi.org/10.1088/1748-0221/3/08/s08005>
- [24] The ALICE Collaboration, “The ALICE experiment at the CERN LHC,” *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 002–S08 002, aug 2008. [Online]. Available: <https://doi.org/10.1088/1748-0221/3/08/s08002>
- [25] F. Pastore, “ATLAS Run-2 status and performance,” CERN, Geneva, Tech. Rep., Sep 2015. [Online]. Available: <http://cds.cern.ch/record/2048973>
- [26] M. E. Pozo Astigarraga, “THE ATLAS DATA ACQUISITION SYSTEM IN LHC RUN 2,” CERN, Geneva, Tech. Rep., Nov 2017. [Online]. Available: <https://cds.cern.ch/record/2292434>
- [27] E. Kay, “Commissioning the atlas liquid argon calorimeter phase-i upgrade,” in *EPJ Web of Conferences*, vol. 253. EDP Sciences, 2021, p. 01007.
- [28] E. A. Smith, “The phase-1 upgrade of the ATLAS level-1 calorimeter trigger,” CERN, Geneva, Tech. Rep., Oct 2021. [Online]. Available: <https://cds.cern.ch/record/2789104>
- [29] W. Qian, “Design and test performance of the ATLAS Feature Extractor trigger boards for the Phase-1 Upgrade,” CERN, Geneva, Tech. Rep., Nov 2016. [Online]. Available: <http://cds.cern.ch/record/2230003>

- [30] Xilinx, Inc, “7 series fpgas data sheet: Overview,” 9 2020, datasheet 180 (v2.6.1). [Online]. Available: https://www.xilinx.com/support/documentation/data_sheets/ds180_7Series_Overview.pdf
- [31] The ATLAS Collaboration, *ATLAS level-1 trigger: Technical Design Report*, ser. Technical design report. ATLAS. Geneva: CERN, 1998. [Online]. Available: <https://cds.cern.ch/record/381429>
- [32] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2015. [Online]. Available: <https://arxiv.org/abs/1512.03385>
- [34] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, “Deep sets,” in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf>
- [35] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001. [Online]. Available: <https://doi.org/10.1023/A:1010933404324>
- [36] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S002200009791504X>
- [37] J. H. Friedman, “Stochastic gradient boosting,” *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.
- [38] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’16. New York, NY, USA: Association for Computing Machinery, 2016, p. 785–794. [Online]. Available: <https://doi.org/10.1145/2939672.2939785>
- [39] The ATLAS Collaboration, “The ATLAS simulation infrastructure,” *The European Physical Journal C*, vol. 70, no. 3, pp. 823–874, sep 2010. [Online]. Available: <https://arxiv.org/abs/1005.4568>
- [40] Agostinelli S. et al, “Geant4—a simulation toolkit,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003. [Online]. Available: <https://infoscience.epfl.ch/record/49909>

- [41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, “Scikit-learn: Machine learning in python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [42] J. Mockus, *Bayesian Approach to Global Optimization: Theory and Applications*, ser. Mathematics and its Applications. Springer Netherlands, 2012. [Online]. Available: <https://books.google.co.il/books?id=VuKoCAAAQBAJ>
- [43] S. Summers, G. D. Guglielmo, J. Duarte, P. Harris, D. Hoang, S. Jindariani, E. Kreinar, V. Loncar, J. Ngadiuba, M. Pierini, D. Rankin, N. Tran, and Z. Wu, “Fast inference of boosted decision trees in FPGAs for particle physics,” *Journal of Instrumentation*, vol. 15, no. 05, pp. P05 026–P05 026, may 2020. [Online]. Available: <https://doi.org/10.1088/1748-0221/15/05/p05026>
- [44] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov *et al.*, “Tmva-toolkit for multivariate data analysis,” *arXiv preprint physics/0703039*, 2007.
- [45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [46] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’16. New York, NY, USA: ACM, 2016, pp. 785–794. [Online]. Available: <http://doi.acm.org/10.1145/2939672.2939785>
- [47] atlas-l1calo efex, “efexfirmware,” <https://gitlab.cern.ch/atlas-l1calo-efex/eFEXFirmware/-/releases/v0.15.0>, 2022.
- [48] T. Barillari, E. Bergeaas Kuutmann, T. Carli, J. Erdmann, P. Giovannini, K. J. Grahm, C. Issever, A. Jantsch, A. Kiryunin, K. Lohwasser, A. Maslennikov, S. Menke, H. Oberlack, G. Pospelov, E. Rauter, P. Schacht, F. Spanó, P. Speckmayer, P. Stavina, and P. Strízenec, “Local Hadronic Calibration,” CERN, Geneva, Tech. Rep., Jun 2008. [Online]. Available: <https://cds.cern.ch/record/1112035>
- [49] C. Deng, O. Litany, Y. Duan, A. Poulenard, A. Tagliasacchi, and L. Guibas, “Vector neurons: A general framework for so(3)-equivariant networks,” 2021. [Online]. Available: <https://arxiv.org/abs/2104.12229>

תקציר

ה-LHC, מאיץ החלקיקים ב-CERN, מייצר כמות אדירה של מידע בכל רגע. בשניה אחת יש כ- 10^{34} התנגשויות בין פרוטונים לסמ"ר בגלאי ATLAS. הטכנולוגיה כיום אינה מסוגלת להתמודד עם כמות כזו של מידע, וכך נוצר הצורך במערכת הטריגר. מערכת הטריגר ב-ATLAS מחליטה אלו אירועי התנגשות מכילים פיזיקה מעניינת ואלו לא, כך שרק אירועי ההתנגשות המעניינים, המהווים חלק קטן מאד מסך האירועים, יישמרו לדיסק. תזה זו עוסקת במערכת הטריגר של חלקיקי טאו, שהיא מערכת שמטרתה להבחין בזמן אמת האם התנגשות נתונה מכילה אות שמתאים לזה של לפטון טאו. עד כתיבת תזה זו, הטריגר התבסס על סף אנרגיה מינימלי והצליח בצורה סבירה לזהות את חלקיקי הטאו. מכיוון שמערכת הטריגר נתונה למגבלות זמן שהיה מחמירות וכן מתוכנתת על חומרה דלת משאבים, תכנון שיטת הבחנה עם ביצועים טובים יותר מהשיטה הקיימת הוא מאתגר. תזה זו מציעה שיטה חדשנית המותאמת לחומרה הקיימת ומבוססת למידת מכונה אשר משפרת את ביצועי מערכת הטריגר לחלקיקי טאו ב-ATLAS. שימוש בשיטה חדשה זו תוביל להגברת קצב שמירת ההתנגשויות המכילות לפטוני טאו שנחוצים למגוון רחב של אנליזות.



School of Physics and Astronomy **בית הספר לפיזיקה ולאסטרונומיה**
The Raymond and Beverly Sackler **הפקולטה למדעים מדויקים**
Faculty of Exact Sciences **ע"ש ריימונד ובברלי סאקלר**
Tel Aviv University **אוניברסיטת תל אביב**

עצי החלטה מחוזקים לזיהוי חלקיקי טאו בקלורימטר רמה 1 ב-ATLAS

אוריאל בר-און

תזת מחקר עבור תואר מגיסטר למדעים

נכתבה בהנחייתה של:

ד"ר לירון ברק

תמוז התשפ"ב