University of Alberta

BDT-Based Single Tau Selection Algorithm for the Level-2 ATLAS Hadronic Tau Trigger

by

Andrew Karamaoun

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Physics

Andrew Karamaoun
Fall 2013
Edmonton, AB

Abstract

We present a Boosted Decision Tree (BDT) based algorithm, with promising potential, that can be used at the level-2 stage of the ATLAS Hadronic Tau Trigger for the selection of a single hadronically decaying tau lepton. The results indicate that our algorithm outperforms the current algorithm on a wide variety of measures. The BDT trigger adds $(8.242 \pm 0.068 \pm 0.321)\%$ to the $Z \to \tau\tau$ signal RoI-wise acceptance rate of the reference trigger and adds $(10.628 \pm 0.069 \pm 0.276)\%$ to the $Z \to \tau\tau$ signal event-wise acceptance rate of the reference trigger. Not only does the BDT trigger outperform the current trigger for Standard Model signal channels, but it also dominates the current trigger for yet undiscovered high mass resonances that will be more accessible in the upcoming (higher energy) running period.

Acknowledgements

I would first like to thank my supervisor, Roger, for his help in reviewing this thesis as it progressed from a rough draft to a final copy. His countless hours spent helping me with questions are greatly appreciated.

Secondly, I would like to thank my professors at the University of Alberta and elsewhere. Their courses provided me with the background needed to even begin this thesis.

Next, I would like to thank all of the members of the ATLAS collaboration who have helped me while working on the ATLAS experiment. This includes Siva, Dugan, Francesc, and members of the Tau Trigger Working Group: Mansoora, Soshi, Philip.

I would also like to thank my family members: my mother, my father, my brother, and my sister. Their continued support helped me get through the most challenging days.

Next, I would like to thank my friends in Edmonton and elsewhere, including Kamil, Nemanja, Sebastian, Alexander, Kingsley, Colin, Pranshu, and Karl. It was a pleasure sharing interests with each of you.

Finally, I would like to thank Nigina for her support.

Contents

List of Tables

List of Figures

Nomenclature

0.1 Mathematics

We define $\mathbb{N} = \{0, 1, 2, 3, \dots\}$, $\mathbb{Z}_+ = \{1, 2, 3, \dots\}$, $\mathbb{Z}_{+n} = \{1, 2, 3, \dots, n\}$, $\mathbb{R}$ as the set of real numbers, and $\mathbb{C}$ as the set of complex numbers.

The cartesian product of sets $A_1, \dots, A_n$ is denoted by $A_1 \times \cdots \times A_n$. If, for some set A, $A_i = A \; \forall i \in \mathbb{Z}_{+n}$, then the product may be simply written as A^n.

A Schwartz class function on $\mathbb{R}^n$ is an element of

$$\left\{ f \in C^\infty(\mathbb{R}^n) : \sup_{x \in \mathbb{R}^n} \left| x^\alpha \partial^\beta f(x) \right| < \infty \; \forall \alpha, \beta \in \mathbb{N}^m \; \forall m \in \mathbb{Z}_+ \right\},$$

where we use the multi-index notation: $x^\alpha \coloneqq x_1^{\alpha_1} \dots x_n^{\alpha_n}$ and $\partial^\beta \coloneqq \partial_1^{\beta_1} \dots \partial_n^{\beta_n}$.

For each $n \in \mathbb{Z}_+$, we define the unitary group $U(n)$ as the set of unitary $n \times n$ matrices with entries from $\mathbb{C}$, the special unitary group $SU(n)$ as the subset of $U(n)$ consisting of the matrices with determinant 1, and $O(n)$ as the subset of $U(n)$ consisting of the orthogonal matrices.

The symbol $\coloneqq$ denotes an equality that is also a definition, i.e. it doesn't follow as a consequence of our assumptions. It is only used when we believe that it would not be clearly implicit from the context.

0.2 Particle Physics

Unless otherwise stated, we work in units where $\hbar = c = 1$. This means that the unit for energy is the electron volt (eV), the unit for momentum is eV/c, and the unit for mass is

eV/c^2. For some figures taken from outside sources, the unit for momentum or mass may simply be written as eV without the c, and thus, the inclusion of c or c^2 is implicit.

If we regard $\mathbb{R}^4$ as spacetime, then its members consist of four-tuples (x_0, x_1, x_2, x_3), where x_0 is the time coordinate and x_1, x_2, x_3 are the spatial coordinates. We typically denote such a four-tuple as x, and $\boldsymbol{x}$ denotes the three-tuple of spatial coordinates (x_1, x_2, x_3). The four spacetime coordinates are indexed with lowercase greek letters such as μ, ν, ρ. A term containing two such indentical indices, with both either lower or upper indices, is implicitly implied to be a summation over the indices. If one of the (identical) indices, μ, is a lower index and the other one is an upper index, then this again implies a summation over the indices, but with each term in the summation multiplied by $g_{\mu\mu}$, with $g_{00} \coloneqq -1$ and $g_{\mu\mu} \coloneqq 1$ otherwise. If x is a member of spactime, then, x^2 is an abbreviation for $x_\mu x^\mu$. Similarly, if p is the four-momentum of a particle, then $\boldsymbol{p}$ denotes its three-momentum, and p^2 is an abbreviation for $p_\mu p^\mu$. For any spacetime coordinate index μ, the partial derivative with respect to the μ^{th} component of spacetime is written as ∂_μ or ∂^μ.

The Fock Space $\mathcal{F}(\mathcal{H})$ for a separable Hilbert Space $\mathcal{H}$ is defined in Section A.1.

Feynman Diagrams (or the Feynman calculus) are explained in Section A.5.

When decribing an elementary particle process, a pair of brackets may be placed to the right of the symbol for a given particle A, and within the brackets, a number denotes the mass of A in GeV/c^2, and a symbol of another particle B denotes that B is one of the decay products of A. The generic symbol l stands for lepton, and h stands for hadron. For example,

$$H(125) \to \tau(h)\tau(\mu) \tag{1}$$

describes a Higgs boson with a mass of $125\,\text{GeV/c}^2$ decaying into two taus, with one tau decaying hadronically, and the other tau decaying into a muon and neutrinos.

0.3 The ATLAS Detector

The z-axis points along the direction of the beamline, in the counter-clockwise direction if seen from above. The angle ϕ is measured around the z-axis. θ is the angle from

the z-axis. The pseudorapidity η is defined as $-\ln\tan(\theta/2)$. The transverse plane is the plane transverse to the beam axis. E_T and p_T denote the transverse energy and momentum, respectively. ΔR, a measure of the distance between regions of interest, is given as $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. The longitudinal impact parameter z_0 of a track is defined as the z coordinate of the point on the track that is closest to the beamline. The pile-up or average number of interactions per bunch crossing for a specified period of time is denoted by μ. Cross sections are typically measured in *barns*, denoted with the symbol b, and defined by $1\,\mathrm{b} = 10^{-28}\,\mathrm{m}^2$. [1][2]

Chapter 1

Introduction

In order to probe smaller distance scales and learn more about the substructure of the universe, it is commonly accepted (via the uncertainty principle) that we must probe higher levels of energy. This may lead to the discovery of new interactions that occur at the smaller distance scales or simply the creation of new particles that were previously undetected due to their large mass. For example, protons, previously believed to be elementary particles that interact via the so-called nuclear force, were discovered, through high energy collisions, to be composed of more elementary constituent particles called up and down quarks that interact via a new force known as the strong force. And it was not until higher energy collisions were studied that we discovered the existence of new quarks that behave in essentially the same way as the other quarks, but are more massive.

The Large Hadron Collider (LHC) is currently at the leading edge of the energy frontier, with a maximum proton-proton collision energy of 14 TeV and an average design rate of one billion collisions per second. After two years of running at about half of its design energy, it has produced enough experimental data to confirm the discovery of a new particle, most likely believed to be the Higgs boson, one of the most significant missing pieces of the Standard Model (SM) that unifies two of the elementary forces and explains how the SM particles obtain mass. On the horizon, we have supersymmetry, a relatively standard theory requiring the existence of new particles that are expected to be within the reach of the LHC

in 2015, when the collision energy is to be increased to its design level. Other theories beyond the SM are being tested as well, including extra dimensions, mini black holes, and dark matter. One of the main goals of this thesis is to efficiently make sense of the tremendous amount of information afforded to us by this machine.

We begin, in Chapter 2, by reviewing the theoretical foundations of particle physics. A high-level overview of the Standard Model is presented, with more details given in Appendix A. Next, an entire section is devoted to deriving the Higgs Mechanism. Lastly, we give an introduction to supersymmetry and explain why it is an important extension to the Standard Model that should be included in experimental searches. We believe that these sections should be completely accessible for those who are comfortable with quantum mechanics and mathematics at the upper-undergraduate level.

In the next two chapters we describe the experimental framework used for our research. This includes the LHC, the ATLAS detector, and the tau trigger. We introduce the tau lepton and motivate the desire to understand elementary interactions with taus. We give a short performance report of the tau trigger in 2011, including a summary of a study we performed on dijet events selected with the tau trigger. Next, we explain the challenges that were faced for the tau trigger in 2012, and describe the improvements done to meet those challenges. We end Chapter 4 with a short outlook for the future of the tau trigger, including potential improvements to the level-2 selection process.

In Chapter 5 we introduce the problem we wish to solve and present our procedure. Our goal is to design an algorithm for single tau selection at the level-2 tau trigger. We begin by describing boosted decision trees and the specific tools that we use to create them. Our procedure is then described in the steps: data preparation, training, testing. Next, we mention the currently implemented algorithm that we intend to replace, and show how it compares with ours. The final three sections describe some of the additional steps that were added to our procedure in order to improve the performance of our algorithm as well as the confidence in our results; namely: error codes, systematic uncertainties, and optimization of variable selection.

Chapter 6 shows the results. We begin by listing the specific datasets that were used for training, and provide a motivation for their inclusion. Then, we state our cut parameters for the initial stage of the selection algorithm, and show the energy dependent acceptance rates for the samples used for training. Next, we state the variables used together with their motivation, and compare the signal and background distributions for these variables. We also show the respective correlation matrices. The next section describes the error code classes that were used, and gives an indication of the fraction of collision data events that belong to each error code class. The next two sections show some of the results used to validate the training procedure; namely the overtraining check and the sequence of decision trees that define the classifier. Next, we have the results for the calorimeter stage of the algorithm, when applied to the training samples. This includes both the overall rates and the pile-up dependent ones. For the remaining sections, we focus on the overall algorithm, rather than just the calorimeter stage. We show results for signal acceptance versus overall rejection, for various classes of error codes. For each training sample, we examine the acceptance rates as functions of transverse energy and pile-up. Finally, we end the chapter by showing the energy and pile-up dependent rates for a variety of other signal and background samples, none of which were used in the training procedure.

The final chapter is the conclusion. Here we summarize the content of this thesis and restate the essential aspects of our results. We also provide suggestions for future extensions of our study.

The nomenclature used is shown before this chapter. The references are listed after the final chapter. We attempt to provide publicly available sources whenever possible, but may provide some sources that are restricted to ATLAS members, as they may be valuable to those engaged in similar studies.

Chapter 2

Theoretical Foundations

2.1 The Standard Model

Elementary particle physics, the study of the fundamental interactions between the fundamental constituents of matter, arose in 1897 with J. J. Thompson's discovery of a negatively charged particle called the electron. By 1911, (due to E. Rutherford) the atomic nucleus was known to be a particle with positive charge with a relatively large separation distance from its corresponding electrons. The model was further refined by the discovery (by J. Chadwick), in 1932, of the neutron, so that the nucleus was viewed as consisting of both positively charged particles called protons and neutral particles called neutrons. In parallel with these discoveries, A. Einstein argued in 1905 that electromagnetic radiation is composed of tiny particles contributing indivisible amounts of energy, and this was confirmed in 1923 by an experminent conducted by A. H. Compton, which is now known as the discovery of the photon. In the years that followed, more elementary particles were discovered, the interactions were revised, and what once were called elementary particles became known to be composites of more elementary constituent particles. Currently, the SM of particle physics consists of three fundamental interactions between 61 species of elementary particles. The two main categories for these particles are fermions and bosons. The fermions obey the *Pauli exclusion principle* stating that no two such particles of the same species (quantum

numbers) can occupy the same state at the same time, and are commonly known as the matter forming particles. By the Spin-Statistics Theorem, any particle with half integer spin is a fermion. However, experimentally, the only observed elementary fermions have spin 1/2, and there are 48 of them. The bosons, on the other hand, are the particles that are not restricted by the Pauli exclusion principle and an arbitrary number of such particles of the same species can occupy the same state simultaneously. They are commonly known as the force carrying particles, and by the Spin-Statistics Theorem, must have integer spin. Experimentally, there are 13 known elementary bosons, and they all have spin 1, except for the recently experimentally confirmed Higgs boson, which has spin 0 [3]. The fermions can be further divided into two groups: leptons and quarks. The leptons consist of the electron, the muon, and the tau, together with their corresponding antiparticles, neutrinos, and antineutrinos. While the electron, muon, and tau are electrically charged, the neutrinos are electrically neutral and only participate in the weak interaction, and the difference between a neutrino and an antineutrino is that the former has positive lepton number while the latter has negative lepton number. The quarks are the strongly interacting particles of types: up, down, charm, strange, top, bottom, where each type consists of three particles of different colour charge and their corresponding antiparticles with opposite electric and colour charge. A particle composed of quarks is called a hadron, and can be either a meson (a quark-antiquark pair), a baryon (three quarks), or an antibaryon (three antiquarks). Finally, the bosons are: the photon, the eight gluons, the Z boson, the two W bosons, and the Higgs boson. The photon mediates the electromagnetic interaction, the gluon mediates the strong interaction, the Z boson and the W bosons mediate the weak interaction, and the Higgs boson provides mass to all of the particles. Of course, gravity is missing from this picture, and that is because we currently do not have any viable theory of gravity that is able to function on the small distance scales present in elementary particle physics. Figure 2.1 shows a table of the elementary particles contained within the SM.

The precise description of the SM is formalized through the language of quantum field theory. In the canonical approach, we start with the classical notion of a field, i.e. a differ-

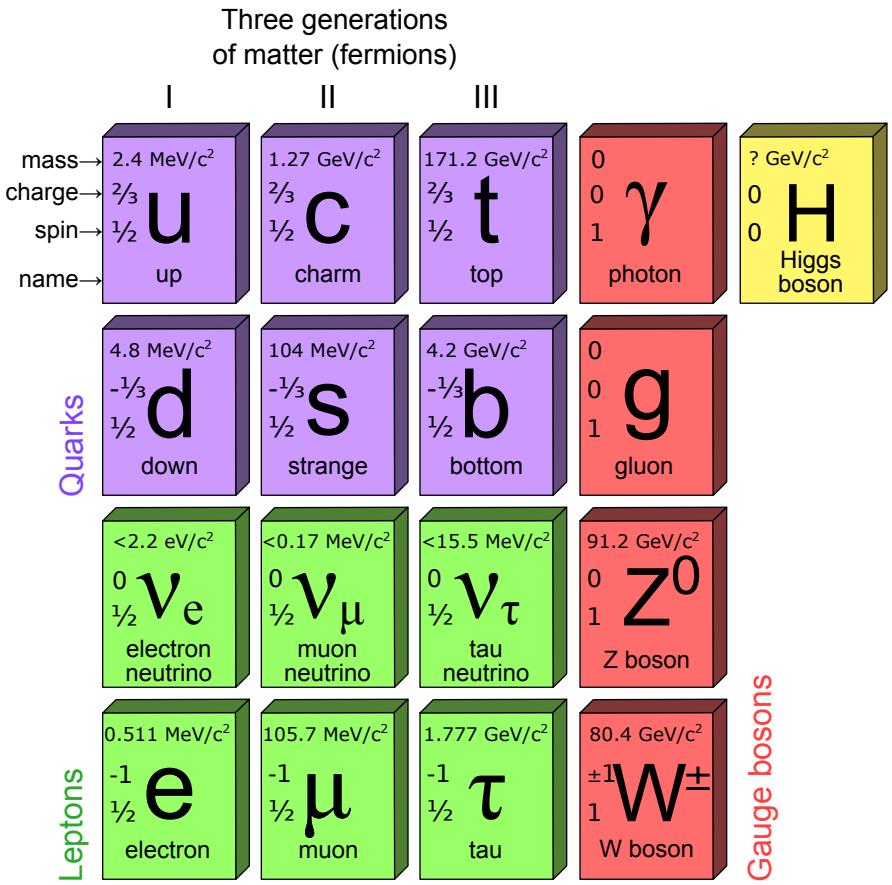


Figure 2.1: The elementary particles in the SM of Particle Physics [4]. Recent experimental evidence suggests a SM Higgs boson in the mass range of 125-127 $\mathrm{GeV/c^2}$ [5].

entiable $\mathbb{C}^n$-valued function φ defined at each point in spacetime $\mathbb{R}^4$, and depending on the particle we wish to describe, we choose an appropriate value of n and free field equation describing the time dependent evolution of the field given an initial value for the field at time $t = 0$. We then expand $\varphi(x)$ as a Fourier integral and replace the Fourier coefficients with linear combinations of annihilation and creation operators indexed by $\boldsymbol{p} \in \mathbb{R}^3$ that act on the Fock state space $\mathcal{F}(\mathcal{H})$ of the field's quanta, for a separable infinite-dimensional Hilbert Space $\mathcal{H}$, with the motivation being that each coefficient represents the time dependent displacement of a harmonic oscillator with angular frequency $\omega_{\boldsymbol{p}} = \sqrt{m^2 + \boldsymbol{p}^2}$, which corresponds to the quantum mechanical position operator $X_{\boldsymbol{p}} = \frac{1}{\sqrt{2\omega_{\boldsymbol{p}}}}(A_{\boldsymbol{p}} + A^\dagger_{\boldsymbol{p}})$, where $A_{\boldsymbol{p}}$ can be identified as an operator that annihilates a particle with momentum $\boldsymbol{p}$ and $A^\dagger_{\boldsymbol{p}}$ as

an operator that creates a particle with momentum $\boldsymbol{p}$. Since the integral may now be divergent, we instead regard φ as an operator-valued *distribution* that maps each real-valued Schwartz class function f on $\mathbb{R}^4$ to the operator

$$\int_{\mathbb{R}^4} f(x)\varphi(x)\mathrm{d}^4x\,. \tag{2.1}$$

The details of the construction of the quantum field are shown in Appendix A. Each free field equation has a corresponding free *Lagrangian density* $\mathcal{L}_0$, a polynomial of arbitrary degree in terms of the field φ and $\partial_\mu\varphi$, that generates the free field equation when one applies the generalized Euler-Lagrange equation,

$$\partial^\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi)}\right)=\frac{\partial\mathcal{L}}{\partial\varphi}\,, \tag{2.2}$$

for $\mathcal{L}=\mathcal{L}_0$. This equation is interpreted by regarding φ and $\partial_\mu\varphi$ as real-valued variables of $\mathcal{L}$ when taking a partial derivative with respect to $\mathcal{L}$, regarding the values of the partial derivatives as the μ^{th} components of spacetime $\mathbb{R}^4$, and finally regarding φ as an operator after summing over the indices. If we add interaction terms (functions of φ, other fields $\varphi_1,\ldots,\varphi_n$, and their derivatives) to $\mathcal{L}_0$, describing the interaction between the fields, we obtain the complete Lagrangian density $\mathcal{L}$ for the system. By definition, the *Lagrangian* $L(t)$ is the spatial integral $\int\mathcal{L}(\varphi)(t,\boldsymbol{x})d^3\boldsymbol{x}$. The terms of the Lagrangian involving time derivatives of fields are referred to as the *kinetic energy* T, while the rest is referred to as the negative of the *potential energy* V, so that $L=T-V$. The *Hamiltonian* is defined as $H=T+V$ and is typically decomposed into its free part H_0 and its interaction part H_I, so that $H=H_0+H_I$. From quantum mechanics, we know that the probability amplitude for a system to begin in state $|a\rangle$, and after time t, be observed in state $|b\rangle$ is $\langle b|\,e^{-itH}\,|a\rangle$, where H is the quantum mechanical Hamiltonian. The same principle is used in quantum field theory, but the states are now members of the multi-field Fock space, and the Hamiltonian is the one for quantum fields described above. Typically, in a particle physics experiment, we have particles with localised momenta that begin sufficiently far apart from each other where their interactions are negligible, approach each other, interact, and then the resulting particles move away from each other so that they are again sufficiently far

apart to have negligible interactions. The probability amplitude for such an experiment is written as $\langle b| S |a\rangle$, where $|a\rangle$ is the incoming state, $|b\rangle$ is the outgoing state, and S, called the *scattering operator*, is given by

$$S = \lim_{\substack{t_0 \to -\infty \\ t_1 \to +\infty}} e^{it_1 H_0} e^{-i(t_1 - t_0)H} e^{-it_0 H_0} \,. \tag{2.3}$$

It can be shown that

$$S = I + \sum_{n=1}^{\infty} \frac{1}{i^n n!} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathcal{T}[H_I(\tau_1) \dots H_I(\tau_n)] d\tau_1 \dots d\tau_n \,. \tag{2.4}$$

where

$$\mathcal{T}[H_I(\tau_1) \dots H_I(\tau_n)] \coloneqq H_I(\tau_{i_1}) \dots H_I(\tau_{i_n}) \text{ for } \tau_{i_1} > \tau_{i_2} > \dots > \tau_{i_n} \,. \tag{2.5}$$

Typically, the assumption is that the first few terms of the series for S is a good approximation to the value of S, and the calculation of probability amplitudes in quantum field theory reduces to calculating these terms through the aid of Feynman diagrams, which encapsulate the needed information in a simpler format. Other quantities such as the differential cross section or lifetime of an elementary particle process can be obtained from the corresponding probability amplitudes using straighforward formulas. More details regarding the scattering operator are shown in Section A.4.

Quantum electrodynamics (QED) is the theory that describes the electromagnetic interaction between the elementary particles predicted by the SM. It is the small scale generalisation of the more familiar theory of classical electrodynamics, which consists of charged masses interacting via the electric and magnetic forces. The Lagrangian density for the theory is

$$\mathcal{L}(\psi, A^\mu) = \sum_{j=1}^{N} \psi_j^\dagger \gamma^0 (i\gamma^\mu \partial_\mu - m)\psi_j - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - q_j g_e A_\mu \psi_j^\dagger \gamma^0 \gamma^\mu \psi_j \,, \tag{2.6}$$

where $F^{\mu\nu} \coloneqq \partial^\mu A^\nu - \partial^\nu A^\mu$, γ^μ are P. Dirac's gamma matrices, g_e is the electromagnetic coupling constant, the charged masses now correspond to $N = 21$ types of elementary charged spin-1/2 particles with bispinor fields ψ_j and charges $q_j \in \pm\frac{1}{3}\{1, 2, 3\}$, and the electromagnetic force corresponds to the photon's vector field A^μ. The corresponding free

field equation is called the *Dirac Equation*. The states of the quanta of each ψ_j consist of bispinors, commonly referred to as simply spinors or Dirac spinors, which are four component objects where two of the components represent the state of a particle, and the other two represent the state of the corresponding antiparticle. Figure 2.2 shows a Feynman diagram for a typical QED interaction.

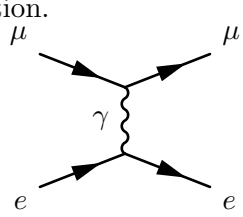


Figure 2.2: A Feynman diagram of an electron interacting with a muon via the photon.

On the other hand, quantum chromodynamics (QCD) is the theory that describes the strong interaction between the elementary particles, which is the force that is well known for holding together atomic nuclei. It is mediated by eight types of gluons, and since gluons, like photons, are massless spin 1 particles, their Feynman propagator is equivalent to that of the photons (modulo the color factor). Gluons are exchanged by each of the 36 quarks, and peculiarly, by gluons themselves. In addition to charge, quarks contain a conserved quantity called colour, which can be one of red, green, blue, anti-red, anti-green, or anti-blue. Gluons may also contain colour, but in colour-anticolour pairs. This allows for colour to be conserved when a quark changes colour at an elementary QCD vertex, as shown in Figure 2.3.

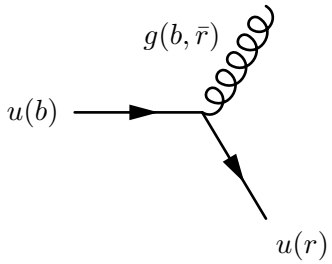


Figure 2.3: A blue up quark converting into a red up quark and emitting a blue-antired gluon.

Unlike QED, where the effective electromagnetic coupling constant decreases as the spatial separation of the particles participating in the interaction increases, the effective

strong coupling constant increases as the spatial separation of the particles participating in the interaction increases. Since the vacuum produces electron-positron pairs, we can view the vacuum as a dielectric material with some dielectric constant ϵ, which tends to give a particle an effective electromagnetic charge of

$$q_{\text{eff}} = q/\epsilon\,, \tag{2.7}$$

where ϵ approaches one as the separation distance approaches zero. Similarly for QCD, the vacuum produces quark-antiquark pairs with opposite colour charge, which has the effect of decreasing the effective strong coupling constant between quarks as the separation distance increases. But since gluons interact with each other, the vacuum produces gluon loops, which have the opposite and dominant effect. What is troubling is that for long enough distances, the coupling constant is greater than one, and so in the QCD Feynman calculus, as the complexity of the Feynman diagram increases, its contribution to the amplitude increases as well, which makes calculating accurate interaction amplitudes much more difficult as compared with QED. Another interesting property of QCD is *colour confinement*, which states that individual quarks/antiquarks cannot be isolated and are only observed as groups of quarks/antiquarks bound together and forming a colour singlet, which is a colourless state consisting of two or more quarks/antiquarks with the same probability for each colour state combination. It has yet to be analytically proven that colour confinement follows as a consequence of QCD, but the explanation is widely believed to be that as two quarks of unlike colour charge separate, the gluons between them form narrow tubes of colour charge, which tend to bring the quarks together, until some point where it is more energetically favourable for a new quark-antiquark pair to spontaneously appear rather than to allow the tube to extend further. The process of hadron formation from quark and gluon interactions is called *hadronisation*. It occurs after high energy particle collisions create free quarks that may exist for a short time, and by colour confinement, combine to form hadrons. The hadrons tend to travel in narrow cones called *jets*. Note that a single quark can be indirectly observed as is the case with the top quark. Due to its short lifetime, it typically decays before it has time to form a hadron with other quarks. Also note that a gluon

cannot exist in a colour singlet state, and this can be seen as a result of the experimentally proved $SU(3)$ symmetry of QCD. This implies that colour singlets cannot exchange gluons to interact strongly between each other, but rather must interact through the exchange of other colour-singlets, and this interaction is called the *residual strong force*, which is typically seen as the force between atomic nuclei. Since the residual strong force is mediated by a massive particle, it has a very short range unlike, for example, the electromagnetic force mediated by the (massless) photon.

Thirdly, we have the weak interaction. It is the fundamental force that is responsible for the radioactive beta decay of atomic nuclei. It is mediated by the W^+,W^- and Z bosons with free field equation

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = 0\,, \tag{2.8}$$

which is equivalent to the photon's free field equation if we set the mass of the boson, $m = 0$. These bosons are exchanged by each of the 48 fermions as well as amongst themselves. While the vertex factor for the electromagnetic force between charged leptons u_a, u_b is $u_a^\dagger i g_e \gamma^0 \gamma^\mu u_b$, the vertex factor for the weak interaction between a charged lepton u_a and a neutrino u_b is $u_a^\dagger \left(\frac{-ig_w}{2\sqrt{2}}\right) \gamma^0\gamma^\mu(1-\gamma^5)u_b$ where g_w is the weak coupling constant. One interesting fact about the weak interaction is that it is not parity invariant. By definition, an interaction is *parity invariant* if the probability of a certain process occuring is the same as the probability of the same process occuring with the spatial coordinates of the particles inverted (take the negative of the spatial coordinates). For if we take

$$x^i \mapsto -x^i\,, \tag{2.9}$$

then by the Dirac Equation, the leptonic state transforms as

$$u(x) \mapsto \gamma^0 u(x)\,, \tag{2.10}$$

so that for leptons with states u_a and u_b,

$$u_a^\dagger \gamma^0 \gamma^\mu u_b \mapsto -u_a^\dagger \gamma^0 \gamma^\mu u_b \tag{2.11}$$

and

$$-u_a^\dagger \gamma^0 \gamma^\mu \gamma^5 u_b \mapsto -u_a^\dagger \gamma^0 \gamma^\mu \gamma^5 u_b \,, \tag{2.12}$$

which shows that the electromagnetic vertex factor is parity invariant, while the leptonic weak vertex factor is not since when you add (2.11) to (2.12), you do not get a factor of unit modulus multiplied by the original interaction amplitude as you do with (2.11), and therefore the probability amplitude is not invariant. Another important concept related to the weak force is *handedness*. A particle is said to be right handed if the component of its spin along its direction of motion is parallel to its direction of motion, and left handed if it is antiparallel to its direction of motion. Experiments in the past have consistently shown neutrinos to be left handed and (actually implied by the Dirac Equation) antineutrinos to be right handed. Note that for massive particles, handedness is not Lorentz invariant, thus if we assume neutrinos keep the same handedness, we must assume that they are massless. It turns out that this assumption is not entirely correct (neutrino oscillations), but it works to a good approximation, and unless otherwise stated, it is the assumption we will make.

Fortunately, the electromagnetic and weak interactions can be unified. The results in Section A.6 show that for a massless particle, it must be left handed in order to participate in weak interactions, and for a massless antiparticle, it must be right handed in order to participate in weak interactions. Thus, according to this model, even if right handed neutrinos existed, they would not interact, so we can assume that a right handed neutrino field does not exist. Further, for $L = \frac{1-\gamma^5}{2}$, $R = \frac{1+\gamma^5}{2}$, $\bar{L} = \frac{1+\gamma^5}{2}$, $\bar{R} = \frac{1-\gamma^5}{2}$, we can decompose u as

$$u = Lu + Ru \tag{2.13}$$

and call $u_L \coloneqq Lu$ the left part of u and $u_R \coloneqq Ru$ the right part of u. We can view u_L and u_R as distinct particles and say that u_L participates in weak interactions while u_R does not. Similarly, for an antiparticle $\bar{u}$, call $\bar{u}_{\bar{L}} \coloneqq \bar{L}\bar{u}$ the left part of $\bar{u}$, $\bar{u}_{\bar{R}} \coloneqq \bar{R}\bar{u}$ the right part of $\bar{u}$, and say that $\bar{u}_R$ participates in weak interactions while $\bar{u}_L$ does not. For any leptons with spinors a, b, this allows us to replace our weak vertex factor

$$a^\dagger g_w \gamma^0 \gamma^\mu (1-\gamma^5) b \tag{2.14}$$

with

$$2a_L^\dagger g_w \gamma^0 \gamma^\mu b_L \,, \tag{2.15}$$

and replace our electromagnetic vertex factor

$$a^\dagger g_e \gamma^0 \gamma^\mu b \tag{2.16}$$

with

$$a_L^\dagger g_e \gamma^0 \gamma^\mu b_L + a_R^\dagger g_e \gamma^0 \gamma^\mu b_R \,. \tag{2.17}$$

This is nice because we can now look at the weak and electromagnetic interactions as having the same vertex factors modulo a coupling constant, except that the weak interaction only acts on the "left" particles. The next section completes the unification through the use of the Higgs Mechanism. [6][7]

2.2 The Higgs Mechanism

The current standard model that unifies QED with weak interactions is the one developed by A. Salam and S. Weinberg in 1967. Assuming for simplicity that the only fermions are electrons with fields denoted by e and electron neutrinos with fields denoted by ν, the model builds upon our previous discussion by introducing a Lagrangian density for free massless left electrons, right electrons and neutrinos:

$$\mathcal{L}_0(\nu, e_L, e_R) = i\nu^\dagger \gamma^0 \gamma^\mu \partial_\mu \nu + i e_L^\dagger \gamma^0 \gamma^\mu \partial_\mu e_L + i e_R^\dagger \gamma^0 \gamma^\mu \partial_\mu e_R \,. \tag{2.18}$$

This Lagrangian density is invariant under the global symmetry group $U(3)$ of unitary 3×3 matrices U acting on (the transpose of) a triple (e_L, ν, e_R). By global we mean that the matrices have no spatial or temporal dependence. Next, a subgroup of $U(3)$ is carefully chosen such that the Lagrangian density is invariant under this subgroup, and if we promote this global symmetry group to a local (spacetime dependent) one, the interaction terms between electrons and neutrinos are obtained.

It turns out that the subgroup used is $SU(2) \times U(1)$, which consists of matrices of the form

$$\begin{pmatrix} e^{-i\theta}a & e^{-i\theta}b & 0 \\ e^{-i\theta}c & e^{-i\theta}d & 0 \\ 0 & 0 & e^{-2i\theta} \end{pmatrix}. \tag{2.19}$$

The details of the motivation for this group are ommited, but the main idea is that $SU(2)$ exploits the symmetry between left electrons and neutrinos under the weak interaction, while $U(1)$ exploits the so called weak hypercharge. The behaviour of strong interactions plays a large role in motivating this construction.

Now, we promote the global symmetry group $SU(2) \times U(1)$ to a local symmetry group. This requires introducing a gauge field

$$\frac{1}{2i}(\mathbf{W}_\mu \cdot \boldsymbol{\sigma}, X_\mu)\,, \tag{2.20}$$

where $\mathbf{W}_\mu = (W_{\mu 1}, W_{\mu 2}, W_{\mu 3})$, $\boldsymbol{\sigma}$ is the triple of Pauli spin matrices $(\sigma_1, \sigma_2, \sigma_3)$, and each $W_{j\mu}$ and X_μ is real valued. This gauge field generates the covariant derivatives:

$$D_\mu \begin{pmatrix} \nu \\ e_L \end{pmatrix} = \left(\partial_\mu + \frac{1}{2i} g \mathbf{W}_\mu \cdot \boldsymbol{\sigma} - \frac{1}{2i} g' X_\mu\right) \begin{pmatrix} \nu \\ e_L \end{pmatrix} \tag{2.21}$$

and

$$D_\mu e_R = \left(\partial_\mu - \frac{1}{i} g' X_\mu\right) e_R\,, \tag{2.22}$$

where g, g' are coupling constants, $\mathbf{W}_\mu \cdot \boldsymbol{\sigma}$ acts on the column vector by matrix multiplication, and ∂_μ and X_μ act componentwise. If we replace the ordinary derivatives in our Lagrangian density (2.18) with the covariant ones, we get the Lagrangian density

$$\mathcal{L}_{e\nu} = \begin{pmatrix} \nu^\dagger & e_L^\dagger \end{pmatrix} \gamma^0 \gamma^\mu \left(i\partial_\mu + \frac{1}{2} g \mathbf{W}_\mu \cdot \boldsymbol{\sigma} - \frac{1}{2} g' X_\mu\right) \begin{pmatrix} \nu \\ e_L \end{pmatrix} + e_R^\dagger \gamma^0 \gamma^\mu (i\partial_\mu - g' X_\mu) e_R\,. \tag{2.23}$$

Since we introduced a new field, we must add its free Lagrangian density

$$\mathcal{L}_F = -\frac{1}{4}\langle F_{\mu\nu} | F^{\mu\nu} \rangle\,, \tag{2.24}$$

where $F_{\mu\nu}$ is the $\mathbb{C}^3 \times \mathbb{C}$ valued function

$$F_{\mu\nu} = (\partial_\mu \mathbf{W}_\nu - \partial_\nu \mathbf{W}_\mu + g\mathbf{W}_\mu \times \mathbf{W}_\nu, \partial_\mu X_\nu - \partial_\nu X_\mu)\,, \tag{2.25}$$

and the inner product is the standard inner product on $\mathbb{C}^4$. Note that the requirement of local invariance, used here and elsewhere, is a fundamental statement about the uniformity of spacetime. The elementary laws of physics should not depend upon the specific time and location.

The Lagrangian density $\mathcal{L}_{e\nu} + \mathcal{L}_F$ now consists of interacting electrons, neutrinos, and gauge fields, but the problem is that they are all massless. To resolve this issue, we add another field to the Lagrangian density that enables mass to be generated by a process known as spontaneous symmetry breaking.

To understand how symmetry breaking works, suppose $\Phi = (\phi_1, ..., \phi_n)$ is an n-tuple of real scalar fields with a Lagrangian density given by

$$\mathcal{L}(\Phi) = \frac{1}{2}(\partial_\mu \Phi) \cdot (\partial^\mu \Phi) - \mathcal{V}(\Phi)\,, \tag{2.26}$$

where $\frac{1}{2}(\partial_\mu \Phi)\cdot(\partial^\mu \Phi)$ is called the kinetic energy density at Φ and $\mathcal{V}(\Phi)$ is called the potential energy density at Φ. If, instead, we view (in the natural way) the domain of $\mathcal{V}$ as the set of all $\boldsymbol{x} \in \mathbb{R}^n$ (not to be confused with position), then $\mathcal{V}$ is the real valued function given by $\mathcal{V}(\boldsymbol{x})$. From now on, the precise choice of $\mathcal{V}$ will be implicit from the context. Typically, as in our previous analysis, $\mathcal{L}$ is such that $\mathcal{V}(\boldsymbol{x})$ is minimized at a unique point $\boldsymbol{x}_0 \in \mathbb{R}^n$. From the Lagrangian density we deduce the vertex and propagator factors that are used to calculate transition rates. To derive these factors, we perform a small perturbation about the vacuum state $|0\rangle$ of the system. But, it can be shown that the vacuum expectation value $\langle 0|\, \Phi(0)\, |0\rangle = \boldsymbol{x}_0$, since we need a stable point for the perturbation series to converge. If there is more than one point that minimizes $\mathcal{V}(\boldsymbol{x})$, then there is no unique choice for our vacuum state, and the interactions we observe between the fields included in the Lagrangian density can change drastically depending on our choice of vacuum state. In principle, even if our system is at a specific vacuum state, given enough energy, we should be able to surpass whatever potential is keeping us from reaching other vacuum states. However, due

to energy limitations in the laboratory, it may be quite difficult for us to observe transitons from one vacuum state to another and see the full picture. Also, our perturbation theory relies on the assumption that the perturbation is small enough that it does not take us to a new vacuum state.

Now suppose that the system has vacuum state $|0_1\rangle$ such that $\langle 0_1| \Phi(0) |0_1\rangle = \boldsymbol{x}_1$, where $\boldsymbol{x}_1$ is not necessarily a minimizer of $V(\boldsymbol{x})$, and $\mathcal{L}(\Phi)$ is invariant under a certain set of transformations $\mathfrak{T}$ acting on the deviations $\Phi - \boldsymbol{x}_1 I$. If $\boldsymbol{x}_2$ is a minimizer of $V(\boldsymbol{x})$, it is not guaranteed that $\mathcal{L}(\Phi)$ is invariant under $\mathfrak{T}$ acting on the deviations $\Phi - \boldsymbol{x}_2 I$. These transformations are commonly referred to as symmetries of the system with vacuum state $|0_1\rangle$. If the system transitions to having a different vacuum state $|0_2\rangle$ where these symmetries are lost, we say that the original symmetries were broken. As mentioned above, it may be possible to regain these symmetries by transitioning back to having vacuum state $|0_1\rangle$, but the cost in energy may be too high to perform in a laboratory, and in our perturbation theory, once we know the system to have vacuum state $|0_2\rangle$, we must assume that the perturbation is small enough that it does not take our system to or past a local maximum of the potential energy.

For example, consider the Lagrangian density given by

$$\mathcal{L}(\Phi) = \frac{1}{2}(\partial_\mu \Phi) \cdot (\partial^\mu \Phi) + 10|\Phi|^2 - |\Phi|^4 \,, \tag{2.27}$$

where $\Phi = (\phi_1, \phi_2)$. For $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$, the potential (energy) is given by

$$\mathcal{V}(\boldsymbol{x}) = -10|\boldsymbol{x}|^2 + |\boldsymbol{x}|^4 \,. \tag{2.28}$$

Figure 2.4 shows an illustration of $\mathcal{V}$.

This potential has a continuous set of minimizers given by the circle:

$$(\sqrt{5}, \theta) \text{ for } \theta \in [0, 2\pi) \text{ in polar coordinates} \,. \tag{2.29}$$

The potential also has an unstable maximizer at $\boldsymbol{x}_1 \coloneqq 0$. Let $\mathfrak{T}$ be the set of transformations T_θ that map

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \mapsto \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \,. \tag{2.30}$$

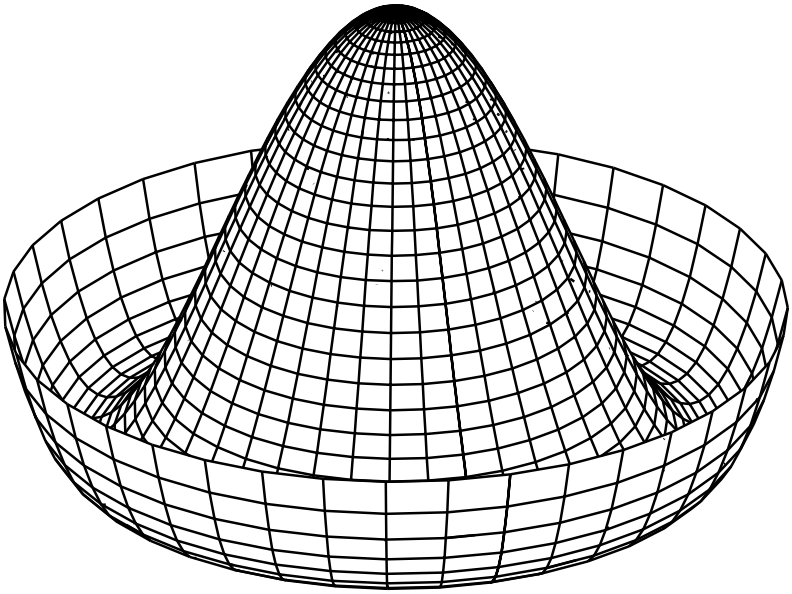


Figure 2.4: An illustration of the Mexican Hat Potential [8].

$\mathfrak{T}$ simply consists of the rotation operators, and it is clear from the figure that $\mathcal{L}$ is invariant under such transformations. In other words, the transformation

$$(\Phi - \boldsymbol{x}_1 I) \mapsto T_\theta(\Phi - \boldsymbol{x}_1 I) \tag{2.31}$$

leaves $\mathcal{L}(\Phi)$ invariant. Thus if the vacuum state of the system is $|0_1\rangle$ and $\langle 0_1| \Phi(0) |0_1\rangle = \boldsymbol{x}_1$, we say that it posseses rotational symmetry. Physically this means that it is equally likely to transition into any direction. Since it is an unstable point, any perturbation of the system will result in a transition to a vacuum state expectation value being one of the minimizers of $\mathcal{V}(\boldsymbol{x})$, and each θ is equally likely. Assume that the system transitioned to having its vacuum expectation value equal to one of these minimizers, $\boldsymbol{x}_2$. Then clearly, the system no longer posseses rotational symmetry, or more formally,

$$(\Phi - \boldsymbol{x}_2 I) \mapsto T_\theta(\Phi - \boldsymbol{x}_2 I) \tag{2.32}$$

does not leave $\mathcal{L}(\Phi)$ invariant. This example illustrates the idea of spontaneous symmetry breaking, where a symmetry is broken in a nondeterministic fashion by simply perturbing it in the slightest possible way. Note that in order to measure the state of a system, we

must perturb it, thus we can think of spontaneous symmetry breaking as the collapsing of a wavefunction into a ground state that does not posses the symmetries of the inital wavefunction.

Now, the question is: What does spontaneous symmetry breaking have to do with generating mass? To answer this question, we use the example Lagrangian density (2.27), choose the minimizer $\boldsymbol{x}_0 = (\sqrt{5}, 0)$, and let $\tilde{\Phi} = \Phi - \boldsymbol{x}_0 I$. Then,

$$\begin{aligned}
\mathcal{L}(\Phi) &= \mathcal{L}(\boldsymbol{x}_0 I + \tilde{\Phi}) \\
&= \frac{1}{2}(\partial_\mu \tilde{\Phi}) \cdot (\partial^\mu \tilde{\Phi}) + 10((\sqrt{5} + \tilde{\phi_1})^2 + \tilde{\phi_2}^2) - ((\sqrt{5} + \tilde{\phi_1})^2 + \tilde{\phi_2}^2)^2 \\
&= \frac{1}{2}(\partial_\mu \tilde{\Phi}) \cdot (\partial^\mu \tilde{\Phi}) + 25 - 30\tilde{\phi_1}^2 - 4\sqrt{5}\tilde{\phi_1}^3 - \tilde{\phi_1}^4 - \tilde{\phi_2}^4 - 4\sqrt{5}\tilde{\phi_1}\tilde{\phi_2}^2 - 2\tilde{\phi_1}\tilde{\phi_2}^2 \\
&\quad - 2\tilde{\phi_1}^2\tilde{\phi_2}^2 \\
&= (\frac{1}{2}(\partial_\mu \tilde{\phi_1})(\partial^\mu \tilde{\phi_1}) - 30\tilde{\phi_1}^2) + (\frac{1}{2}(\partial_\mu \tilde{\phi_2})(\partial^\mu \tilde{\phi_2})) \\
&\quad + \text{interactions between } \tilde{\phi_1} \text{ and } \tilde{\phi_2}\,. \qquad (2.33)
\end{aligned}$$

In general, if m_i is the mass of a real-valued field ϕ_i, then the ϕ_i^2 term of the corresponding Lagrangian density must be $-\frac{1}{2}m_i^2$. Thus, we read from above that $\tilde{\phi_1}$ has mass $m_1 = \sqrt{10}$, and $\tilde{\phi_2} = \phi_2$ has mass $m_2 = 0$.

Moreover, it can be shown that for any Lagrangian density of the form (2.26), if it is invariant under a global symmetry group $G \subset O(n)$, (continuous) sponataneous symmetry breaking leads to massless scalar particles called Goldstone bosons [7].

But, we want to generate mass, not massless particles. This is where local symmetry invariance comes in. We will illustrate the concept with the example, but the results are easily generalizable. First, reinterpret the Lagrangian density (2.27) so that its argument is one complex field rather than two real fields:

$$\mathcal{L}(\Phi) = \frac{1}{2}(\partial_\mu \Phi)^*(\partial^\mu \Phi) + 10\Phi^*\Phi - (\Phi^*\Phi)^2\,. \qquad (2.34)$$

The rotational symmetry under the matrices given by (2.30) now becomes invariance under the transformation

$$\Phi \mapsto e^{i\theta}\Phi\,. \qquad (2.35)$$

We promote this transformation to a local one by allowing θ to have a spacetime dependence:

$$\Phi(x) \mapsto e^{i\theta(x)}\Phi(x)\,. \tag{2.36}$$

To make the Lagrangian density invariant under such a transformation we introduce a massless gauge field A^μ and replace the derivatives in the Lagrangian density with covariant ones. This results in the modified Lagrangian density:

$$\mathcal{L}(\Phi) = \frac{1}{2}((\partial_\mu - iA_\mu)\Phi)^*((\partial^\mu + iA^\mu)\Phi) + 10\Phi^*\Phi - (\Phi^*\Phi)^2 - \frac{1}{4}\langle F_{\mu\nu}|F^{\mu\nu}\rangle\,. \tag{2.37}$$

As before we let $\tilde{\Phi} = \Phi - x_0 I$ for $x_0 = \sqrt{5}$. Then the only difference comes from the first term:

$$\begin{aligned}&\frac{1}{2}\left[(\partial_\mu - iA_\mu)\Phi\right]^*\left[(\partial^\mu + iA^\mu)\Phi\right]\\ &= \frac{1}{2}\left[(\partial_\mu - iA_\mu)(x_0 I + \tilde{\Phi})\right]^*\left[(\partial^\mu + iA^\mu)(x_0 I + \tilde{\Phi})\right]\,,\end{aligned} \tag{2.38}$$

which includes a term

$$\frac{1}{2}(x_0^2)A_\mu A^\mu \tag{2.39}$$

that implies that the spin-1 field A^μ is massive with mass $x_0 = \sqrt{5}$. There is also a term

$$-2ix_0(\partial_\mu\phi_2)A^\mu \tag{2.40}$$

for which it does not make sense to interpret ϕ_2 as a fundamental particle of the theory. Actually, it is possible to exploit our local symmetry to remove ϕ_2. Rewrite (2.36) as

$$[\phi_1(x) + i\phi_2(x)] \mapsto [\cos\theta(x) + i\sin\theta(x)]\,[\phi_1(x) + i\phi_2(x)]\,. \tag{2.41}$$

By choosing $\theta(x) = -\tan^{-1}(\phi_2(x)/\phi_1(x))$, we see that the transformed $\Phi(x)$ becomes real and $\phi_2(x) = 0$.

Let us summarize the results of symmetry breaking. If the Lagrangian density is globally, but not locally, invariant under a symmetry group $G \subset O(n)$, spontaneous symmetry breaking creates massless spin 0 particles called Goldstone bosons. If we add massless vector fields to the Lagrangian density to make it locally invariant under the symmetry group,

spontaneous symmetry breaking creates massive spin 1 particles instead of the Goldstone bosons. We say that "the gauge fields eat the Goldstone bosons and become massive", and since the massive vector fields have three degrees of freedom compared to the massless ones that have two, spontaneous symmetry breaking in effect conserves the degrees of freedom of the Goldstone bosons by passing them to the gauge fields.

Now that we understand spontaneous symmetry breaking and how it can generate mass, we will add a field Φ to our electroweak Lagrangian density that acts to break the symmetry. Let

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1^+ + i\phi_2^+ \\ \phi_1^0 + i\phi_2^0 \end{pmatrix} . \tag{2.42}$$

Note that Φ is a doublet of complex-valued scalar fields. It is commonly known as the *Higgs doublet*. The (locally invariant) Lagragian density for Φ is

$$\mathcal{L}_\Phi = \frac{1}{2}(D_\mu \Phi)^\dagger (D^\mu \Phi) + \kappa \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2 , \tag{2.43}$$

where κ, λ are positive constants and

$$D_\mu = \partial_\mu + \frac{1}{2i} g \mathbf{W}_\mu \cdot \sigma + \frac{1}{2i} g' X_\mu . \tag{2.44}$$

Finally, we add an interaction between Φ and the leptons:

$$\mathcal{L}_{e\nu\Phi} = -G \begin{pmatrix} \nu^\dagger \gamma^0 & e_L^\dagger \gamma^0 \end{pmatrix} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e_R - G e_R^\dagger \gamma^0 \begin{pmatrix} \phi^{+*} & \phi^{0*} \end{pmatrix} \begin{pmatrix} \nu \\ e_L \end{pmatrix} , \tag{2.45}$$

where G is a coupling constant. The full Lagrangian density for the Salam-Weinberg Electroweak model (including only electrons and neutrinos as the leptons) is:

$$\mathcal{L} = \mathcal{L}_{e\nu} + \mathcal{L}_F + \mathcal{L}_\Phi + \mathcal{L}_{e\nu\Phi} . \tag{2.46}$$

We will now interpret the effect of this addition to the Lagrangian density. First note that for $\boldsymbol{x} \in \mathbb{C}^2$, the potential of $\mathcal{L}_\Phi$,

$$\mathcal{V}(\boldsymbol{x}) = -\kappa \boldsymbol{x}^\dagger \boldsymbol{x} + \lambda (\boldsymbol{x}^\dagger \boldsymbol{x})^2 , \tag{2.47}$$

is minimized when

$$|x^+|^2 + |x^0|^2 = \kappa / 2\lambda . \tag{2.48}$$

Among the $\boldsymbol{x}$ satisfying this condition, we take

$$x^+ = 0, x^0 = a\,, \tag{2.49}$$

where $a = \sqrt{\frac{\kappa}{2\lambda}}$. As explained in [7], we can perform a gauge transformation $g(x)$ that puts any field with values near $\binom{0}{aI}$ into the form

$$\Phi = \begin{pmatrix} 0 \\ aI + \phi \end{pmatrix}\,, \tag{2.50}$$

where ϕ is a real scalar field. Further, there is a unique $g(x)$ such that $g(x) \in SU(2)\ \forall x$ and this determines the field ϕ known as the Higgs field. Its quanta are (spin-0) particles known as Higgs bosons. Henceforth, we assume this choice of gauge. The covariant derivative

$$D_\mu\Phi = \begin{pmatrix} 0 \\ \partial_\mu\phi \end{pmatrix} + (\frac{g}{2i}\begin{pmatrix} W_{\mu 3} & W_{\mu 1} - iW_{\mu 2} \\ W_{\mu 1} + iW_{\mu 2} & -W_{\mu 3} \end{pmatrix} + \frac{g'}{2i}X_\mu)\begin{pmatrix} 0 \\ a + \phi \end{pmatrix} \tag{2.51}$$

gives

$$\begin{aligned} &(D_\mu\Phi)^\dagger D^\mu\Phi \\ &= \partial_\mu\phi\partial^\mu\phi - \frac{a^2g^2}{4}(W_{\mu 1} + iW_{\mu 2})(W_{\mu 1} - iW_{\mu 2}) - \frac{a^2}{4}(gW_{\mu 3} - g'X_\mu)(gW_3^\mu - g'X^\mu) + \dots\,, \end{aligned} \tag{2.52}$$

where the dots denote terms containing products of three or four fields. Also,

$$\kappa\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2 = \frac{\kappa^2}{4\lambda} - 2\kappa\phi^2 + \dots\,. \tag{2.53}$$

Thus, after removing constant terms,

$$\mathcal{L}_\Phi = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{a^2g^2}{4}W_\mu^\dagger W^\mu - \frac{a^2}{8}(g^2 + g'^2)Z_\mu Z^\mu + \dots\,, \tag{2.54}$$

where

$$W = \frac{W_1 + iW_2}{\sqrt{2}}, Z = \frac{gW_3 - g'X}{\sqrt{g^2 + g'^2}}, A = \frac{g'W_3 + gX}{\sqrt{g^2 + g'^2}} \tag{2.55}$$

are deduced to be distinct particles by the way they combine in the Lagrangian density. The complex value of the W field implies that the W particle is charged. The masses of the particles are read to be

$$m_W = \frac{ag}{2}, m_Z = \frac{a\sqrt{g^2 + g'^2}}{2}, m_\phi = 4\kappa\,. \tag{2.56}$$

The A field remains massless. We now examine the interaction between Φ and the leptons. Applying our gauge condition, we get

$$\mathcal{L}_{e\nu\Phi} = -Ga(e_L^\dagger \gamma^0 e_R + e_R^\dagger \gamma^0 e_L) - G(\phi e_L^\dagger \gamma^0 e_R + \phi e_R^\dagger \gamma^0 e_L)\,. \tag{2.57}$$

The first pair of terms is a mass term for electrons. The mass m_e of an electron should appear in the Lagrangian density as $-m\psi_e^\dagger \gamma^0 \psi_e$ where $\psi_e = e_L + e_R$. Using the fact that $(e_L + e_R)^\dagger \gamma^0 (e_L + e_R) = e_L^\dagger \gamma^0 e_R + e_R^\dagger \gamma^0 e_L$, we get that the mass of an electron is

$$m_e = aG\,. \tag{2.58}$$

Thus, all of the masses in the electroweak theory are generated by the field Φ. We can continue to analyze the electroweak Lagrangian density and deduce the various coupling constants and other properties between the particles involved; however, we will not go into these details here. Also recall that we included only electrons and neutrinos as the leptons of the theory. Adding the other leptons is straightforward. Simply add the analogous $\mathcal{L}_{ab} + \mathcal{L}_{ab\Phi}$ for the muon and tau, where a is a muon or tau and b is the corresponding neutrino. It is also possible to add quarks into the picture by adding extra terms to the Lagrangian density. Finally, we should answer the basic question of how the electromagnetic and weak interactions can be viewed as a single unified force. The answer is that at high enough temperatures (energy density) when electroweak symmetry is no longer broken, all the fundamental particles remain massless and there is no distinction between these interactions. If temperatures drop, the symmetry is spontaneously broken, and the interactions are as we commonly see them. In particular, the fields for the gauge bosons for the electromagnetic and weak interactions are each mixtures of the W_i and X fields. [6][7]

2.3 Supersymmetry

Supersymmetry (SUSY) is a quantum field theory that postulates that there is a transformation linking fermions with bosons, such that the physical laws (action) for the theory is invariant under this transformation. This kind of transformation is called a symmetry

for the theory, and in particular it implies that every fermion has a corresponding bosonic partner and every boson has a corresponding fermionic partner, and when the symmetry is unbroken, the masses of the corresponding fermion-boson pairs coincide. The theory has the potential to solve several important problems in particle physics. [6]

One of them is the problem of renormalizing the mass of the Higgs Boson. While the SM particles gain mass by coupling to the Higgs field, the Higgs field gains a correction to its mass by coupling to the SM particles. Figure 2.5 shows the first three diagrams of an infinite sequence of fermion loop Feynman diagrams that contribute to the mass of the Higgs boson.

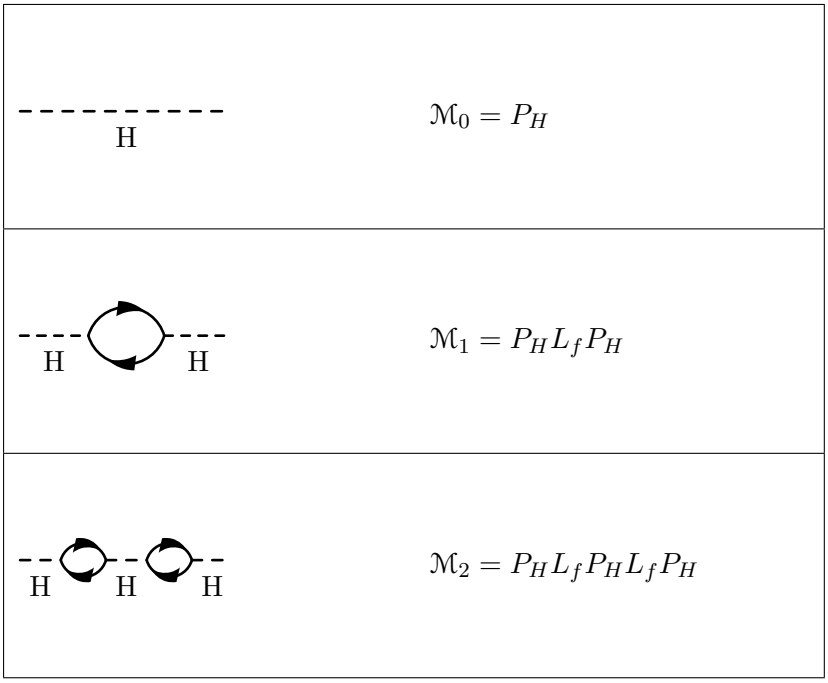


Figure 2.5: The first three fermion loop diagrams for the Higgs boson, along with their corresponding amplitudes. P_H is the propagator factor and L_f is the fermion loop factor.

The net amplitude of the entire sequence is the sum

$$P_H \sum_{j=0}^{\infty} (L_f P_H)^j \,, \tag{2.59}$$

and it can be shown that

$$P_H = \frac{i}{p^2 - m_H^2}, \quad L_f \propto -m_f^2 G_F \Lambda^2 \,, \tag{2.60}$$

such that m_f is the mass of the fermion, G_F is a factor of the coupling constant between the fermion and the Higgs boson, and Λ is the maximum energy threshold. The infinite sum (2.59) is just

$$\frac{P_H}{1 - L_f P_H} = \frac{i}{(p^2 - m_H^2)(1 - L_f P_H)} \tag{2.61}$$

$$= \frac{i}{p^2 - (m_H^2 + iL_f)} \,. \tag{2.62}$$

The correction to the (squared) mass of the Higgs boson due to the fermions is thus

$$\Delta m_H^2 \propto -m_f^2 G_F \Lambda^2 \,. \tag{2.63}$$

On its own, this correction would be troublesome, since it is a negative correction to the mass that grows with the energy threshold. However, we also have the bosons. It turns out that their loop diagrams lead to a correction of

$$\Delta m_H^2 \propto +m_b^2 G_F \Lambda^2 \,, \tag{2.64}$$

where m_B is the mass of the boson. Under SUSY, the pairing between the fermions and bosons would cause the net correction due to all fermions and bosons to converge to a value of the order of the SUSY breaking energy scale ($\sim 1\,\text{TeV}$). [9]

Another application of SUSY includes the modification of the energy dependence of the three fundamental coupling constants so that they perfectly agree at the Grand Unified Theory (GUT) scale. Figure 2.6 shows a comparison between the convergence of the coupling constants at the GUT scale for (a) The Minimal SM (MSM), and (b) The Minimal SM with SUSY (MSSM). [6]

A significant number of theories beyond the SM assume the existence of SUSY. In particular, String Theory, a quantum theory of gravity, is one of them. Also, in most models of SUSY, the lightest supersymmetric particle is a good candidate for dark matter, since it is colourless, neutral, and stable. The stability follows from the concept of R-parity. Supersymmetric particles are given a different value of R-parity than SM particles, and R-parity is assumed to be conserved. [6] [10]

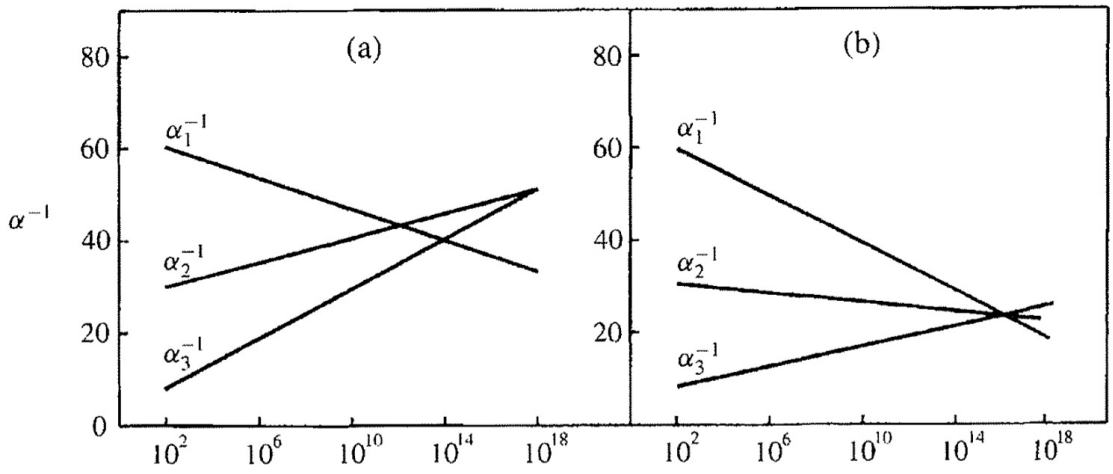


Figure 2.6: Convergence of the coupling constants at the GUT scale for (a) The MSM, and (b) The MSSM. The horizontal axis is energy in GeV. [6]

If SUSY exists, then it must be broken such that the masses of the supersymmetric particles are greater than the masses of the corresponding SM particles. However, most minimal supersymmetric models of SUSY breaking require at least 124 independent parameters, which makes experimental searches for SUSY extremely challenging. [6]

Chapter 3

Experimental Framework

3.1 The Collider

The LHC is a particle accelerator built in 2008 by the European Organization for Nuclear Research (CERN) with the aim of helping scientists to answer key unresolved questions in physics, such as: the origin of mass, SUSY, dark energy, dark matter, the matter-antimatter imbalance, the nature of the early universe, and the existence of extra dimensions. It consists of a 27 km long circular tunnel more than 100 m underground that accelerates protons to a maximum proton-proton collision energy of 14 TeV, making it the world's largest and most energetic particle accelerator. Before entering the main tunnel, protons, isolated from hydrogen gas by stripping the atoms' electrons, are injected into the Proton Synchroton (PS) booster at an energy of 50 MeV per proton. The booster accelerates them to 1.4 GeV at which point they are injected into the PS where they are accelerated to 25 GeV. Next, they travel to the Super Proton Synchroton (SPS) where they are accelerated to 450 GeV. They are then ready to enter the LHC where they are accelerated, both in a clockwise and anticlockwise direction, until they reach the desired energy level. In order to guide the protons along the circular paths defined by the tunnels, they are acted upon by magnetic fields generated by superconducting magnets, which are kept in operation by cooling them to a temperature of about $-271°$C. The strength of the fields vary with the speed of the

protons. When the protons are first injected into the LHC, the field strength is 0.535 T, and as the protons accelerate, this increases to a maximum of 8.33 T. The beam pipes within the tunnels are kept at an ultrahigh vacuum state (10^{-13} atm) so that additional particles within the tunnel do not add a significant amount of background noise to the decay products of the proton collisions. When travelling along the main beam pipe, the protons are spaced out into groups called *bunches*. Each bunch may contain as many as 115 billion protons and the bunches may be separated by as little as 25 ns. When two opposing bunches cross, at most 40 significant proton-proton interactions (or collisions) should be expected to occur, with an average of about 23 at the design luminosity (defined below). If we count only the interactions between protons that originated from the same bunch crossing, then this number is called *in-time pile-up*, while *out-of-time pile-up* refers to the number of interactions per bunch crossing between protons originating from different bunch-crossings. *Pile-up* refers to the sum of in-time and out-of-time pile-up. Each bunch crossing is called an *event*, and during a time interval (t_1, t_2), the number of collisions of a certain type is given by

$$N_{col} = \sigma_{col} \int_{t_1}^{t_2} \mathcal{L}(t) dt \,, \tag{3.1}$$

where σ_{col} is the cross section for the given type of collision, and $\mathcal{L}$ (not to be confused with the Lagrangian density) is the *instantaneous luminosity* of the incoming protons, i.e. the number of protons per unit area per unit time passing the interaction point. The design luminosity is $10\,\mathrm{nb}^{-1}\mathrm{s}^{-1}$. For 7 TeV collisions, the total cross section, that is, the cross section for any kind of collision, has been measured to be 98.0 ± 2.5 mb [11]. At a constant instantaneous luminosity of $1\,\mathrm{nb}^{-1}\mathrm{s}^{-1}$ (typical for 2011 7 TeV collisions), this translates into (approximately)

$$(98 \times 10^6)\,\mathrm{nb} \times 1\,\mathrm{nb}^{-1}\mathrm{s}^{-1} = 98 \text{ million collisions per second} \,. \tag{3.2}$$

[12][13]

The main LHC tunnel contains four interaction points where proton-proton collisions occur. At each point, there is a detector absorbing the particles emanating from the collision

and taking measurements on their energies and trajectories. Currently, the seven detectors placed along the LHC tunnel are: ATLAS, CMS, ALICE, LHCb, TOTEM, LHCf, and MoEDAL. ATLAS and CMS are both general purpose detectors with the same goals but with different technical solutions and design methods for achieving them. ALICE is specifically designed to detect a state of matter known as quark-gluon plasma, which is hypothesized to have existed shortly after the occurence of the Big Bang. LHCb is a forward region detector designed to investigate the matter-antimatter imbalance problem, by focusing on the detection of the b quarks. TOTEM is a forward region detector sharing the same interaction point as CMS with goals such as measuring the size of the proton and accurately monitoring the LHC's luminosity. LHCf is a forward region detector sharing the same interaction point as ATLAS with the goal of simulating cosmic rays in laboratory conditions. MoEDAL, a relatively new addition, is designed to search for the magnetic monopole and other highly ionizing stable massive particles [14]. This thesis will focus on the ATLAS detector. Figure 3.1 shows a diagram of the accelerating tunnels and interaction points for the LHC. [12]

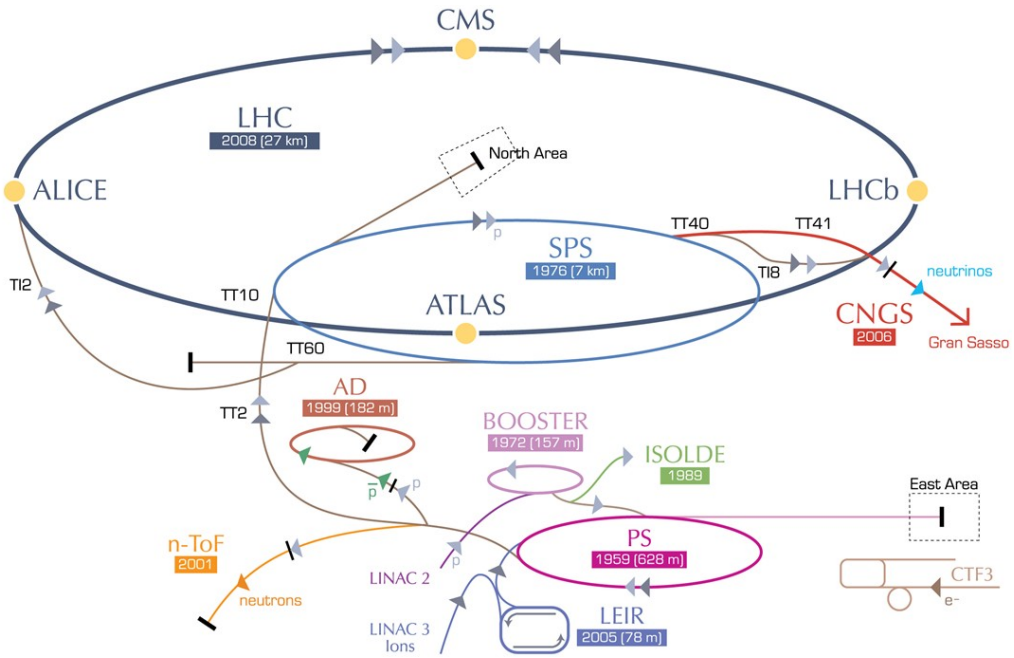


Figure 3.1: A diagram of the accelerating tunnels and interaction points for the Large Hadron Collider [15].

The first beams began circulating in the LHC on September 10, 2008 [16], but 9 days later, operations were halted due to a faulty electrical connection between two of the accelerator's magnets, which resulted in mechanical damage to the superconducting magnets, the release of helium from the magnet cold mass into the tunnel, and contamination of the vacuum beam pipe [17]. The beams resumed circulation on November 20, 2009 [18], and on March 30, 2010 data taking began at a 7 TeV collision energy [19]. This continued until April 2012 (with breaks in the winter), when the energy was increased to 8 TeV [20], and then continued with that level of energy until February 2013, the beginning of the long term shut down period [21].

3.2 The Detector

The ATLAS detector is a cylindrically shaped structure enclosing the LHC beam pipe with a 46 m length and a 13 m radius and consisting of four main components: the inner detector, the calorimeter, the muon spectrometer, and the magnet system. The inner detector measures the momenta of charged particles, the calorimeter measures the energies of charged particles and hadrons, the muon spectrometer focuses on measuring the momenta of muons, and the magnet system curves the trajectories of the particles so that their momenta can be measured based on the radius of the curvature.

The inner detector lies in the innermost 1.15 m radius surrounding the interaction point. It is subdivided into three parts (in order of innermost to outermost): the pixel detector, the semiconductor tracker (SCT), and the transition radiation tracker (TRT). The pixel detector, crucial for vertex reconstruction, consists of three silicon based layers at radii $r = 50.5\,\text{mm}$, $r = 88.5\,\text{mm}$, and $r = 122.5\,\text{mm}$, which provide a three-hit system with 8×10^7 pixels of resolution for particles with $|\eta| < 2.5$ [22]. The SCT consists of 8 layers of silicon microstrip detectors with pairs of layers at radii $r = 30.0\,\text{cm}$, $r = 37.3\,\text{cm}$, $r = 44.7\,\text{cm}$, and $r = 52.0\,\text{cm}$, providing 6.2 million pixels of resolution. The TRT consists of hundreds of layers of straw detectors with a diameter of 4 mm, each containing a gold-plated wire with a diameter of $30\,\mu\text{m}$. The barrel part of the TRT contains about 50000 axially oriented

straws, each divided in two at the centre, and covering the radial range from 56 to 107 cm. The end caps contain 320000 radially oriented straws, and they cover the radial range from 64 to 103 cm for the 14 layers nearest the interaction point and a radial range from 48 to 103 cm for the other 4 layers. The spatial resolution of this part of the tracker, 170 μm per straw, is significantly worse than the others, but the construction is much more cost effective given the larger volume, and as a unique feature, Xenon gas within the straws provides an accurate measurement of transition radiation, which is primarily used to detect the presence of electrons. The inner detector detects the track of any charged particle; neutral particles do not bend under the presence of the magnetic field.

The calorimeter encloses the inner detector and is subdivided into an electromagnetic (EM) calorimeter and a hadronic (HAD) calorimeter. The EM calorimeter measures the energies of charged particles. It uses lead as the material that absorbs the energies of the particles and liquid argon (LAr) as the material that samples their energies. Its granularity varies between 0.003 to 0.1 for $\Delta\eta$ and 0.025 to 0.1 for $\Delta\phi$, depending on the location. The HAD calorimeter measures the energies of particles that pass the EM calorimeter and interact via the strong force. Its main section, the tile calorimeter uses iron as the absorbing material and scintillating tiles to sample the energies. Its end cap and forward sections use liquid argon as the sampling material and copper and tungsten as the absorbing materials. These regions tend to have a higher energy density of colliding particles and hence need more conductive materials to dissipate the resulting heat. Its granularity in $\Delta\eta \times \Delta\phi$ varies between 0.1×0.1 and 0.2×0.1, and is thus coarser than that for the EM calorimeter. All charged particles interact in the EM calorimeter and all hadrons interact in the HAD calorimeter. Particles that are both charged and hadrons, such as the proton, interact in both calorimeters, but primarily in the HAD calorimeter. The lower density of the absorbing material for the EM calorimeter ensures that hadrons pass through it without significantly interacting with its nuclei, due to the short range of the residual strong force. Electrons typically end their journey at the EM calorimeter through the process of bremsstrahlung radiation, while the more massive muons easily pass through both the EM

and HAD calorimeters, since their mass heavily suppresses any possible bremsstrahlung radiation.

The muon spectrometer surrounds the calorimeter and works like the inner detector to track the displacement of particles through it, where in this case the trackable particles that make it past the calorimeters are mostly muons. It has a lower spatial precision than the inner detector trackers and more than one million readout channels.

Finally, the magnet system, crucial for the proper functioning of the tracking system, consists of an inner solenoid magnet surrounding the inner detector and providing a nearly uniform magnetic field of 2 T, and an outer superconducting toroid magnet within the muon spectrometer providing a magnetic field of up to 4 T. Figure 3.3 shows a computer generated image of the ATLAS detector. [1]

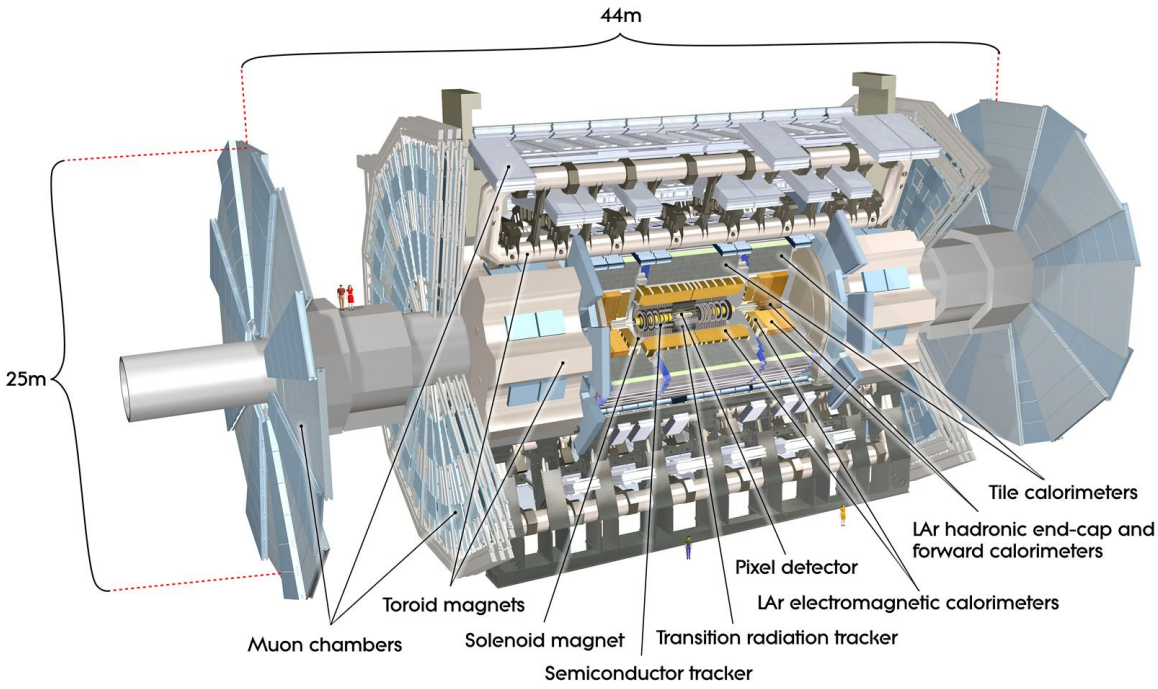


Figure 3.2: A computer generated image of the ATLAS detector [23].

3.3 The Trigger

The LHC proton bunches are designed to cross at a frequency of 40 MHz for more than 10 hours per day. This means that each day, more than 14.4 trillion events are expected to occur at the ATLAS interaction point. The information for each event has a size of about 1.5 MB, so storing all these events would require more than 21.6 million TB of storage space per day, which is far beyond the reasonable limit. In order to select and analyze interesting events in a timely manner, while considering storage constraints, ATLAS uses a system called a trigger to decide which events to accept and which to reject. The trigger is divided into three parts: Level-1 (L1), Level-2 (L2), and Event Filter (EF). L1 is the first part of the trigger to obtain information about an event. It is mostly hardware based, takes less than 2.5 μs to make a decision, has an event output frequency of about 75 kHz, and has the most primitive selection process. It determines regions of interest (RoI), that is regions in the (η, ϕ) space of the detector where it has found interesting features, based on simple energy readouts from the calorimeter, and passes on the information associated with the RoIs to the next level. L2 takes as input the RoIs selected by L1, analyzes them in a more complex manner, and decides, in about 40 ms on average, whether to accept or reject them. This may involve performing some basic track reconstruction in order to determine more information on the dynamics of the particles involved instead of simply considering the energy deposited. The L2 event output frequency is about 3 kHz. The RoIs accepted by L2 are then passed to the EF, which performs an even more complex analysis, in about 4 s on average, to determine whether to accept or reject the event. The event output frequency of the EF is about 200 Hz. If the event is accepted, then all information about the event is put into long term storage for offline analysis. This results in a data storage rate of about 300 MB/s, which yields a reasonable limit for data storage consumption. [24][25]

Since different physics searches have different demands on the properties of an event that are interesting enough to keep, each physics trigger group in ATLAS uses its own trigger criteria. This is done by allowing each group to design their own L1, L2, and EF criteria based on measurable properties (variables) for the physics objects of interest. The

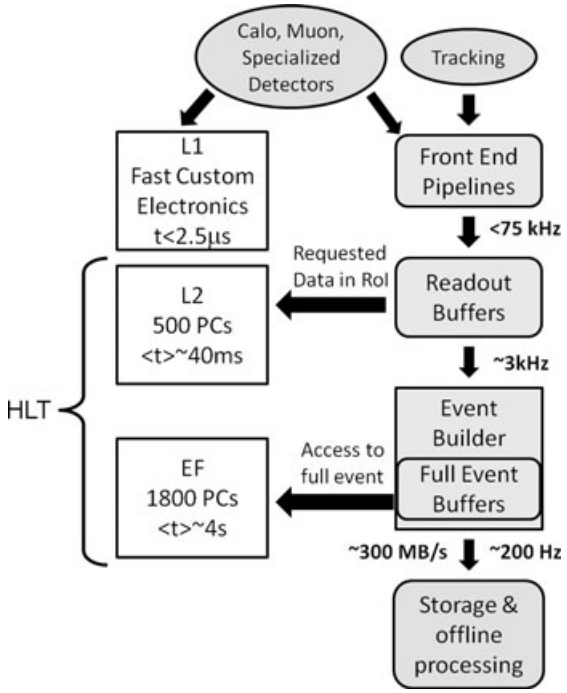


Figure 3.3: A diagram of the ATLAS trigger system [25].

variables are generally more precise the higher the trigger level, and are typically based upon properties of one or more RoIs that represent the candidate particles of interest. Each group may have multiple sets of criteria for each level, resulting in different types of triggers for each level and for each group. Overall trigger resources are divided amongst the different types of triggers by assigning to each type a number s called the *prescale* associated with it, which is defined such that it considers for acceptance only one out of s eligible events. Naturally, the L1 triggers are prefixed with `L1`, the L2 triggers are prefixed with `L2`, and the EF triggers are prefixed with `EF`. The underscore character is used to divide various types of criteria. For example, `L1_TAU8` is the label for the L1 tau trigger (explained in Chapter 4) with a requirement that the input RoI have L1 $E_T > 8\,\text{GeV}$, `L2_TAU16_LOOSE` is the label for the L2 tau trigger with loose cut requirements whose output is the input for the EF tau trigger `EF_TAU16_LOOSE`, which has loose cut requirements and an EF $E_T > 16\,\text{GeV}$ requirement. `L1_TAU11` is just like `L1_TAU8`, except that the energy threshold is 11 GeV rather than 8 GeV; `L2_TAU16_MEDIUM` is just like `L2_TAU16_LOOSE`, except that the cuts are

tighter; and similarly for `EF_TAU29_MEDIUM`. Note: By tighter we mean that the set of RoIs that result from those cuts are a subset of the set of RoIs that result from the looser cuts. Generally, loose cuts have a higher prescale than tight cuts, in order to keep the overall physics trigger frequency below the preassigned limit. For reference, a *trigger chain* is a triple consisting of an L1, L2, and EF trigger such that the L2 trigger operates on the L1 trigger's output and the EF trigger operates on the L2 trigger's output. [25][26]

3.4 The Event Data Model

The digital representation of events detected by ATLAS is commonly called the *Event Data Model*. After an event has been accepted, it is stored in a format called the *Raw Data Format*, which uses about 1.5 MB per event. This format is mainly used for backup and testing purposes. Events in the Raw Data Format are then converted into the *Event Summary Data* (ESD) format or the *Analysis Object Data* (AOD) format. The ESD format contains the detailed output of the detector reconstruction and uses about 500 kB per event. It contains sufficient information to allow for particle identification, track-refitting, and jet calibration. The AOD format contains a less detailed summary of the reconstructed events and uses about 100 kB per event. It contains sufficient information for common analyses, and is typically divided into various streams that are each tailored for a different type of physics analysis. Events in the AOD format can be created from events in the ESD format, and thus it is generally unnecessary to process the raw data in order to create events in the AOD format. Another commonly used format is the *Derived Physics Data* (DPD) format. It comes in various forms, but is generally derived from raw, ESD, or AOD data by removing certain unneccessary information. The *D3PD* format is a type of DPD format that uses the ROOT format to encapsulate the data rather than the POOL format that is used for the other formats. [27] The ROOT format allows for convenient access to the event information using the ROOT software [28]. [29]

Chapter 4

The Tau Trigger

4.1 The Tau Lepton

The tau lepton, with a mass of $1.8\,\mathrm{GeV/c^2}$, is the heaviest of all known leptons, and the only lepton that can decay hadronically. It decays into two neutrinos and a lepton 35% of the time, while 65% of its decays involve a neutrino and hadrons, where the hadrons are mostly pions. The principal modes of tau production at the LHC are $W^{\pm} \rightarrow \tau^{\pm}\nu_{\tau}^{\mp}$ and $Z \rightarrow \tau^{+}\tau^{-}$. At a collision energy of 8 TeV, the cross section for the former is 9.1081 nb and the cross section for the latter is 8.7804×10^{-1} nb [30][31]. While the former does have a greater cross section, the presence of a neutrino causes a greater uncertainty on the mass of the W boson, which negatively affects the certainty of the resulting tau reconstruction. [32]

Understanding events with taus will not only help in understanding SM processes, but will also contribute greatly to the discovery potential of the ATLAS detector. For example, recent experimental evidence suggests a SM Higgs boson in the mass range of 125-127 $\mathrm{GeV/c^2}$ [5], which would decay into tau lepton pairs with a branching fraction of about 6% [33]. Figure 4.1 shows a Feynman diagram for a typical Higgs boson production and decay of this type, and Figure 4.2 shows the branching ratios for several common Higgs boson decays as functions of the Higgs boson mass.

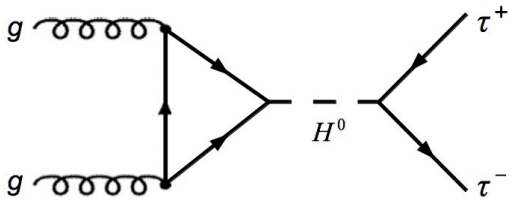


Figure 4.1: A Higgs boson created through gluon-gluon fusion and decaying into two taus. [34]

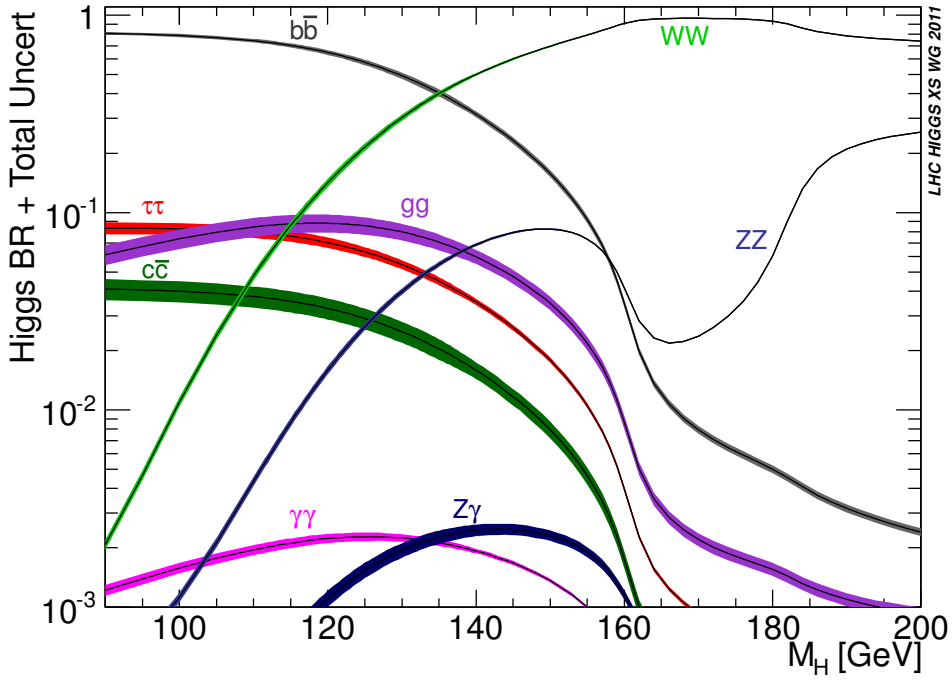


Figure 4.2: The branching ratios for several common SM Higgs boson decays as functions of the Higgs boson mass. The $H \to \tau\tau$ branching ratio is shown in red. [33]

Furthermore, the MSSM requires the existence of two Higgs doublets rather than one. With this requirement, electroweak symmetry breaking leads to the existence of five spin-0 Higgs bosons: h^0, H^0, A^0, H^+, H^-. The first three are neutral, the H^+ is positively charged, and the H^- is negatively charged. The parameter $\tan\beta$, defined as the ratio of the vacuum expectation values of the doublets, has a large effect on the phenomenology of the model. For $\tan\beta \gtrsim 2$, charged Higgs bosons decay dominantly to tau leptons through the $H^\pm \to \tau^\pm\nu$ channel, making the identification of taus important for searching for charged Higgs bosons. The coupling of neutral MSSM higgs bosons to taus is quite significant as well [35]. [36][10]

4.2 2011 Tau Trigger

Due to its short lifetime (2.9×10^{-13} s) [32], the tau lepton decays inside the beam pipe of the LHC, and hence the identification of taus is done through their decay products. The events where taus decay into charged leptons can be triggered by low E_T threshold electron triggers and low p_T threshold muon triggers. Taus that decay into hadrons are called hadronic taus, and for events containing them, we use a dedicated hadronic tau trigger, which as the only trigger dedicated to taus, is referred to as the tau trigger. [26]

By the nature of the collisions involved, jets of hadrons created through QCD interactions are an overwhelmingly large background to hadronic taus passing through the ATLAS detector. Discriminating between a hadronic tau decay and a QCD (gluon-produced) jet can be very challenging. In general, hadronic taus are identified by selecting jets that are narrow, well collimated, have low track multiplicity, and small energy in their outer region. The source of QCD jets tends to have a greater invariant mass than the mass of the tau lepton, and this gives them more room in phase space for a wider and more complex shower of hadrons. Figure 4.3 shows a visual comparison between QCD jets and tau-produced jets. [26]

The three stages of the 2011 tau trigger can be summarized as follows.

At L1, the tau trigger defines RoIs using EM and HAD calorimeter trigger towers with

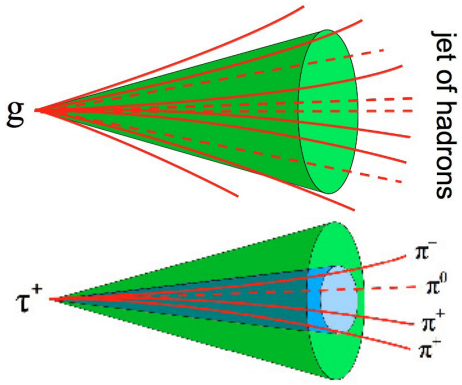


Figure 4.3: Comparison between QCD jets and tau-produced jets. Notice that the tau jets lie in a narrower cone and have a lower number of tracks. [34]

granularity $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. The hadronic tau decay modes are identified by: the energy in 2×1 pairs of EM towers, the energy in 2×2 HAD towers behind the EM cluster, and the energy in a 4×4 (isolation) ring around the 2×2 core. The L1 rate can be tuned by adjusting the threshold on the transverse energy of the 2×2 core, or the maximum allowed energy in the isolation ring. Figure 4.4 shows the calorimeter towers used at L1. [37][26]

At L2, a calorimeter-based algorithm refines the position of the RoI and obtains total transverse energy and cluster shape variables using full detector granularity within a region containing the L2 RoI. The computing time required to reconstruct the tracks is significantly greater than that required to reconstruct the calorimeter-based variables. Therefore, a transverse energy cut is performed before track reconstruction. Once the tracking variables have been calculated, the cuts are split into the cases: single-prong and multi-prong. [26]

Finally, the EF stage selects taus based on track multiplicity and both cluster and track shape variables with high discrimination power, calculated using track and calorimeter based algorithms similar to those used offline. [26]

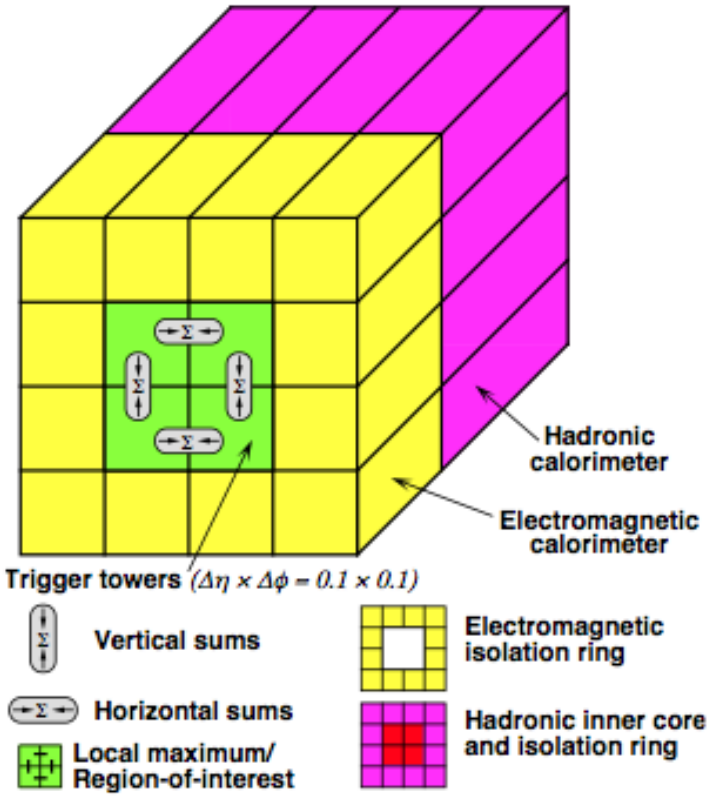


Figure 4.4: The calorimeter towers used at the L1 stage of the tau trigger. [37]

4.3 Offline Reconstruction

Recall that the events that pass the EF of the ATLAS trigger are put into long-term storage for offline analysis. This analysis typically includes the application of a reconstruction algorithm for the objects of interest to the events. For each event, the reconstruction algorithm finds RoIs that contain the object(s) of interest and calculates variables for the RoIs that may be used when later performing selection cuts. For tau-based analyses, the most common objects of interest are jets (for background) and taus (for signal). [24]

A jet reconstruction algorithm typically starts with the formation of *topological clusters* of calorimeter cells, which are 3-dimensional connected sets of cells in the calorimeter. The formation of each cluster begins with a seed cell, and neighbours of cells already in the cluster are iteratively added to the cluster provided that the energy in the new cell is above a threshold defined as a function of the expected noise. The clusters are then separated such that they may only share cells with other clusters at their borders. [38] Once the clusters are formed, they are passed on to a jet-finding algorithm, which can be classified as either a *cone algorithm* or a *clustering algorithm*. Cone algorithms define jets as sets of clusters within a $\Delta R < r$ cone, where r is typically 0.4 or 0.7. The jets are then separated with a split-merge procedure. Clustering algorithms construct jets via a clustering procedure which depends on the relative transverse momenta of the topological clusters, and ensure a unique association of these clusters to the jets without requiring a split-merge procedure. The jets resulting from a jet-finding algorithm are stored as digital objects containing properties such as energy and direction, and these objects are commonly referred to as *offline jets*. [39][40]

The tau reconstruction algorithm, `tauRec`, reconstructs hadronically decaying taus (candidates) from two possible objects (seeds): a calorimeter jet or a good quality track. The calorimeter jet must be an offline jet reconstructed using a cone algorithm that is run over topological clusters, and must have $E_T > 10\,\text{GeV}$ and $|\eta| < 2.5$. Tracks with $p_T > 1\,\text{GeV/c}$ are associated to the calorimeter seed if they are within a $\Delta R < 0.3$ cone and satisfy certain quality criteria. The track seed must have $p_T > 6\,\text{GeV/c}$ and pass the same quality

criteria. Tracks with $p_T > 1\,\mathrm{GeV/c}$ that are within a $\Delta R < 0.2$ cone are associated to the track seed as well provided they satisfy a different set of quality criteria. If exactly three tracks are found, their net charge must have absolute value 1. If more than 8 tracks are found, the track seed is not considered. Calorimeter seeds and track seeds are matched in a $\Delta R < 0.2$ cone, and if they overlap, only one tau candidate is built for them. The digital representation of a tau candidate is called an *offline tau*. This object contains the information for both the tracks associated to the track seed and the tracks associated to the calorimeter seed. It also contains calorimeter-based information and discrimination variables that use both the calorimeter and tracking information. For offline taus with a calorimeter seed, their energies are calculated by iterating over the cells associated with the offline jet and summing up their energies with cell weights calibrated with a standard jet calibration technique, and later corrected to more accurately model taus. For offline taus with a track seed, an algorithm calculates the so-called *e-flow* energy of the candidate. The algorithm is based on the momenta of the tracks as measured in the inner detector and the energy of the clusters measured in the EM calorimeter, and is optimized for calculating the energies of taus, while slightly underestimating the energies of QCD jets. If the offline tau has no calorimeter seed, then the e-flow energy is the baseline definition of its energy. It should be made clear that offline taus are merely offline jets with a loose bias towards taus and with energy variables calibrated to more accurately model taus rather than QCD jets. The creation of offline taus is the main purpose of the reconstruction stage. The *offline identification* stage selects taus by performing cuts on the discrimination variables that are stored inside the offline taus. [41][42]

Related to reconstruction is the concept of simulation. Physical processes are often simulated in order to compare theoretical predictions with experimental measurements. In ATLAS, various Monte-Carlo (MC) based algorithms are used to produce simulated events that can be compared with the real events detected by ATLAS. The data representing the simulated events is commonly referred to as *MC*, while the data representing the real events is commonly referred to as simply *data*. Just like data, MC contains complete

trigger information that includes what triggers fired, the discrimination variables for each RoI at each trigger level, the detailed calorimeter and tracking information, and the offline reconstructions. [43]

4.4 Performance

The study of the performance of the tau trigger generally follows the steps: (1) measure L1 cross section, (2) measure EF frequencies, (3) validate variables, and (4) measure efficiencies. [44]

The L1 cross section is calculated as the ratio of the frequency of L1 accepted events to the instantaneous luminosity of the LHC. Figure 4.5 shows the L1 cross section as a function of pile-up for a variety of tau triggers, with events collected in 2011. [45]

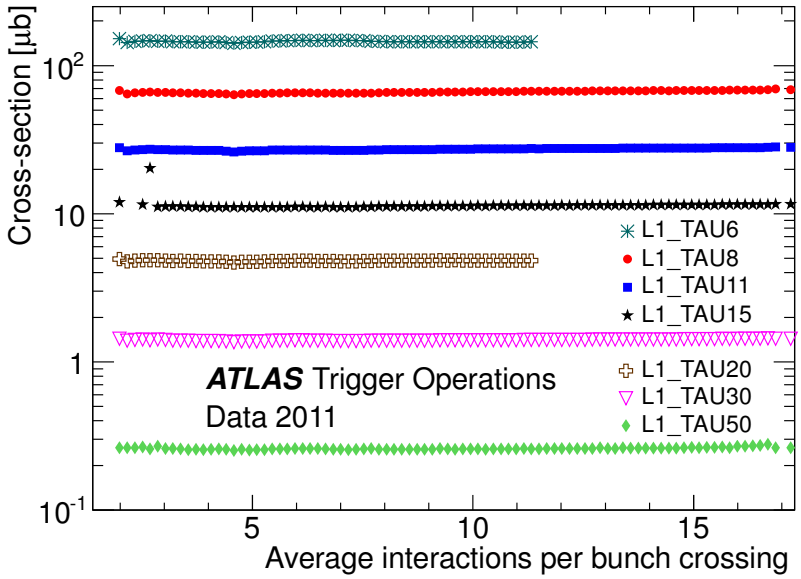


Figure 4.5: L1 cross section as a function of pile-up for a variety of tau triggers, with events collected in 2011 [45].

As desired, the cross section is relatively constant with changing pile-up conditions.

The EF frequency as a function of instantaneous luminosity for a variety of tau triggers, with a portion of the events collected in 2011, is shown in Figure 4.6.

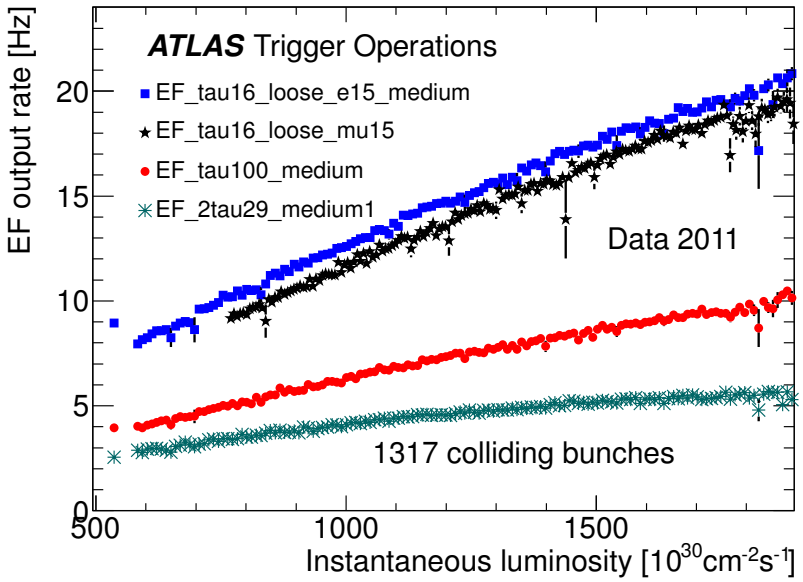


Figure 4.6: EF rate (frequency) as a function of instantaneous luminosity for a variety of tau triggers, with events collected from September to October 2011. [45]

Some of the triggers shown in the figure are combinations of a single-tau trigger with another single-object trigger. In this case, the trigger requires both of the single-object triggers to accept the event in order for the event to be accepted. For example, `e15_medium` represents the trigger that requires electrons with $E_T > 15\,\text{GeV}$ and medium-type cuts. Similarly, `mu15` requires a muon with $E_T > 15\,\text{GeV}$. [45] Overall, the results show that the EF frequency scales linearly with instantaneous luminosity; as desired.

Next, for each level of the trigger, we show in Figures 4.7, 4.8, and 4.9, one (of several) validation plots used to demonstrate the consistency of MC with data for the variables used for tau selection. Offline candidates were selected using a tag-and-probe analysis with $Z \to \tau\tau$ events collected by ATLAS during Summer 2011, where one of the taus decayed leptonically [46]. The offline candidates were then matched to the candidates for the corresponding level of the trigger chain containing `EF_TAU20_MEDIUM1`. This chain requires $E_T > 20$ GeV, medium cuts, and at most 3 tracks at the EF. [45]

The *efficiency* of a tau trigger is generally defined as the probability that an offline identified tau passes the corresponding trigger requirements. Physics analyses using data

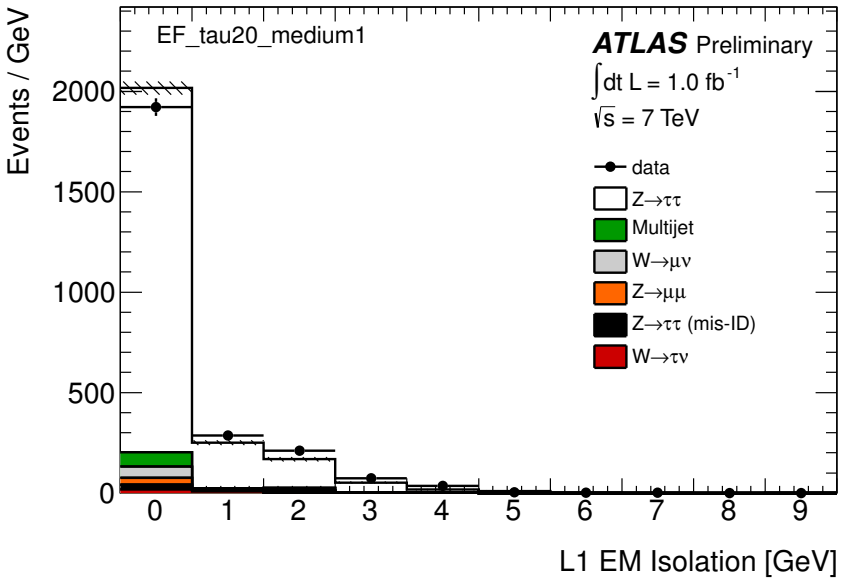


Figure 4.7: Histogram of L1 EM isolation, defined as the energy deposited in the EM isolation ring (shown in Figure 4.4), as calculated by L1. Events were selected with a $Z \to \tau\tau \to \mu$h tag-and-probe analysis on events collected by ATLAS during Summer 2011. [45]

collected with a tau trigger need to understand the trigger efficiency in order to properly correct the estimations from the simulation samples. Figure 4.10 shows the efficiency of the same EF trigger as a function of offline p_T for the same offline selections as above.

The data-to-MC ratio together with its uncertainty can be used to scale MC histograms by p_T bin in order to accurately model the data. The close agreement of data with MC in Figure 4.10 demonstrates the consistency of MC with data, and the relatively high efficiency shows that the tau trigger has indeed performed well in the year 2011. [26] [34]

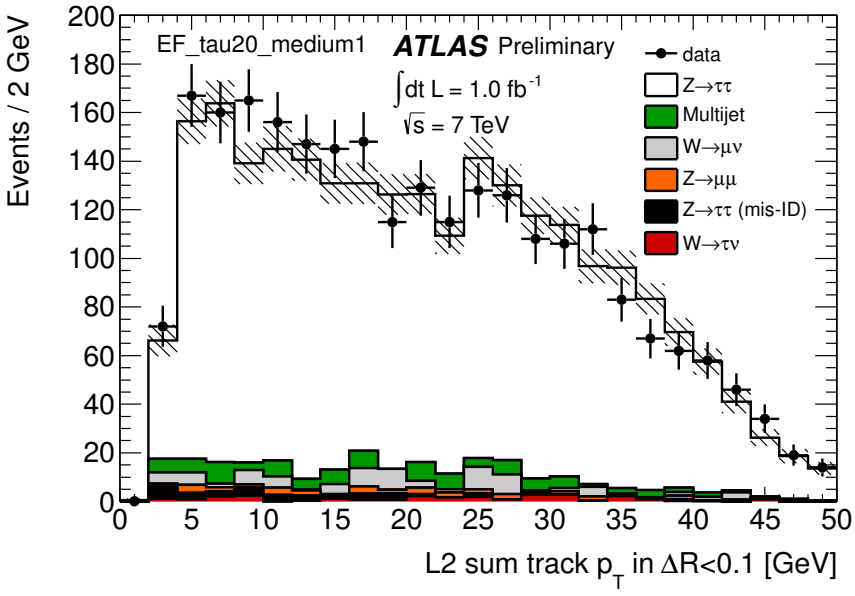


Figure 4.8: Histogram of L2 sum track p_T, defined as the summed p_T of the tracks in the region $\Delta R < 0.1$, as calculated by L2. Events were selected with a $Z \to \tau\tau \to \mu$h tag-and-probe analysis on events collected by ATLAS during Summer 2011. [45]

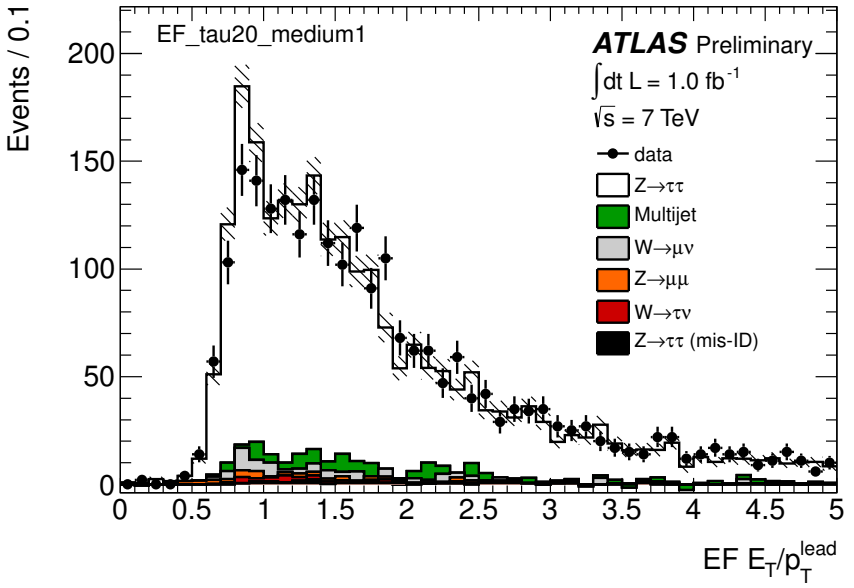


Figure 4.9: Histogram of EF E_T/p_T^{lead}, where p_T^{lead} is the p_T of the track with highest p_T, as calculated by the EF. Events were selected with a $Z \to \tau\tau \to \mu$h tag-and-probe analysis on events collected by ATLAS during Summer 2011. [45]

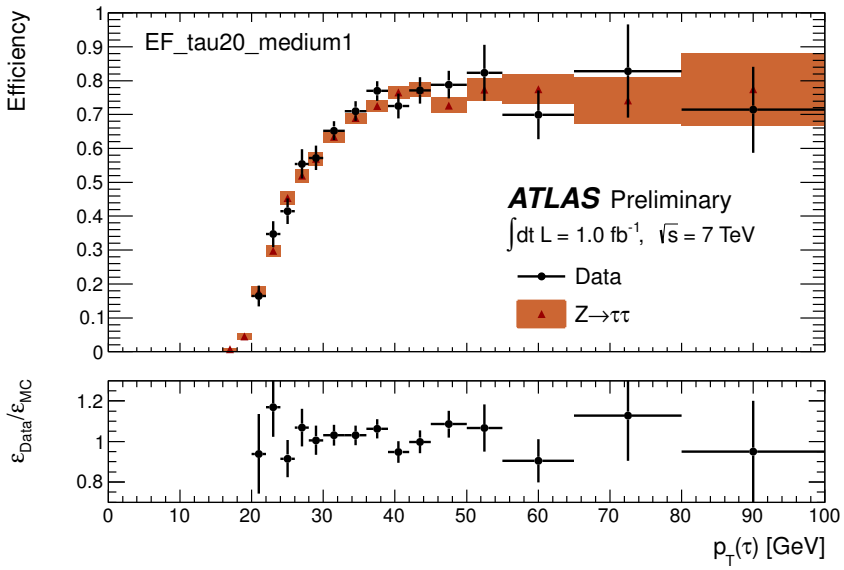


Figure 4.10: Efficiency of `EF_TAU20_MEDIUM1` as a function of offline p_T. $\varepsilon_{\text{Data}}$ is the efficiency for data and ε_{MC} is the efficiency for MC. Events were selected with a $Z \to \tau\tau \to \mu\text{h}$ tag-and-probe analysis on events collected by ATLAS during Summer 2011. [45]

4.5 Dijet Study

Understanding how QCD jets (dijets) pass the tau trigger is an important input to controlling the trigger frequency as a function of luminosity. In this study, dijet events are selected by requiring at least two offline taus, each with $p_T > 15\,\mathrm{GeV/c}$, $|\eta| < 2.5$, and $|\Delta\phi| > 2.7$ between the leading (p_T) and subleading offline taus (back-to-back). For reference, we also select $Z \to \tau\tau$ events by requiring at least one $\Delta R < 0.2$ truth-matched offline tau with $p_T > 15\,\mathrm{GeV/c}$. No offline identification is performed in either case.

For the dijet events, we use collision data from periods F to K of the 2011 runs, as well as a representative set of dijet MC data. For the $Z \to \tau\tau$ events, we use $Z \to \tau\tau$ MC data. The datasets were skimmed to only include events that passed the `EF_tauNoCut` trigger. This implies that `L1_TAU6` and `L2_tauNoCut` were also fired. The reason for choosing this trigger chain is to obtain a minimally biased set of events. Essentially, the only selection performed is the `L1_TAU6` requirement, which only requires an L1 RoI with L1 $E_T > 6\,\mathrm{GeV}$. `L2_tauNoCut` accepts any L2 RoI and `EF_tauNoCut` accepts any EF RoI. Of course, these triggers are heavily prescaled, so after skimming, the luminosity of the collision data is only about $15\,\mathrm{nb}^{-1}$.

For each event passing the requirements, we determine whether the leading offline tau passed `L1_TAU11` by matching it to an L1 tau through $\Delta R < 0.2$ and checking whether this L1 tau passes L1 $E_T > 11\,\mathrm{GeV}$. If that passes, we then determine whether it also passed `L2_TAU29_MEDIUM` by matching it to an L2 tau through $\Delta R < 0.2$ and checking whether the emulation of `L2_TAU29_MEDIUM` accepts the L2 tau. If that passes, we then determine whether it also passed `EF_TAU29_MEDIUM` by matching it to an EF tau through $\Delta R < 0.2$ and checking whether the emulation of `EF_TAU29_MEDIUM` accepts the EF tau. For each stage, we also fill histograms for offline p_T in order to be able to calculate the corresponding efficiencies.

Figure 4.11 shows the efficiency for `L1_TAU11`, Figure 4.12 shows the efficiency for `L2_TAU29_MEDIUM`, and Figure 4.13 shows the efficiency for `EF_TAU29_MEDIUM`. We found a reasonable agreement between the dijet events selected from collision data and those

selected from MC data.

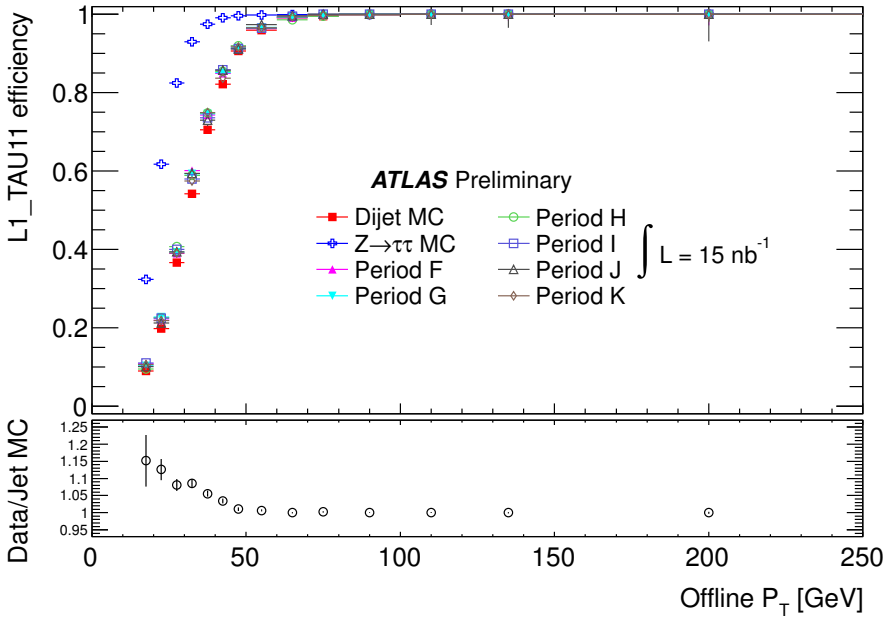


Figure 4.11: L1 trigger efficiency as a function of offline tau p_T. [26]

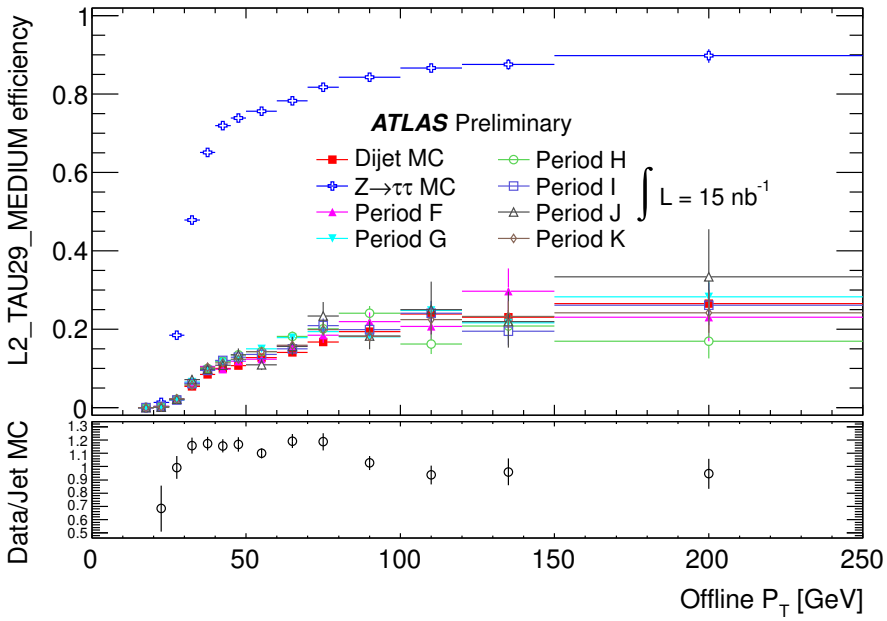


Figure 4.12: L2 trigger efficiency as a function of offline tau p_T. [26]

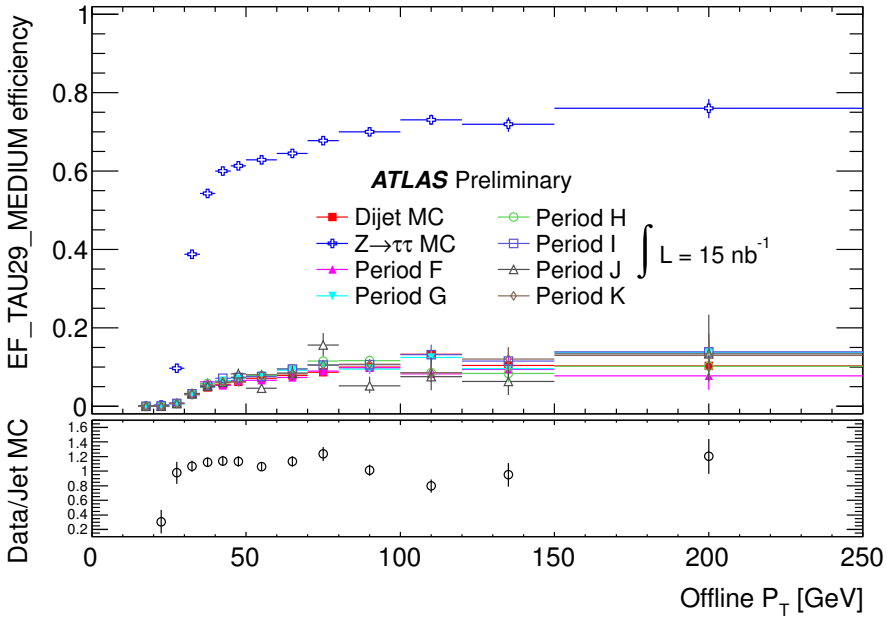


Figure 4.13: EF trigger efficiency as a function of offline tau p_T. [26]

4.6 2012 Improvements

As the energy of the bunch crossings increases or the number of protons per bunch crossing increases or the bunch spacing decreases, the number of significant proton interactions per bunch crossing increases as well. This number is also strongly correlated with the number of *primitive vertices* in the corresponding event, which are the starting points of the offline reconstructed tracks that each correspond to a single proton-proton collision. The dependence of the trigger efficiencies on pile-up was apparent near the end of the 2011 running period. The number of protons per bunch crossing was increasing, which led to an increase in in-time pile-up; a situation that was negatively impacting the efficiencies of the tau trigger in its then current form, as seen in Figures 4.14 and 4.15. [26]

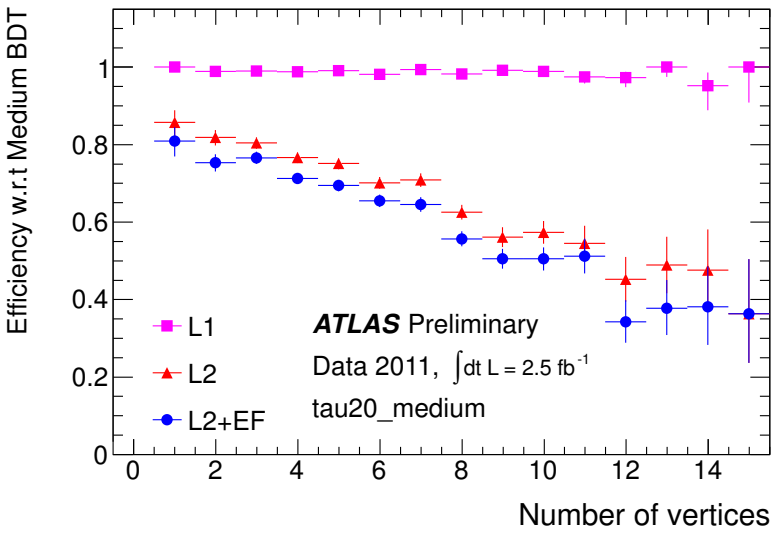


Figure 4.14: 2011 dependence of tau trigger efficiency on the number of (primitive) vertices in an event, a measure of pile-up. $Z \to \tau\tau \to \mu h$ events collected in 2011 were selected using a tag-and-probe technique. A medium offline identification was used. [45]

The preparation for the even higher energies and luminosities of 2012 was hence focused on modifying the tau trigger to account for the expected increased in-time pile-up. For L1, the main modification, not entirely pile-up motivated, was the addition of EM isolation cuts in order to reduce the rates as needed. For L2, a narrower region of (η, ϕ) space was used to calculate the L2 E_T variable which is used for the calorimeter cut stage ($\Delta R < 0.2$

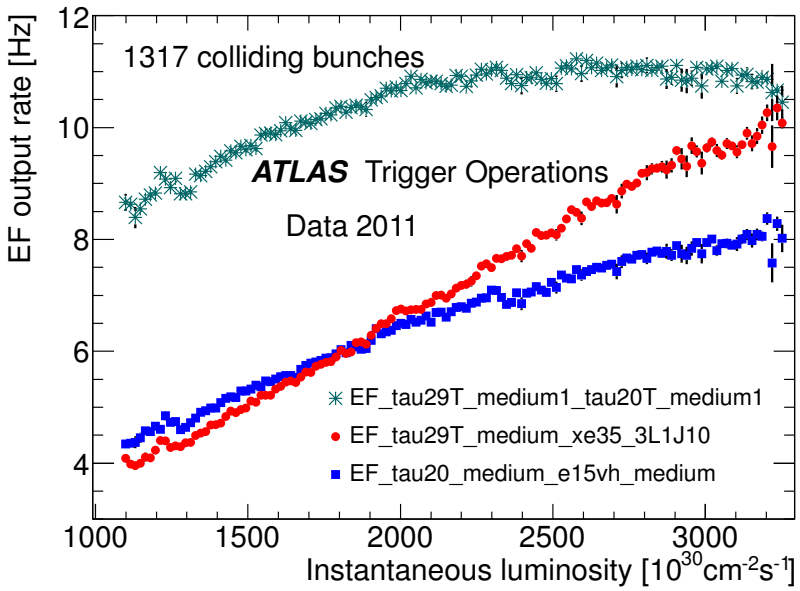


Figure 4.15: EF rate as a function of instantaneous luminosity for a variety of tau triggers, with events collected from September to October 2011. Notice that the EF rate for the 2011 ditau trigger `EF_tau29T_medium1_tau20T_medium1` does not scale linearly with instantaneous luminosity. This demonstrates the effect of pileup in widening the shower shape, and making it more difficult to resolve between two taus in the same event. [45]

instead of 0.4), a constraint of $\Delta z_0 < 2\,\text{mm}$ w.r.t. the leading p_T track was used for the tracks in the tracking stage, and the energy threshold for tracks was lowered to 1 GeV (from 1.5 GeV). In addition to the E_T cut, a cut on another calorimeter-based variable (`CoreFrac`) is performed before the L2 track reconstruction, in order to keep within timing constraints. The EF also obtained pile-up robust modifications such as a smaller Δz_0, and now uses two multivariate methods to perform its selection: boosted decision trees (BDT), and log likelihood (LLH) (explained in Chapter 5). The expected efficiency with respect to the number of vertices in an event for a 2012 tau trigger is shown in Figure 4.16 and can be compared with Figure 4.14 in order to see the improvement. Similarly, the EF rate as a function of instantaneous luminosity for a variety of tau triggers, with a portion of the events collected in 2012 is shown in Figure 4.17 and can be compared with Figure 4.15 in order to see the improvement. [44][47]

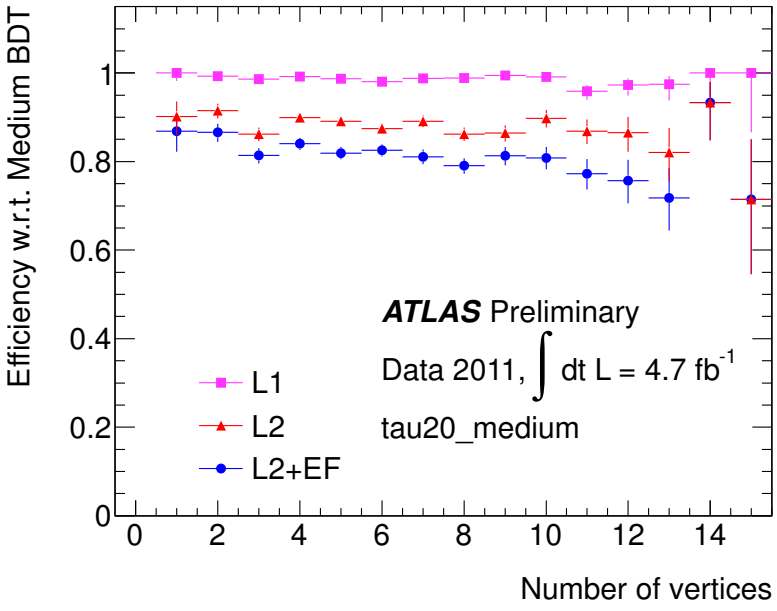


Figure 4.16: Expected 2012 dependence of tau trigger efficiency on the number of (primitive) vertices in an event, a measure of pile-up. $Z \rightarrow \tau(\mu)\tau(h)$ events collected in 2011 were selected using a tag-and-probe technique, and the 2012 version of the trigger was emulated. The BDT medium offline identification was used. [45]

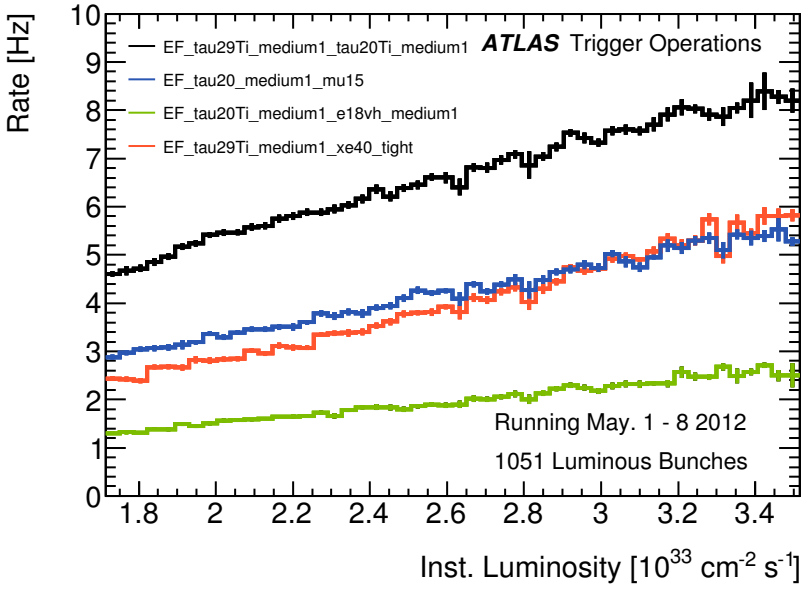


Figure 4.17: EF rate (frequency) as a function of instantaneous luminosity for a variety of tau triggers, with events collected from September to October 2012. [45]

4.7 Looking Forward

In order to reduce rates and improve the sensitivity of taus to new physics, hadronic tau triggers will continue to be combined with other signature triggers (e.g. electron, muon, missing E_T, a second hadronic tau), depending on the physics analysis. The benefits of such a strategy have already been shown: a tau + missing E_T trigger was used in tau polarization measurements in $W \to \tau\nu$ decays [48] as well as in the search for a charged Higgs boson [49]; a ditau trigger with asymmetric energy thresholds of 29 GeV and 20 GeV and medium ID was used in the search for the SM Higgs boson in the double hadronic decay channel [50]; a tau + lepton trigger was used in the $Z \to \tau\tau$ cross section measurement [51] as well as in the search for the SM Higgs boson in the single hadronic decay channels [50]. Furthermore, multivariate techniques, such as the one presented in the next chapter, will continue to be exploited in order to effectively maximize the discrimination power of single tau triggers.

At L2, there is the possibility of eliminating the two step (calorimeter + tracking) selection, assuming the currently developed ATLAS FastTracker (FTK) is built. The FTK is a hardware based track finder that can reconstruct all tracks with a transverse momentum greater than 1 GeV/c up to luminosities of $3 \times 10^{34}\,\mathrm{cm}^{-1}\mathrm{s}^{-1}$ at an event input frequency of 100 kHz and a latency of a few hundred microseconds. The plan is to have this system ready to use for the 2015 running period. [52]

Chapter 5

Procedure

5.1 Introduction

While the tau trigger has performed remarkably well in all respects, the success of multivariate techniques at the offline level has sparked an interest in attempting to apply these techniques to the trigger level. Not only would this raise the possibility of increasing the efficiency of the selection of taus, but it would also help to synchronize the work done offline with the work done at the three stages of the trigger. For the tau trigger, the currently implemented multivariate techniques for the offline identification of taus are BDTs and LLH. Both of these techniques have recently been added to the EF stage of the tau trigger for operation during the 2012 8 TeV running period. The LLH technique defines a likelihood function, for both signal and background, that maps the values of a set of discrimination variables into a probability density for the occurance of an event with those values given the signal or background hypothesis. The conditional probability density functions are based on the empirical distributions of the data, and the value of the LLH classifier is the logarithm of the ratio of the likelihood for signal to the likelihood for background. [46] Here we describe a BDT-based algorithm designed to be implemented at the L2 stage of the tau trigger. The software used for creating the BDTs is *the Toolkit for Multivariate Data Analysis with ROOT* (TMVA) [53].

5.2 Boosted Decision Trees

5.2.1 Decision Trees

BDTs build upon the concept of a decision tree. A *decision tree* is a binary tree structured classifier that performs a "yes" or "no" decision at each node depending on the value of a single variable, until a leaf node is reached. The events that end up in a "yes" leaf node are classified as "signal", while the others are classified as "background". The number of variables to look at is a parameter n for the decision tree and each event that passes through the classifier has the values for these n variables precomputed. At each node, one of the n variables is read and compared to a cutoff value, and depending on whether or not the variable is less than the cutoff value, the decision "yes" or "no" is made, and the event is passed to the corresponding next node, which performs this decision based on another variable. Figure 5.1 shows a schematic diagram of a decision tree.

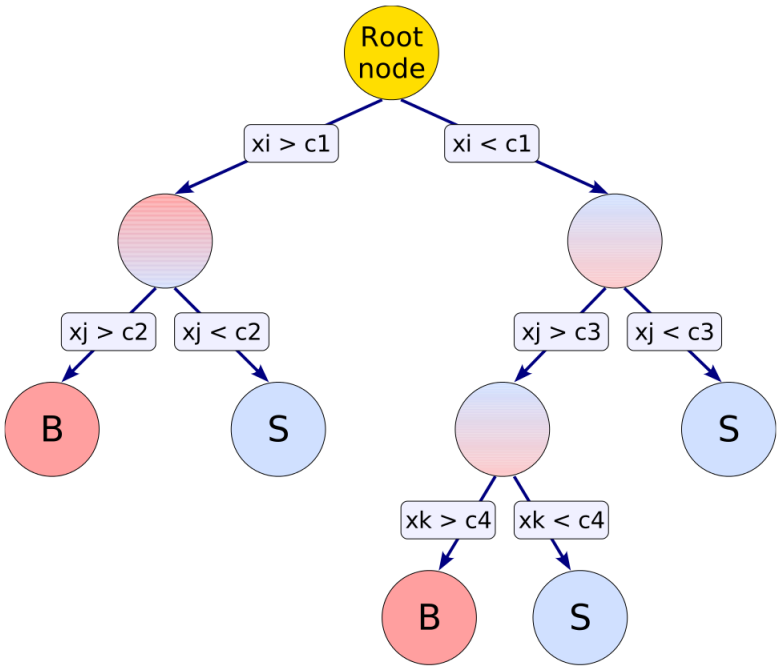


Figure 5.1: A schematic diagram of a decision tree. The case where a variable's value equals the cutoff is not shown, but can be chosen arbitrarily without loss of generality. [54]

A decision tree can be created by *training* it with samples of events that are each already

known to be of type signal or background. In TMVA, the training program takes as input a sample of events, with each event prelabeled as either signal or background, places the sample into the root node, and calculates a quantity known as the *separation* of the node. It then scans over each of the n variables, and for each variable, it scans over all cut values for the variable (within some limit of precision), splits the node into two temporary nodes according to the cut criteria, and calculates the event-weighted average separation of the temporary nodes. It compares this quantity with the separation of the root node, and for each variable, it chooses the cut that maximizes the difference between the separation of the root node and the average separation of the temporary nodes. Then, it compares the maximum separation difference of each variable, and chooses the variable that provides the greatest separation difference. This variable together with the chosen cut on the variable define the splitting decision at the root node of the decision tree. The nodes below the root node are split recursively in the same manner, until some halting condition is met. This may be that the total number of nodes has reached a certain value, or that the depth of the tree has reached a certain value, or that the number of events in a given node falls below a certain value. The *purity* of a node is defined as S/N where S is the number of signal events in the node and N is the total number of events in the node. The leaf nodes with purity greater than the `NodePurityLimit` parameter are classified as signal, while the others are classified as background. Each event may be weighted according to a custom formula, so that events with a higher weight are more likely to be properly classified.

In TMVA, there are five different options for how to define the separation:

- *Gini index*, defined by $p(1-p)$,
- *Gini index with Laplace's correction*, defined by $(SB+S+B+1)/(S+B+2)^2$,
- *Cross entropy*, defined by $-p\log_2(p)-(1-p)\log_2(1-p)$,
- *Misclassification error*, defined by $1-\max(p,1-p)$,
- *Statistical significance*, defined by $S/\sqrt{S+B}$,

where S is the number of signal events in the node, B is the number of background events in the node, and p is the node's purity.

5.2.2 Pruning

After a decision tree has been created, it may be pruned in order to remove statistically insignificant nodes. The two methods used in TMVA for pruning a decision tree are *expected error pruning* and *cost complexity pruning*.

The expected error method removes all subtree pairs for which the error-adjusted misclassification rate of their parent node is smaller than their combined error-adjusted misclassification rate. If p is the node's purity, the misclassification rate of a node is defined as $1-p$ if p is greater than `NodePurityLimit`, and p otherwise. The error-adjusted misclassification rate of a node is defined by

$$\mu_a = \min(1, \mu + s\sqrt{p(1-p)/N}) \,, \tag{5.1}$$

where μ is the node's misclassification rate, p is its purity, N is the number of events it contains, and s is the `pruneStrength` parameter. If the subtrees are leaves, the combined error-adjusted misclassification rate of the left and right subtrees of a node is defined as

$$\frac{N_L * r_L + N_R * r_R}{N_L + N_R} \,, \tag{5.2}$$

where N_L is the number of nodes in the left subtree, r_L is the error-adjusted misclassification rate of the left subtree, N_R is the number of nodes in the right subtree, and r_R is the error-adjusted misclassification rate of the right subtree. For more general subtrees, this combined rate is calculated in the same way, except that r_L and r_R are obtained through recursion. Note: In TMVA, with this pruning method, when a node is marked for pruning, it still counts as an existing node when later determining whether ancestors of the node should be pruned. We modified this behavior so that nodes that are marked for pruning are not counted when calculating the combined error-adjusted misclassification rate of subtrees rooted at their ancestors.

The cost complexity method removes all subtree pairs for which the cost-complexity index of its parent node is below a certain threshold. The cost-complexity of a node is defined by

$$\alpha = \frac{R(\text{node}) - R(\text{subtree below that node})}{\#\text{leaves(below that node)} - 1}\,, \tag{5.3}$$

where $R(\text{node}) \coloneqq N\delta$ such that N is the number of events in the node and δ is the separation of the node, and $R(\text{subtree})$ is defined recursively as $R(\text{left subtree}) + R(\text{right subtree})$ such that the subtree of a leaf is the leaf itself. The algorithm first computes α for each node, and then temporarily prunes each node by order of increasing α. Each time a subtree pair is temporarily pruned, α is recalculated for the nodes above it. The algorithm continues the temporary pruning until there are no more internal nodes remaining. If `pruneStrength` is set to a nonnegative value, and n is the number of subtree pairs that were temporarily pruned, the first $k-1$ of these subtree pairs actually get pruned, where $k =$ `int(PruneStrength/100.0*n)`. If `pruneStrength` is negative, the algorithm additionally calculates a measure of quality for the temporarily pruned tree each time a subtree pair is temporarily pruned. The quality is defined recursively as the sum of the qualities of the left and right subtrees, where the quality of a leaf node is defined as

$$q = \frac{p * B + (1-p) * S}{N}\,, \tag{5.4}$$

where p is the node's purity, B is the weighted total number of background events in the node, S is the weighted total number of signal events in the node, and N is the weighted total number of events in the tree. After the algorithm finishes temporarily pruning the tree, it finds the index k of the ordered subtree pairs that were temporarily pruned that corresponds to the minimum tree quality, and prunes the first $k-1$ subtree pairs.

5.2.3 Boosting

Boosting is a process that greatly enhances the performance of typically weak multivariate methods, by sequentially applying a multivariate algorithm to weighted (boosted) versions of the training sample. The two different methods for boosting a decision tree in TMVA are the *adaptive boost* algorithm and the *gradient boost* algorithm.

In the adaptive boost algorithm, events that were used to train the decision tree and were misclassified are given a higher event weight in the training of the following tree. The weight given is

$$\omega \coloneqq \left(\frac{1-\mu}{\mu}\right)^{\beta}, \tag{5.5}$$

where μ is the misclassification rate of the tree, defined as the weighted sum of the misclassified events in the tree divided by the total number of events in the tree, and β is the `AdaBoostBeta` parameter. The weights of the entire training sample are then renormalized so that their sum remains the same. The number of trees that are created by this process can be set by defining the variable `NTrees`. The *classifier* for a decision tree is a function h that maps an n-tuple of values for each of the variables, $\boldsymbol{x}$, into $\{-1, 1\}$ or $[0, 1]$, depending on the setting of `UseYesNoLeaf`. If `UseYesNoLeaf` is `true`, then the function maps into $\{-1, 1\}$ such that the value -1 means that the event is classfied as background and the value 1 means that it is classified as signal. Otherwise, the function maps into $[0, 1]$ where its value is equal to the purity of the resulting leaf node. If M is the number of trees that were created in the boosting process, and for each tree indexed by m, ω_m is the boost weight given to the tree and h_m is the tree's classifier, then the classifier for the *forest* of trees is the weighted average

$$H(\boldsymbol{x}) = \frac{1}{M}\sum_{m=1}^{M} \ln(\omega_m) h_m(\boldsymbol{x})\,. \tag{5.6}$$

As explained in Section 5.5, we only consider the adaptive boost algorithm as our boosting method, and hence, we omit the details of the other method.

A decision tree that has been trained on a boosted sample of events (and possibly pruned) is called a *boosted decision tree*. In TMVA, the variable `PruneBeforeBoost` controls whether it is pruned before boosting or not. Typically when we speak of BDTs, we mean a boosted sequence of BDTs, such that the corresponding BDT classifier is a weighted average of the classifiers for each BDT in the sequence. The value of a BDT classifier applied to a considered event is called the *BDT score* of the event. In order to use a BDT classifier, one typically chooses a cutoff for the BDT score, so that events with a BDT score greater than the cutoff are accepted, while otherwise, they are rejected.

5.2.4 Variable Importance

During the (TMVA) training of a BDT classifier, a useful measure of variable importance is calculated for each discriminating variable. For any discriminating variable x, the *absolute importance* of x for a decision tree is defined as

$$I_x = \sum_{i=1}^{n} \delta_i N_i \,, \tag{5.7}$$

where the sum is over each node that splits with variable x, δ_i is the separation gain of the split, and N_i is the number of events in the node. The absolute importance of x for a BDT-classifier is defined as

$$\sum_{m=1}^{M} \ln(\omega_m) I_{m,x} \,, \tag{5.8}$$

where the sum is over each tree in the classifier, ω_m is the tree's boost weight, and $I_{m,x}$ is the importance of variable x in the tree. The *importance* is the absolute importance divided by the sum of the importances for each variable. In order to simplify a BDT classifier, one may remove variables of low importance (and retrain), since they should not significantly affect the classifier's performance as compared with the other variables. [54]

5.3 Overview of the Algorithm

The algorithm to be described is one that takes as input an L2 tau RoI and determines whether or not it contains a tau. Those RoIs which L2 thinks contain a tau are passed to the EF for further analysis. Since the current L2 tau trigger is divided into both calorimeter and tracking stages, the algorithm for tau selection must be constrained by this requirement as well. Instead of using simply one BDT classifier for the L2 RoIs, it uses two, where one uses only variables that were derived from the calorimeter, and the other uses both calorimeter and tracking derived variables. The four stages of the algorithm, in sequential order, are thus: Pre-BDT-1, BDT-1, Pre-BDT-2, and BDT-2. Each stage determines whether to accept or reject an RoI, and if it is accepted, it is passed (as input) to the next stage. The Pre-BDT-1 stage is a non BDT-based algorithm that performs simple cuts on calorimeter

variables in preparation for input to the BDT-1 stage. BDT-1 is the application of the BDT classifier that only uses calorimeter variables to make its decision. Pre-BDT-2 performs simple cuts on both tracking and calorimeter variables in preparation for input to the BDT-2 stage. Finally, BDT-2 is the application of the BDT classifier that uses both calorimeter and tracking variables to make its decision. We refer to the combination of Pre-BDT-1 and BDT-1 as the *calorimeter stage* and the combination of Pre-BDT-2 and BDT-2 as the *tracking stage*. The reason we have non-BDT stages is that we do not want our BDTs to be dependent on certain variables, such as E_T and η, that are not theoretically dependent on the presence of taus. Instead, we would like to control the cuts on these variables more directly, depending on the beam conditions.

The Pre-BDT-1 stage is currently set to require $|\eta| < 2.2$ and $E_{T,\text{med}} >$ `etCut`, where η is the pseudorapidity of the centre of the RoI, $E_{T,\text{med}}$ is defined as

$$\sum_{i \in C_{0.2}} E_i / \cosh(\eta)\,, \tag{5.9}$$

where $C_{0.2}$ is the set of cells in the calorimeter inside the $\Delta R < 0.2$ cone about the centre of the RoI, and E_i is the energy measured by i. `etCut` is optimized for the collision energy of the training sample, with higher energy collisions generally requiring higher values in order to keep the trigger rates low. The η requirement provides some assurance that the RoI is in a region of the detector that has a sufficiently high level of measurement accuracy. Pre-BDT-1 also calculates some calorimeter based variables for use in the BDT-1 stage. Pre-BDT-2 currently makes no cuts at all and simply calculates some tracking based variables for use in the BDT-2 stage. The currently active L2 tau trigger that we use to compare our trigger with follows `L1_TAU8` and has an overall calorimeter-stage *acceptance rate* (fraction of considered events that are accepted) of $(2.9368 \pm 0.0057) \times 10^{-1}$ (by testing on the representative sample described later). `L1_TAU8` is currently the lowest energy threshold L1 tau trigger, and outputs L2 tau RoIs with the most flexibility and the least bias (loosest cuts). For this reason, we focus on `L1_TAU8` derived RoIs and set the cutoff score for BDT-1 to be such that BDT-1 accepts as many events as possible with the limitation that the event-wise acceptance rate of the calorimeter stage is less than 0.29, with 3σ certainty. We

require a 3σ significance level because it is traditionally used as the threshold for publishing a preliminary scientific discovery. Note: We may use the terms *efficiency* and acceptance rate interchangeably, even when testing a non-signal sample.

The procedure detailed below was mostly implemented with the C++ and Python programming languages. The C++ code is written at a relatively low level and is used for most of the CPU intensive processes. Python is only used for the high level or non CPU intensive processes.

5.4 Data Preparation

Data is prepared in various stages starting from samples of ATLAS events in the D3PD format. Each D3PD sample is slimmed by selecting only ROOT branches that are needed for the analysis and skimmed to only include events that pass `L1_TAU8`, `L2_tauNoCut` and `EF_tauNoCut` (the `tauNoCut` trigger chain). The `tauNoCut` criterion simply allows all RoIs to pass the respective level. This ensures that the events are unbiased with respect to an L2 or EF tau selection algorithm, and essentially, the only requirement is that the events each pass `L1_TAU8`.

The resulting sample then passes through a selection process that begins by assigning a number called the *event type* to each event. Events numbered as multiples of 4 are given an event type of 2; otherwise if they are multiples of 2, they are given an event type of 1. The other events are given an event type of 0. Events of type 1 are to be used for training BDT-1, events of type 2 are to be used for TMVA testing BDT-1, and events of type 0 are to be used for other purposes. By TMVA testing we mean passing RoIs for TMVA to test alongside the training RoIs. Events for TMVA testing are also used for more general testing, as described later. Events of type 0 or type 2 are possible candidates for training BDT-2. Thus, for all events of type 0, for every 4th such event, we change its type to 4, and for every 2nd such event, we change its type to 3, and for all events of type 2, every 4th becomes of type 6 and every 2nd becomes of type 5. Later, when training BDT-1, events of type 1 will be used for TMVA training, and events of type 2, 5 or 6 will be used for TMVA

testing. When training BDT-2, events of type 3 or 5 will be used for TMVA training, and events of type 4 or 6 will be used for TMVA testing. Note: By the design of the ATLAS data acquisition system, adjacent events are assumed to be uncorellated between each other. Also, if the sample is not intended for training, then the event types will have no effect. Table 5.1 gives a summary of the event types with their respective proportions and uses.

Event type	**% of events**	**Use**
0	25	testing BDT-1
1	25	training BDT-1
2	12.5	TMVA testing BDT-1
3	12.5	testing BDT-1 & training BDT-2
4	12.5	testing BDT-1 & TMVA testing BDT-2
5	6.25	TMVA testing BDT-1 & training BDT-2
6	6.25	TMVA testing BDT-1 & TMVA testing BDT-2

Table 5.1: The event types and their respective proportions and uses.

The rest of the selection process depends on whether the sample is to be used as a signal sample or as a background sample, and whether it consists of real collision data or MC simulated data. For signal MC, we require the event to contain at least one good primitive vertex with at least four tracks. Further, we look for all L1 RoIs with $E_T > 8\,\text{GeV}$, match them to the corresponding L2 and EF RoIs for the `tauNoCut` chain, perform an EF-Offline $\Delta R < 0.2$ match, require the offline object to have $E_T > 15\,\text{GeV}$, $|\eta| < 2.2$, at least one track, and medium offline BDT identification, and finally, require an EF-Truth match such that the true tau object has visible $E_T > 10\,\text{GeV}$ and visible $|\eta| < 2.5$. The L2 RoIs with matched offline objects that passed the stated requirements, referred to as *clean* RoIs, are stored as leaves of vectors of floats in a ROOT `TTree` structure, with each event in the tree containing only RoIs that came from the same event. The other RoIs, referred to as *dirty* RoIs, are also stored as long as the event contains at least one clean RoI. The dirty RoIs are stored in a separate tree that has the same number of events as the tree for the clean RoIs, and is constructed so that if some dirty RoI belongs to event i of its tree, then the corresponding clean RoIs belong to event i of their tree. The selection for signal data is the same except that no truth matching is involved. For background MC, we simply select

all L2 RoIs for the `tauNoCut` chain that are matched with an L1 RoI with $E_T > 8\,\text{GeV}$. For background data, we perform the additional steps of matching the L2 RoIs to EF RoIs for the `tauNoCut` chain, requiring an EF-Offline jet $\Delta R < 0.2$ match, such that the jet has $p_T > 15\,\text{GeV/c}$, $|\eta| < 2.2$, and there must be another offline jet separated from the offline jet by $|\Delta\phi| > 2.7$ and satisfying the same p_T and η requirements. In this case the clean RoIs are defined as the offline jet matched RoIs whose corresponding offline jets passed the stated requirements, while the others are the dirty RoIs. Back-to-back jets is a key characteristic of the hadronization process resulting from the proton-proton hard scattering process, and so it is highly unlikely that we would select tau jets with this criteria. These requirements are not necessary for background MC, since this data is created exclusively from QCD dijet processes. A summary of the selection criteria is given in Table 5.2, with the technical details shown in Appendix B.

The result of this selection process becomes the input to the Pre-BDT-1 stage. After performing its selection, the Pre-BDT-1 stage divides the sample into three parts: RoIs for TMVA training (TRN), events for testing (TST), and RoIs for TMVA testing (TRT). The TRN RoIs must be clean and each of their variables are simply stored as `TTree` leaves of floats. In order to prevent biasing our results by testing on events that were used for training, the TRN RoIs must come from separate events than those used for TST. The TST events remain in the leaves-of-vectors format so that the event-wise acceptance rates can be calculated. The TRT RoIs come from a subset of the TST events, must be clean, and are stored in the same format as the TRN RoIs. They are used as the members of the testing sample of the TMVA training program, which is used by TMVA to automatically perform various tests of the trained classifier that we may later use for analyzing our results. If the sample is to be used for training, the splitting is generally determined by sending the events of type 0, 3 or 4 into the TST sample, sending the events of type 2, 5 or 6 to both the TRT sample and the TST sample, and sending the remaining events (type 1) to the TRN sample. Otherwise, all of the events are passed into the TST sample. The TST sample is the input to the BDT-1 stage, which stores its output in the same data format as that for

TST, but only includes RoIs that passed the stage's requirements. The resulting sample becomes the input to the Pre-BDT-2 stage, which splits the events as does Pre-BDT-1, except that now events of type 0 or 2 are sent to the TST sample, events of type 4 or 6 are sent to both the TRT sample and the TST sample, and events of type 3 or 5 are sent to the TRN sample. BDT-2 then takes as input the TST output of Pre-BDT-2, and stores the resulting sample in the same format as does BDT-1. During each stage, histograms for various trigger variables and overall acceptance rates are put into storage for later analysis. Figure 5.2 shows a diagram of the data preparation procedure.

Sample type	**Selection criteria**
Signal MC	• one good primitive vertex with at least four tracks • offline tau $E_T > 15\,\text{GeV}$, $\|\eta\| < 2.2$ • at least one track, medium offline BDT ID • truth match with visible $E_T > 10\,\text{GeV}$ and visible $\|\eta\| < 2.5$.
Signal data	• one good primitive vertex with at least four tracks, • offline tau $E_T > 15\,\text{GeV}$, $\|\eta\| < 2.2$, • at least one track, medium offline BDT ID.
Background MC	None
Background data	• back-to-back jets with offline jet $p_T > 15\,\text{GeV/c}$, $\|\eta\| < 2.2$.

Table 5.2: The signal/background selection criteria. By default, the events come from the `tauNoCut` chain and the L2 RoIs must be matched to an L1 RoI with $E_T > 8\,\text{GeV}$ and an EF RoI from the same chain. Offline matching is performed with the EF RoIs and $\Delta R < 0.2$. More details are shown in Appendix B

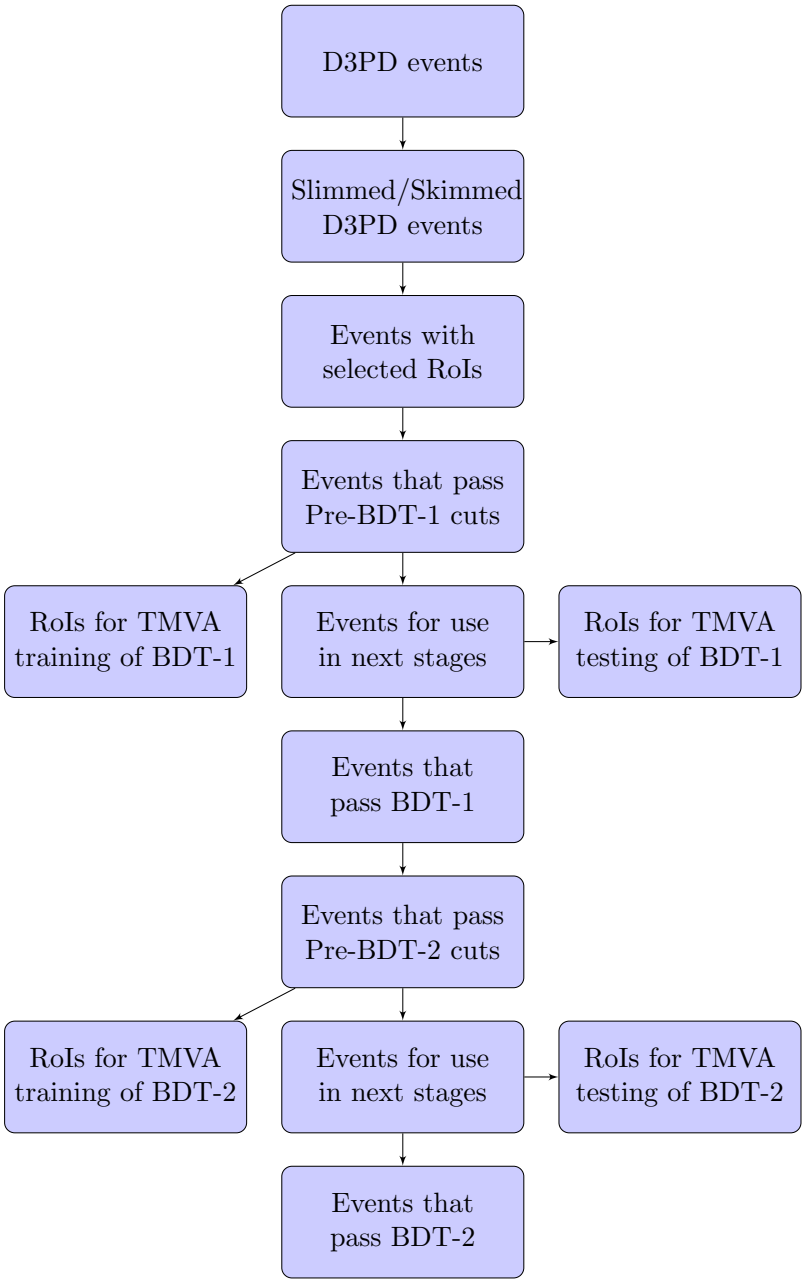


Figure 5.2: A diagram of the data preparation procedure. Note: By "events with selected RoIs" we mean a signal or background sample selected according to the criteria in Table 5.2, with the clean RoIs as the selected RoIs.

5.5 Training

The training of the BDT classifiers is performed by executing the TMVA training program for BDT-based classification. By virtue of the data selection procedure, the training sample for BDT-1 is the TRN output of Pre-BDT-1, and the training sample for BDT-2 is the TRN output of Pre-BDT-2. The weight for each RoI is set as the ratio of the cross section of the dataset it is a member of to the number of entries in the sample. The cross section of a dataset is typically taken from the ATLAS Metadata Interface (AMI) database [30]. Rather than allowing the TMVA Factory to normalize the events so that each event has an average weight of 1 (the default), no normalization is performed at all (`NormMode=None`). This results in a tree that is trained with a realistic proportion of signal to background events. One general principle used is that we do not add complexity where simplicity provides an insignificantly worse level of performance. For this reason we use the Gini index, the simplest measure of separation (a 2^{nd}-order polynomial function of a real-valued variable), as the separation index. We set `nCuts=-1`, so that the maximum level of precision is used optimize the cuts. The maximum depth of the tree is set to 6. This allows us to constrain the time complexity of the classifier to a reasonable amount, while not sacrificing any significant gain in discrimination power. We set `UseYesNoLeaf` to `false`, so that the classifier for each tree outputs the purity of the terminal node, which is more meaningful than simply a binary value. We prune each tree *before* it is boosted, so that statistically insignificant nodes are not counted in the misclassification rate. Using the simplicity principle, we use the expected error algorithm for pruning, and set `PruneStrength` to 1.0. Likewise, we use the adaptive boost algorithm for boosting and set `AdaBoostBeta` to 0.5. For similar reasons as for our choice of tree depth, the number of trees is set to 50. Depending on the specific conditions of our training sample, both the maximum tree depth and the number of trees may be lowered, as shown in the next chapter.

For the training of BDT-1, we consider 8 different calorimeter variables: `HADRadius`, `CaloRadius`, `CoreFrac`, `IsoFrac`, `EMFrac`, `stripWidth`, `HADtoEMenergy`, and `EnergyTonCells`. All of these variables are available in the Tau Working Group (WG) D3PDs, except for

HADtoEMenergy and EnergyTonCells, which are custom calculated for the purpose of training BDT-1. They were selected from the set of all calorimeter-derived L2 tau variables available in the Tau WG D3PDs [55] by removing those that are highly correlated with one of the others. Table 5.3 gives the definitions of each of these variables.

We consider 5 different tracking variables: nCoreTracks, trkAvgDist, etOverPtLeadTrk, scalarPtSumIsoToCore, and chargeTrans. All of these variables are available in the Tau WG D3PDs, except for scalarPtSumIsoToCore and chargeTrans, which are custom calculated for the purpose of training BDT-2. As before, they were selected from the set of all tracking-derived L2 tau variables available in the Tau WG D3PDs by removing those that are highly correlated with one of the others. Table 5.4 gives the definitions of each of these variables.

Variable	Definition	Description
`HADRadius`	$\frac{\sum_{i\in L_{4-6}\cup T} E_{0.2,i} R_i}{\sum_{i\in L_{4-6}\cup T} E_{0.2,i}}$	Energy-weighted hadronic radius $[\Delta R]$
`CaloRadius`	$\frac{\sum_{i\in L\cup T} E_{0.2,i} R_i}{\sum_{i\in E\cup T} E_{0.2,i}}$	Energy-weighted calorimeter radius $[\Delta R]$
`CoreFrac`	$\frac{\sum_{i\in L\cup T} E_{0.1,i}}{\sum_{i\in L\cup T} E_{0.2,i}}$	Fraction of energy that is in the core region
`IsoFrac`	$1 - \frac{\sum_{i\in L_2} E_{0.1,i}}{\sum_{i\in L_2} E_{0.2,i}}$	Fraction of energy in the 3$^{\text{rd}}$ layer of the EM calorimeter that is in the isolation region
`EMFrac`	$\frac{\sum_{i\in L_{0-3}} E_{0.2,i}}{\sum_{i\in L\cup T} E_{0.2,i}}$	Fraction of energy that is in the EM calorimeter
`stripWidth`	$\sqrt{\frac{\sum_{i\in L_{2,6}} \eta_{0.2,i}^2 E_{0.2,i}}{\sum_{i\in L_{2,6}} E_{0.2,i}} - \left[\frac{\sum_{i\in L_{2,6}} \eta_{0.2,i} E_{0.2,i}}{\sum_{i\in L_{2,6}} E_{0.2,i}}\right]^2}$	Energy weighted standard deviation of cell η
`HADtoEMenergy`	$\frac{\sum_{i\in L_{4-6}\cup T} E_{0.2,i}}{\sum_{i\in L_{0-3}} E_{0.2,i}}$	HAD to EM energy *
`EnergyTonCells`	$\frac{\sum_{i\in L\cup T} E_{0.2,i}}{\sum_{i\in L\cup T} 1}$	Energy to number of cells [MeV] *

Table 5.3: The definitions of the variables used for the training of BDT-1. For $j \in 0, 1, 2, 3$, L_j is the set of all activated cells in the $j + 1^{\text{th}}$ layer of the (LAr) EM Calorimeter. For $j \in 4, 5, 6$, L_j is the set of all activated cells in $j - 3^{\text{th}}$ layer of the LAr hadronic calorimeter. For $j \in 0, 1, 2$, T_j is the set of all activated cells in the $j + 1^{\text{th}}$ layer of the tile (hadronic) calorimeter. $L \coloneqq \cup_{l=0}^{6} L_j$ and $T \coloneqq \cup_{l=0}^{2} T_j$. For each $\alpha \in \mathbb{R}_+$ and $i \in C \subset L \cup T$, if i is in the $\Delta R < \alpha$ cone about the centre of the RoI, then $E_{\alpha,i}$ is the energy measured by i and $\eta_{\alpha,i}$ is the pseudorapity of i; otherwise, $E_{\alpha,i} = 0$ and $\eta_{\alpha,i} = 0$. R_i is the ΔR distance between i and the centre of the RoI. (continued on next page)

Table 5.3: (continued) A division by 0 typically results in an error code, as described in Section 5.8. The units for the variables are shown in square brackets. [56]
* The precise definitions of these variables are shown in Section C.1.

Variable	**Definition**	**Description**
`nCoreTracks`	$\sum_{i \in R_{0.1}} 1$	Number of core tracks
`trkAvgDist`	$\frac{\sum_{i \in \bar{R}_{0.3}} p_{T,i} \bar{R}_i}{\sum_{i \in \bar{R}_{0.3}} p_{T,i}}$	Transverse-momentum-weighted track radius [ΔR]
`etOverPtLeadTrk`	$\frac{E_{T,\mathrm{med}}}{p_{T,\mathrm{lead}}}$	Medium transverse energy to momentum of leading track [c]
`scalarPtSumIsoToCore`	$\frac{\sum_{i \in R_{0.3}} p_{T,i}}{\sum_{i \in R_{0.1}} p_{T,i}}$	Sum of isolated transverse momenta to sum of core transverse momenta *
`chargeTrans`	$\|\|q\| - 1\|$	Transformed charge [e]

Table 5.4: The definitions of the tracking-derived variables used for the training of BDT-2. Due to the 2012 improvements, all tracks used for these measurements must have $p_T > 1\,\mathrm{GeV/c}$ and $|\Delta z_0| < 2\,\mathrm{mm}$ between themselves and the leading (p_T) track for the RoI. L, T, and $E_{T,\alpha,i}$ are as defined in the caption of Table 5.3. For each $\alpha \in \mathbb{R}_+$, R_α is the set of tracks contained in the $\Delta R < \alpha$ cone about the centre of the leading track, and for each $i \in R_\alpha$, $p_{T,i}$ is the transverse momentum of i and R_i is the ΔR distance between the centre of i and the centre of the leading track; $p_{T,\mathrm{lead}}$ is the transverse momentum of the leading track. For each $\alpha \in \mathbb{R}_+$, $\bar{R}_\alpha$ is the set of tracks contained in the $\Delta R < \alpha$ cone about the centre of the RoI, and for each $i \in R_\alpha$, $p_{T,i}$ is the transverse momentum of i and $\bar{R}_i$ is the ΔR distance between the centre of i and the centre of the RoI. q is the net electromagnetic charge of the tracks. The units for the variables are shown in square brackets. [57][58]
* The precise definition of this variable is shown in Section C.2

5.6 Testing

The main purpose of the algorithm is to select as much signal as possible while keeping within the trigger rate limitations. By rate we mean the average acceptance rate for all types of events that our trigger may see. To determine whether our algorithm is worthy of implementation, we compare its signal acceptance rates with those of the currently implemented simple cut-based L2 tau trigger that follows the same L1 output as our trigger, and that we use to compare with our trigger. We call this the *reference trigger*. As already mentioned, the event-wise rate limitation for the calorimeter stage, based on the reference trigger, is set to 0.29, with a 3σ certainty threshold. We want the rate for this stage to be as close as possible to its limit, since it is not until after this stage that we can make use of the discrimination power of both the calorimeter and tracking variables, instead of simply the calorimeter variables. For the combined calorimeter and tracking stages, the reference trigger has an event-wise acceptance rate of $(8.673 \pm 0.035) \times 10^{-2}$. Therefore, we set the event-wise rate limitation for the overall algorithm to 0.085, again with a 3σ certainty threshold. We also examine other values for the tracking-stage rate limitation in order to see the relationship between signal acceptance and overall rejection.

At each of the four stages, we record various types of acceptance rates for each sample of events that is used for testing (rather than training). The acceptance rates can be divided into two types: event-wise acceptance rates and RoI-wise acceptance rates. An event is accepted if at least one RoI (clean or dirty) is accepted. The *event-wise* acceptance rate is defined as the ratio of the number of accepted events to the number of events considered. For the RoI-wise acceptance rate we consider only the clean RoIs, so that the *RoI-wise* acceptance rate is defined as the ratio of the number of accepted clean RoIs to the number of clean RoIs considered. For each of these types of acceptance rates, we analyze their dependence with respect to pile-up by filling a histogram with a measure of pile-up (average number of interactions per bunch crossing during the event) each time an event is accepted or a clean RoI is accepted. We also want to determine the E_T dependence of our acceptance rates. For RoI-wise rates, we fill the histograms with the RoI's value of the L2 E_T variable,

while for the event-wise rates, we fill the histograms with the sum over the L2 E_T values of each RoI (clean or dirty), and note that this sum is precalculated before any part of the algorithm has the chance to eliminate some RoIs from an event. We also fill the same kind of histograms when selecting a signal or background sample that becomes an input to Pre-BDT-1, so that we can later calculate efficiencies for these variables.

In addition to acceptance rates, we must also keep track of the statistical errors of those acceptance rates. In general, our assumption is that the number of events or RoIs, N, passing a certain condition is poisson distributed so that the standard deviation of N is $\sqrt{N}$. When taking the ratio, we use the Bayesian-based approach described in [59], so that if N_1 is the number of considered events or RoIs with standard deviation $\sigma_1 = \sqrt{N_1}$ and N_2 is the number of selected events or RoIs with standard deviation $\sigma_2 = \sqrt{N_2}$, then the acceptance rate, $\rho = N_2/N_1$ has standard deviation

$$\sigma = \sqrt{\frac{(N_2+1)(N_1-N_2+1)}{(N_1+2)^2(N_1+3)}}\,, \tag{5.10}$$

and the estimator for the mean is

$$\rho = (N_2+1)/(N_1+2)\,. \tag{5.11}$$

For histograms, we use ROOT's `TGraphAsymmErrors->Divide()` method to divide the histograms with the option parameters `"cl=0.683␣b(1,1)␣mode"`. This uses standard Bayesian techniques to calculate the error bars corresponding to a confidence level of 0.683 (1σ). If we wish to calculate the overall acceptance rate, ρ, for a sequence of two stages of the algorithm, we may need to multiply the acceptance rate for one stage, $\rho_1 \pm \sigma_1$ with the acceptance rate for the other stage, $\rho_2 \pm \sigma_2$, to obtain $\rho = \rho_1\rho_2$ with standard deviation calculated as

$$\sigma = \rho\sqrt{\sigma_1^2/\rho_1^2 + \sigma_2^2/\rho_2^2}\,. \tag{5.12}$$

(standard error propagation). If the error bars are asymmetric, then we apply the above equation to the upper and lower error bars separately. The multiplication of acceptance rates is frequently used when generating μ or E_T dependent efficiency plots from the μ

or E_T dependent histograms for samples that are used for training a BDT classifier. This is because some of the events that appear in the histogram for one stage may be used as events for training a BDT at a later stage, and when testing that BDT, those events will not be considered as input. We also use standard error propagation when calculating averages of acceptance rates. For each signal or background sample, an effective cross section is calculated by multiplying the sample's cross section with an estimator for the fraction of events that are accepted during the signal/background selection process. The estimator and its uncertainty are simply given by (5.10) and (5.11). The cross section is taken as a parameter with no uncertainty, so the standard deviation for the effective cross section is simply the standard deviation of the estimator multiplied by the cross section. The effective cross sections are used as the weights for the average acceptance rates.

When plotting the pile-up dependence of an acceptance rate, we also perform a linear fit to the data using `TGraph`'s `Fit` method with the `"pol1"` parameter. This will automatically find a best linear fit to the data using the default χ^2-based fitting algorithm, and return the numerical value of the slope and error bar for the slope of the fitted line.

Before we can test the acceptance rate of a BDT classifier (BDT-1 or BDT-2), we must choose a cutoff score. If we want a rate that is less than but as close as possible to the value ρ_0, then we scan a discrete set of cutoff scores, and for each cutoff score, test the acceptance rate $\rho \pm \sigma$ of the classifier on a representative sample of events. If ρ is sufficiently close to ρ_0 and $\rho + 3\sigma < \rho_0$, then we record the cutoff score for later use.

Finally, it is important to ensure that the selection algorithm can be performed within the timing constraints of the L2 tau trigger system. Due to our choice of a maximum tree depth of 6 and a maximum number of trees of 50, each execution of the algorithm will require at most $6 \times 50 = 300$ variable comparisons, 50 floating point multiplications (by boost weights), and the addition of 50 floating point numbers (contribution from each tree). We believe that, compared to the calorimeter and track reconstruction performed at L2, the time require to perform these steps is quite insignificant.

5.7 Reference Trigger

The reference trigger that we use is `L2_TAU20_MEDIUM1`. It is one of the two main triggers that are used in conjunction with `L1_TAU8`; the other being `L2_TAU20_MEDIUM`. The only difference between the `MEDIUM` trigger and the `MEDIUM1` trigger is that `MEDIUM` requires less than eight core tracks while `MEDIUM1` requires less than five [60]. We choose `L2_TAU20_MEDIUM1` because the `MEDIUM1` requirement is currently being used for the search for the SM Higgs boson. [61] The reference trigger is applied to our signal and background samples in order to compare its performance with the performance of our BDT-based trigger. The result is a set of analogous rates, histograms, and efficiency graphs as described above.

The calorimeter stage for the reference trigger uses only two variables: `etMedium` ($E_{T,\text{med}}$) and `CoreFrac`. The requirement is that `etMedium` must be greater than 15.199 GeV and `CoreFrac` must be greater than 0.7485. At the tracking stage, the trigger uses the variables `nCoreTracks`, `scalarPtSumIso`, `scalarPtSumCore`, `etOverPtLeadTrk`, and `trkAvgDist` to make the final selection. The precise algorithm for this trigger, for any input RoI, is shown in Section C.3.

5.8 Error Codes

Some of the variables used for splitting the decision trees contain so-called error code values. These are values that are, in a sense, unphysical, discontinuous from the rest of the distribution, and typically occur in exceptional cases where the variable cannot be calculated from the information in the RoI. For example, if a variable is the ratio of two variables, it cannot be calculated when the variable in the denominator is 0; so in this case, we may set its value to an error code. To deal with error codes, we train events (RoIs) with error codes separately from those without error codes. Further, we divide the events with error codes into different classes depending on which variables in the event have an error code, and train each of these seperately as well. For some error code classes of events, there may be an insufficient amount of statistics to properly train a BDT or say anything significant

about its acceptance rate during the testing phase, and for such cases, we automatically reject the events. The goal is to be robust with respect to different error code classes. We want each class to have acceptance rates that are below the maximum limits, and we want to maximize the signal acceptance for each class given that limitation. The acceptance rates are tuned by choosing an appropriate cutoff score for each error code class (for both BDT-1 and BDT-2), and possibly adjusting the cutoff scores to ensure an appropriate acceptance rate for the overall algorithm applied to all of the events. The BDT classifier for each class is trained by restricting its training sample to only events of that class and removing the variables with error codes from the list of variables used. An alternative way to train such a BDT classifier would be to only restrict the training samples to the events that do not contain error codes for the variables in this shortened list, and this would have the advantage of introducing more statistics into the training program. We decided to proceed with the former method, since a BDT classifier typically performs better on continuous distributions of variables, and when events of different error code classes were mixed together, we observed the presence of significant discontinuities. For example, for a few of the samples we used, whenever the variable `etOverPtLeadTrk` has an error code, all of the other tracking variables are 0-valued.

After examining the training samples, we determined that there are 8 variables that are prone to error codes, and at most one error code class per variable was found. The variables with error codes are: `HADRadius` (-99), `CoreFrac` (-99), `IsoFrac` (-99), `stripWidth` (-99), `CaloRadius` (-99), `EMFrac` (-99), `trkAvgDist` (-111), and `etOverPtLeadTrk` (-111).

When calculating an RoI-wise acceptance rate for an error code class, we consider only RoIs from that class as input. When calculating an event-wise acceptance rate for an error code class, we consider only events with at least one RoI being a member of that class as input, and events with at least one such RoI being accepted as output for acceptance. The $\sum E_T$ variable for the event-wise E_T dependent histograms and efficiencies for an error code class is calulated by only adding the E_T of RoIs from that class. Also, when creating a signal or background sample with a specific error code class, the event type is determined

before selecting events containing RoIs from the error code class.

5.9 Systematic Uncertainties

The specific algorithm that a BDT-classifier uses depends on the training sample it is given. Changing the training sample can have an effect on performance measures such as the acceptance rates for our algorithm. This effect results in a systematic uncertainty due to our training procedure. To measure this effect, we apply our training procedure for three different cases for the training conditions: (1) the entire training sample for both BDT-1 and BDT-2, as determined from the event types, is used for training; (2) only the even numbered halves of the training samples of BDT-1 and BDT-2 are used for training; and (3) only the odd numbered halves of the training samples of BDT-1 and BDT-2 are used for training. Note: If a part of a training sample is not used for training, then it is not even used for testing. This is to ensure that there is no correlation between cases (2) and (3). When showing the results, *BDT trigger* will refer to our trigger trained under case 1, *BDT even trigger* will refer to our trigger trained under case 2, and *BDT odd trigger* will refer to our trigger trained under case 3. We define the *systematic uncertainty* of an acceptance rate as the absolute difference between the acceptance rates for the even and odd triggers. We write the measured value of an acceptance rate as

$$\rho \pm \sigma_{\text{stat}} \pm \sigma_{\text{sys}} \,, \tag{5.13}$$

where ρ is the estimator for the mean of the acceptance rate, σ_{stat} is the statistical uncertainty of ρ, and σ_{sys} is the systematic uncertainty of ρ. Evidently, there are more sources of systematic uncertainty than the one we use, but in order to simplify the analysis, we solely measure the systematic uncertainty due to this source.

5.10 Optimization

Typically, adding more variables to a BDT classifier causes it to perform better, at the cost of more training time and a need for more training statistics. But if BDT-1 output

is an input to BDT-2, adding more variables to BDT-1 does not necessarily increase the performance of the combined algorithm. Since BDT-2 has more information, it may be more effective overall to leave out some calorimeter variables from BDT-1.

For this reason, we scan over the combinations of variables for BDT-1, and for each tested combination, we compare the results with other combinations to determine whether there is any statistically significant gain in using the given combination. In order to make this computationally feasible, we perform the testing of each combination in parallel, using Python's `multiprocessing` library to manage the concurrent subprocesses (we also use concurrency for training and testing different error code classes at once). In addition, we set conditions in the algorithm that make it so that the algorithm does not necessarily scan over all possible combinations of the variables. If A and B are subsets of the set of all possible variables for BDT-1, such that $A \subset B$, and if B leads to significantly worse results than A, then we do not test any combinations that are supersets of B; otherwise, we test all combinations of the form $B \cup \{x\}$ where $x \notin B$.

Chapter 6

Results

6.1 Introduction

Here we show the results of applying the procedure described in the previous chapter. We begin with a specification of the training samples. Then, we specify the cutoff parameter for the Pre-BDT-1 stage and show the performance of this stage for these samples by plotting the E_T dependent efficiencies. Next, we list the variables used for the BDT stages. We compare their distributions for signal and background and show the relevant correlation matrices. The error code classes that each correspond to a trained BDT classifier are subsequently listed, together with an indication of the fraction of events that each class consists of. Next, we explain the concept of overtraining, and show the relevant overtraining plots for each error code class. We then look at the main result of the BDT training procedure: a set of boosted decision trees. We show a diagram of one representative example tree from each BDT stage, and list the importance indices for the classifiers that contain those trees. Next, we examine the efficiencies of the calorimeter stage of the algorithm by showing the pile-up dependent acceptance rate plots for each of the training samples. We also list the corresponding overall acceptance rates for the calorimeter stage. The overall algorithm is examined next by plotting the main signal acceptance versus overall rejection, for each of the principal error code classes. Next, we study the energy dependence of

the acceptance rates by showing the E_T dependent plots for each of the training samples. Similarly, the pile-up dependence is analyzed next by showing the pile-up dependent plots for the same samples, together with linear fits to the data shown. Finally, we iterate through the rest of the samples tested and show for each the relevant energy and pile-up dependent plots. These samples were not used for training and include both signal and background from real collision data, and signal from $W \to \tau\nu$, $Z \to \tau\tau$. $H \to \tau\tau$, and $Z' \to \tau\tau$ MC. Finally, the numerical values for the overall acceptance rates for all samples tested are summarized in a table. Note: Each time we show results for acceptance rates, we show both the RoI-wise and event-wise acceptance rates, and where possible, we show the corresponding acceptance rates for the reference trigger.

6.2 Training Samples

For the following results, we use the ATLAS datasets

- `mc12_8TeV.147818.Pythia8_AU2CTEQ6L1_Ztautau`
  `.merge.NTUP_TAU.e1176_s1479_s1470_r3553_r3549_p1130/`
- `mc12_8TeV.170204.Pythia8_AU2CTEQ6L1_Zprime1000tautau`
  `.merge.NTUP_TAU.e1176_s1479_s1470_r3553_r3549_p1130/`

to create the signal training samples, and the ATLAS datasets

- `data12_8TeV.period*.physics_JetTauEtmiss`
  `.PhysCont.NTUP_TAU.grp13_v01_p1130/`

to create the background training samples. The `*` character indicates a wildcard that ranges in $\{\texttt{B}, \texttt{C}, \texttt{D}\}$. Recall that the samples are created by applying the selections described in Table 5.2 to the (skimmed) datasets. By training sample we mean a set of events with a subset that is used for training a BDT classifier; other subsets of the sample are used for testing.

The first signal dataset is 2012 MC data consisting of $\mathrm{Z} \to \tau\tau$ events (with a cross section of 8.7804×10^{-1} nb), and the second one is 2012 MC data consisting of $\mathrm{Z}' \to \tau\tau$

events (with a cross section of 1.4845×10^{-4} nb) with the mass of the Z′ set to 1000 GeV [30]. The Z′ is a hypothetical particle with properties that are similar to those of the Z, but with a much greater mass [62]. Using both of these signals allows us to take into account a wide mass spectrum for the resonances that decay into the tau leptons, including the masses of particles that are yet to be discovered with higher collision energies. The background datasets consist of real collision data from the 2012 ATLAS data taking periods B-D. We do not include Period A because it was found to contain some anomalies as compared with the other periods, and since the other periods that we use contain a sufficient number of events, we decided not to investigate this further. We reserve Period E exclusively for testing. This allows us, in a way, to demonstrate the ability of our algorithm to perform well on the future beam conditions after training on past data. The rest of the periods are not included since they were not available for the particular version of D3PDs that we are using. Both the signal and background datasets correspond to 8 TeV $\mathrm{p}-\mathrm{p}$ collisions. Recall that we need the cross sections of the datasets in order to properly weigh the training samples. We use 8.7804×10^{-1} nb as the cross section for the $\mathrm{Z} \to \tau\tau$ dataset and 1.4845×10^{-4} nb as the cross section for the $\mathrm{Z}' \to \tau\tau$ dataset. Both of these values have been obtained by looking up the cross sections for the particular samples in the AMI database. A study performed by the TOTEM Collaboration determined that the luminosity-independent 8 TeV LHC total cross section is 101.7 ± 2.9 mb [63]. Thus, we use 1.017×10^{8} nb as the cross section for the background datasets.

For choosing the cutoff scores, we use as the representative sample the sample resulting from the union of the background datasets when *all* L2 RoIs seeded by `L1_TAU8` (and match `L2_tauNoCut` and `EF_tauNoCut`) are considered without restriction. This effectively simulates the average acceptance rate during real data taking, without making assumptions on the cross sections of various event classes. The representative sample does not include events that were used for training. This is done by performing the same event type selection that would be performed if the sample were used for training.

Table 6.1 shows, for each type of training sample, the number of events in its dataset (after skimming) and the number of events it consists of.

Sample	**No. of events in dataset**	**No. of events in sample**
$Z \to \tau\tau$ Signal	1744013	818242
$Z' \to \tau\tau$ Signal	386380	252425
Dijet Background	1304789	425431
Representative	1304789	1304555

Table 6.1: A summary, for each type training sample, of the number of events in its dataset and the number of events it consists of. Periods B-D are used for both the dijet background and representative samples. The number of events in a dataset is counted after skimming the dataset to only include events that pass the `tauNoCut` chain.

6.3 Pre-BDT-1

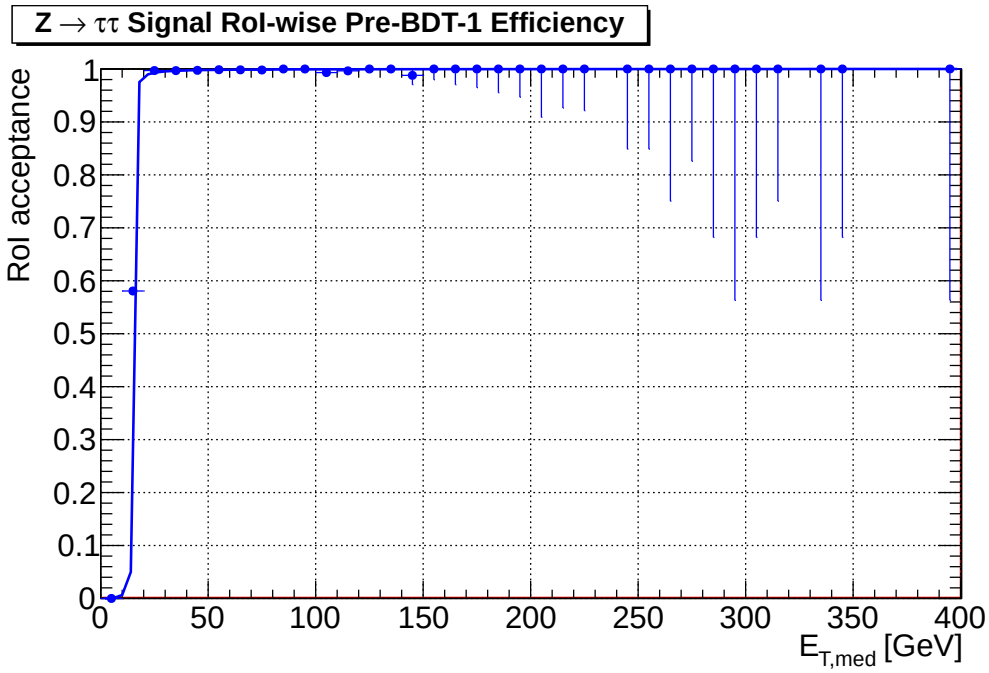
For Pre-BDT-1, we use 15.199 GeV as our `etCut`. This is the same value that the reference trigger uses for cutting on $E_{T,\text{med}}$, which makes for a fair comparison between the triggers. Now we show the Pre-BDT-1 acceptance rates for the training samples. Figure 6.1 shows the E_T dependent rates for the $Z \to \tau\tau$ sample, Figure 6.2 shows the E_T dependent rates for the $Z' \to \tau\tau$ sample, Figure 6.3 shows the E_T dependent rates for the dijet background sample, and Figure 6.4 shows the E_T dependent rates for the representative sample.

The curve used to fit to the data is the graph of the function

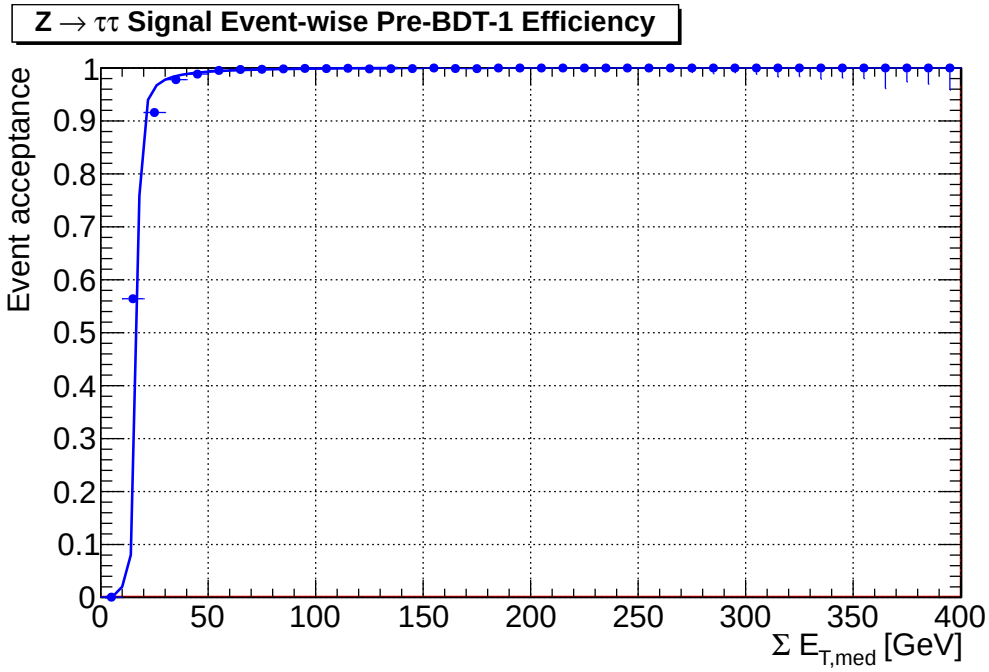
$$f(x) = C_0 \arctan\left(\frac{x - C_1}{C_2}\right) + C_3\,, \tag{6.1}$$

where x is $E_{T,\text{med}}$ or $\sum E_{T,\text{med}}$, and C_0, C_1, C_2, C_3 are real-valued parameters. This is the fitting function used for the official tau trigger efficiency measurements (where x is offline p_T) [26]. We use it because it models our data well and it allows us to be consistent with the methodology used by the tau trigger group. This type of curve is typically called a *turn-on* curve, a name derived from its shape [34].

The close agreement of the data with the turn-on curves shows that the rates follow the expected distributions, and in general, increase with increasing values of E_T. Recall that the background training sample is the subset of the representative sample that requires the RoIs to be matched to offline jet RoIs with $p_T > 15\,\text{GeV/c}$ and $|\eta| < 2.2$; hence, it is not surprising that it has a noticeably higher acceptance rate at this stage.

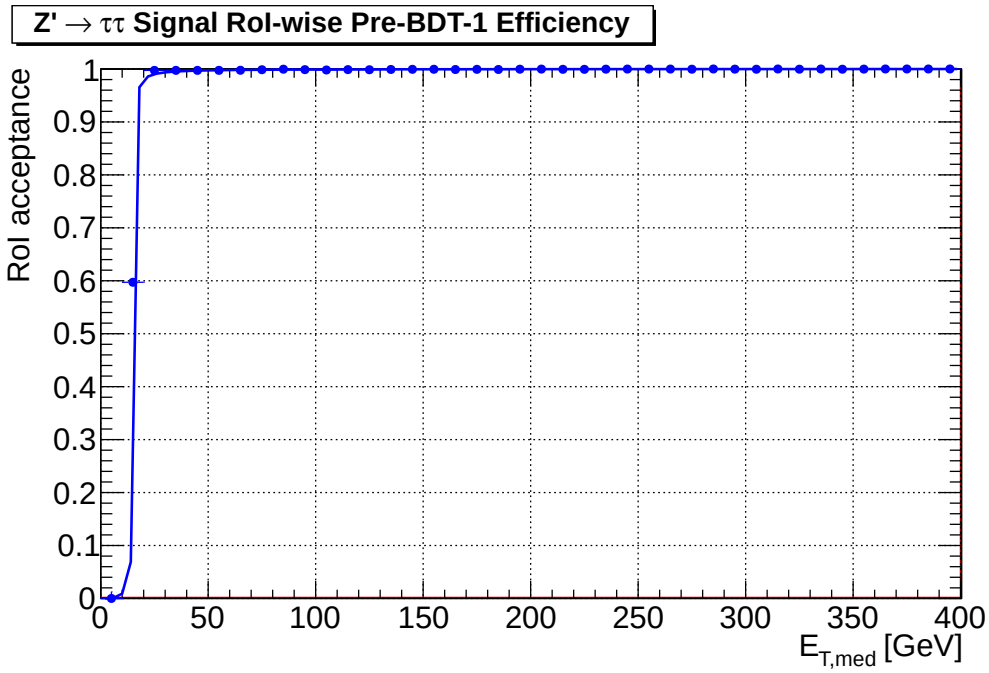


(a)

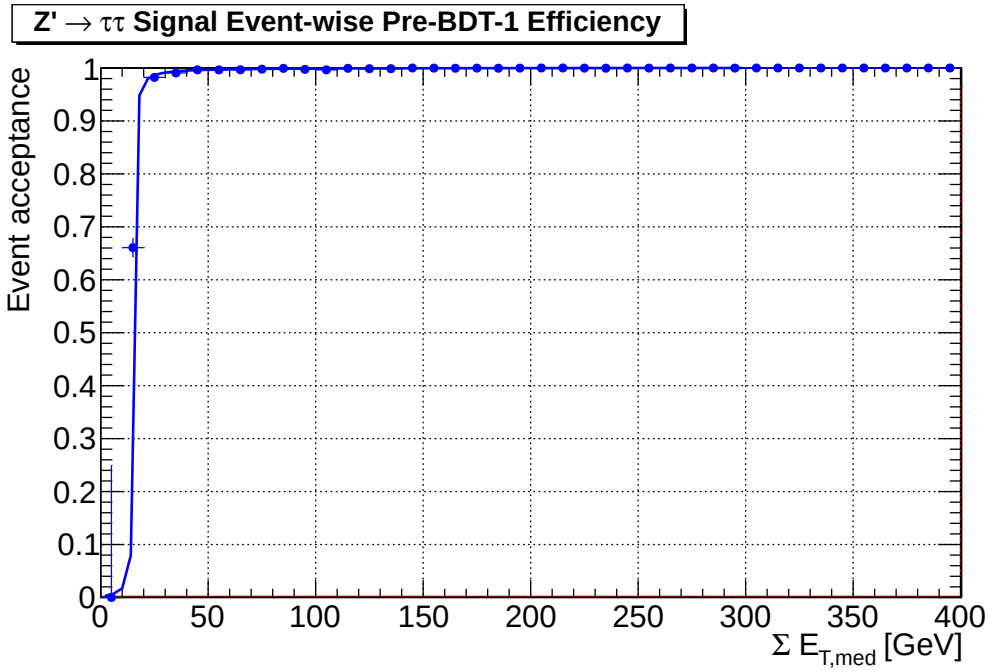


(b)

Figure 6.1: The Pre-BDT-1 acceptance rate for $Z \to \tau\tau$ signal for two different cases. (a) shows the RoI-wise acceptance rate w.r.t. L2 $E_{T,\text{med}}$, (b) shows the event-wise acceptance rate w.r.t. Σ L2 $E_{T,\text{med}}$.

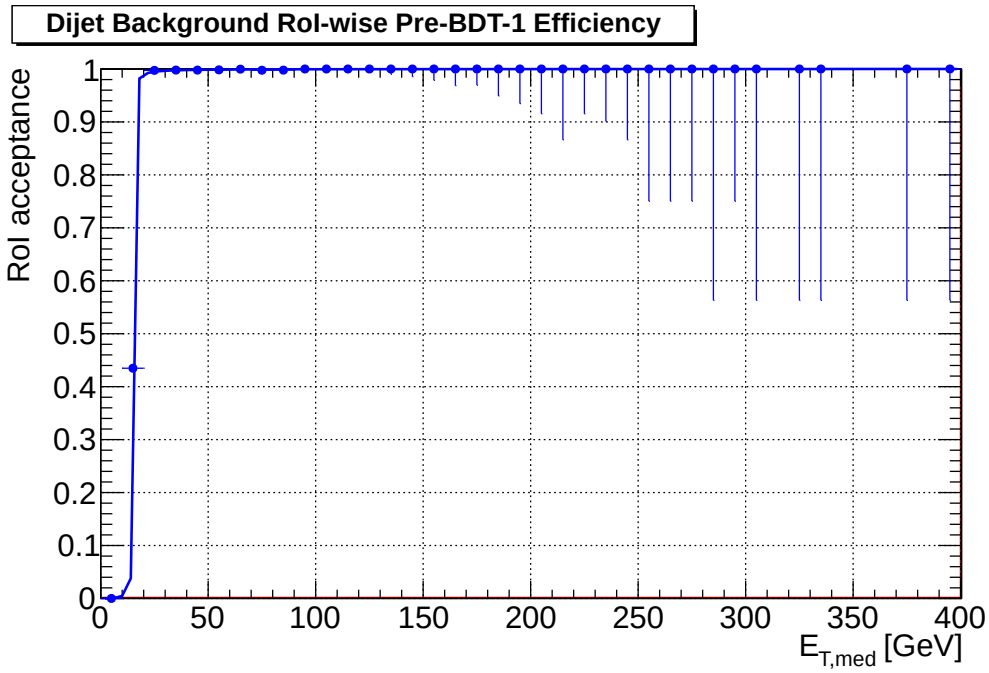


(a)

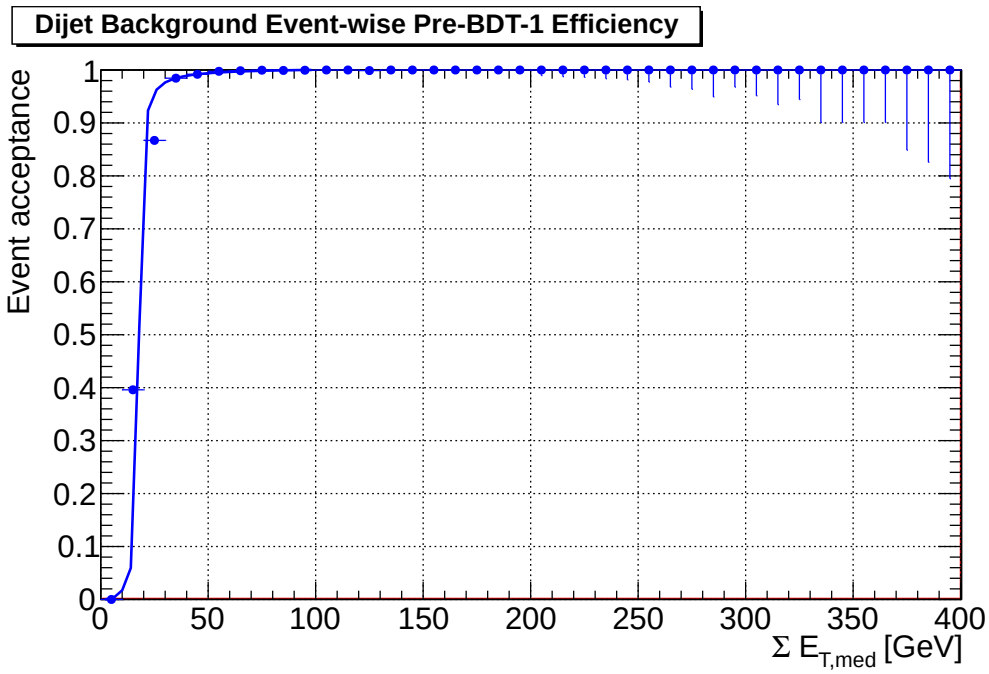


(b)

Figure 6.2: The Pre-BDT-1 acceptance rate for $Z' \to \tau\tau$ signal for two different cases. (a) shows the RoI-wise acceptance rate w.r.t. L2 $E_{T,\text{med}}$, (b) shows the event-wise acceptance rate w.r.t. Σ L2 $E_{T,\text{med}}$.

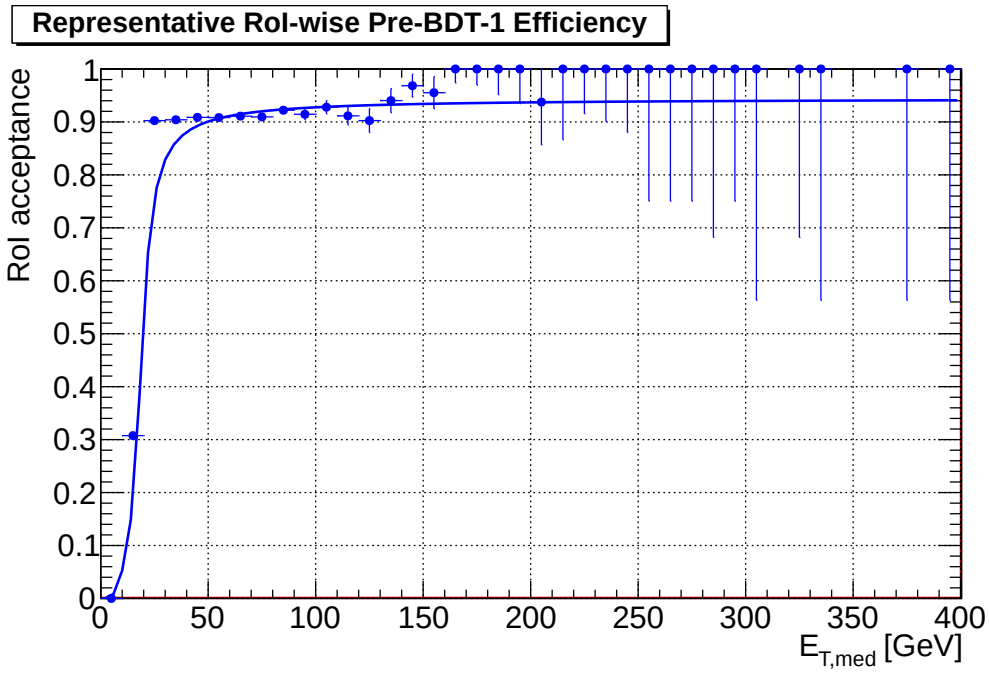


(a)

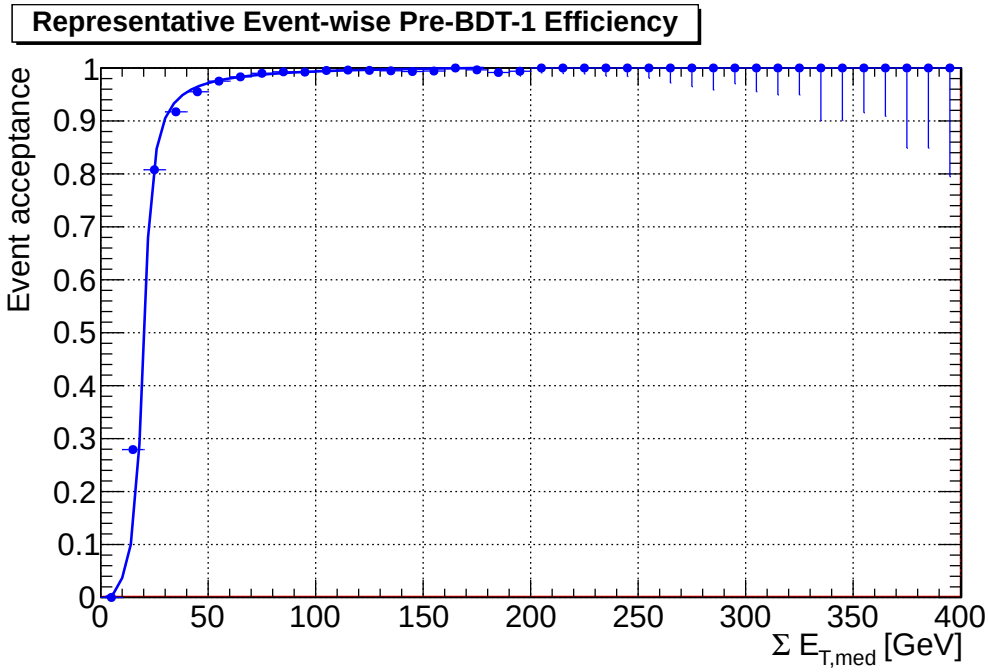


(b)

Figure 6.3: The Pre-BDT-1 acceptance rate for dijet background for two different cases. (a) shows the RoI-wise acceptance rate w.r.t. L2 $E_{T,\mathrm{med}}$, (b) shows the event-wise acceptance rate w.r.t . Σ L2 $E_{T,\mathrm{med}}$.



(a)



(b)

Figure 6.4: The Pre-BDT-1 acceptance rate for the representative sample for two different cases. (a) shows the RoI-wise acceptance rate w.r.t. L2 $E_{T,\mathrm{med}}$, (b) shows the event-wise acceptance rate w.r.t . Σ L2 $E_{T,\mathrm{med}}$.

6.4 BDT Variables

For BDT-1, we use the variables: `CoreFrac`, `CaloRadius`, `EnergyTonCells`, and `HADtoEMenergy`. `CoreFrac` was included by default since it is currently the primary discriminating variable for the calorimeter stage of the L2 tau trigger, and it provides a clean and clear separation. After scanning over all combinations of variables for this stage (that contain `CoreFrac`), it was found that for all combinations that are statistically equivalent (within 3σ) to the combination with the highest signal acceptance (RoI-wise, class of events with no error codes from the $Z \to \tau\tau$ training sample), they contain the `CaloRadius` variable. We eliminated combinations that contain the `HADRadius` variable, since this variable presents a significantly large fraction of error codes. This left us only with combinations that contain the `EnergyTonCells` variable. From these combinations, all but two of them contain the `HADtoEMenergy` variable, but these two contain `EMFrac`, which is similar to `HADtoEMenergy`, but possibly has error code values. For simplicity, we choose the combination that contains these four mentioned variables and no others.

BDT-2 uses all of the calorimeter and tracking variables, as listed in Section 5.5.

Figure 6.5 shows the distributions of the values of the variables used for BDT-1 for the sample of events that are passed to its training procedure, when none of the values are error codes.

Moreover, Figure 6.6 shows the correlation matrices for these same variables.

Clearly, all variables are significantly uncorrelated, except for the correlation between `CoreFrac` and `CaloRadius`. It may still be acceptable to include both of these variables, since there still may be a significant amount of uncorrelation between these variables that provides a significant amount of additional separation power to the classifier, and the evidence of this comes from our previous mention of how `CaloRadius` was required on top of `CoreFrac` in order to have a high level of signal acceptance in comparison to the other combinations of variables.

Figures 6.7, 6.8, and 6.9 show the distributions of the values of the variables used for BDT-2 for the sample of events that are passed to its training procedure, when none of the

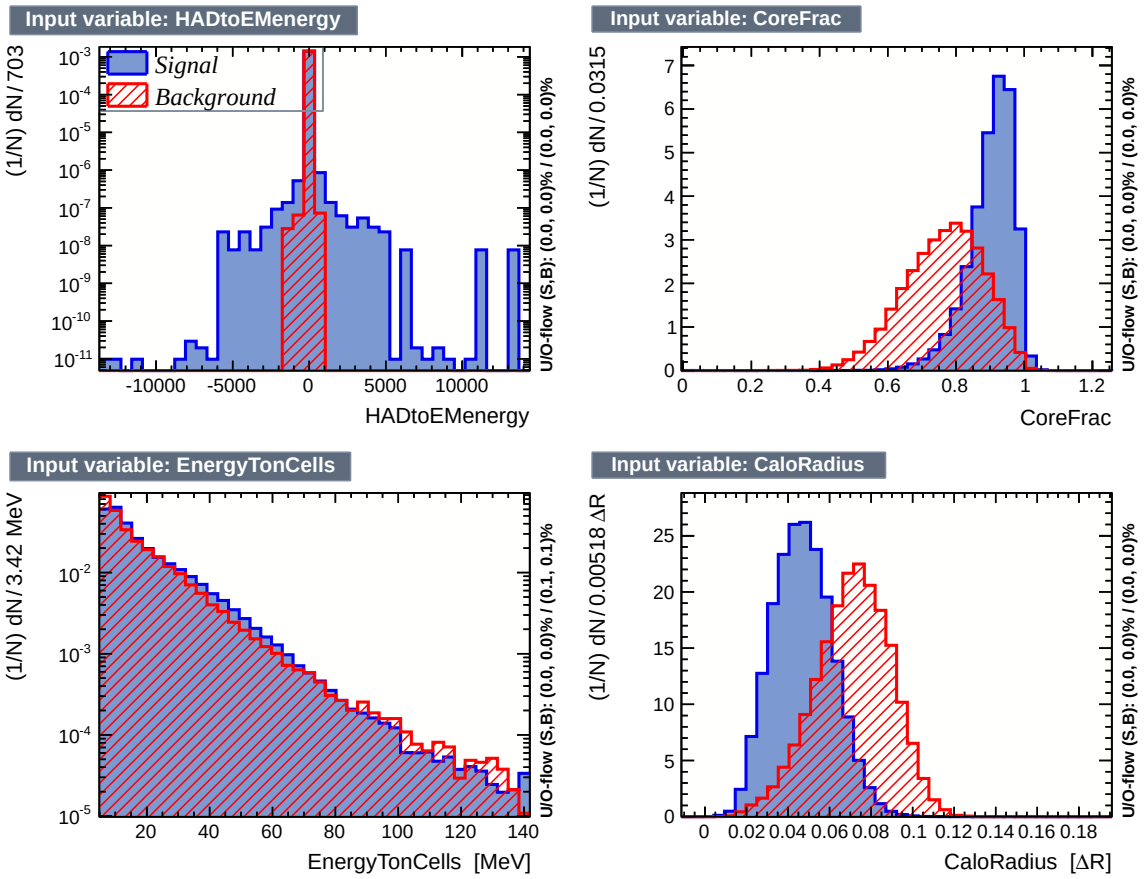


Figure 6.5: TMVA generated histograms of the values of the variables used for BDT-1 for the sample of events that are passed to its training procedure, when none of the values are error codes.

values are error codes.

Figure 6.10 shows the corresponding correlation matrix.

Again, we can see that the variables have acceptable levels of correlation between each other, which justifies using all of them.

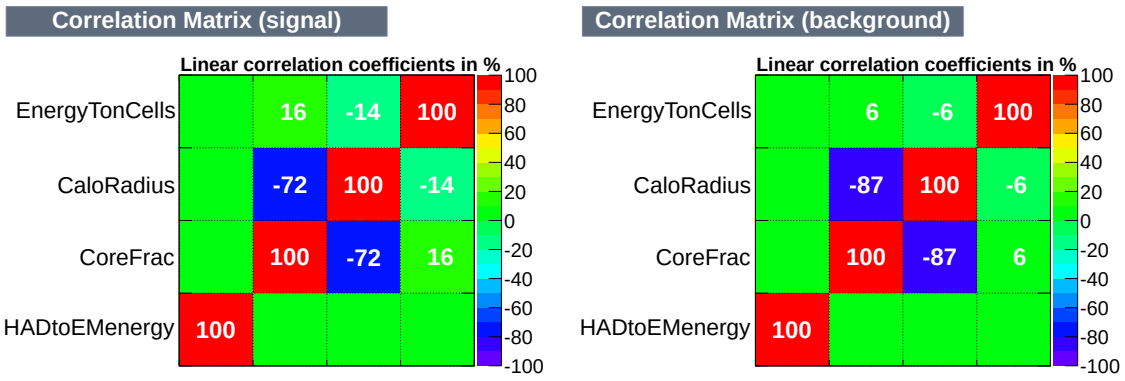


Figure 6.6: The TMVA generated correlation matrices for the variables used for BDT-1 for the sample of events that are passed to its training procedure. Each entry is the sample correlation coefficient (scaled by 100) for the variables associated with the corresponding row and column. Implicitly, the columns correspond to the same ordered set of variables as the rows. An entry without a shown value means that the value is 0. (a) shows the signal correlation matrix and (b) shows the background correlation matrix.

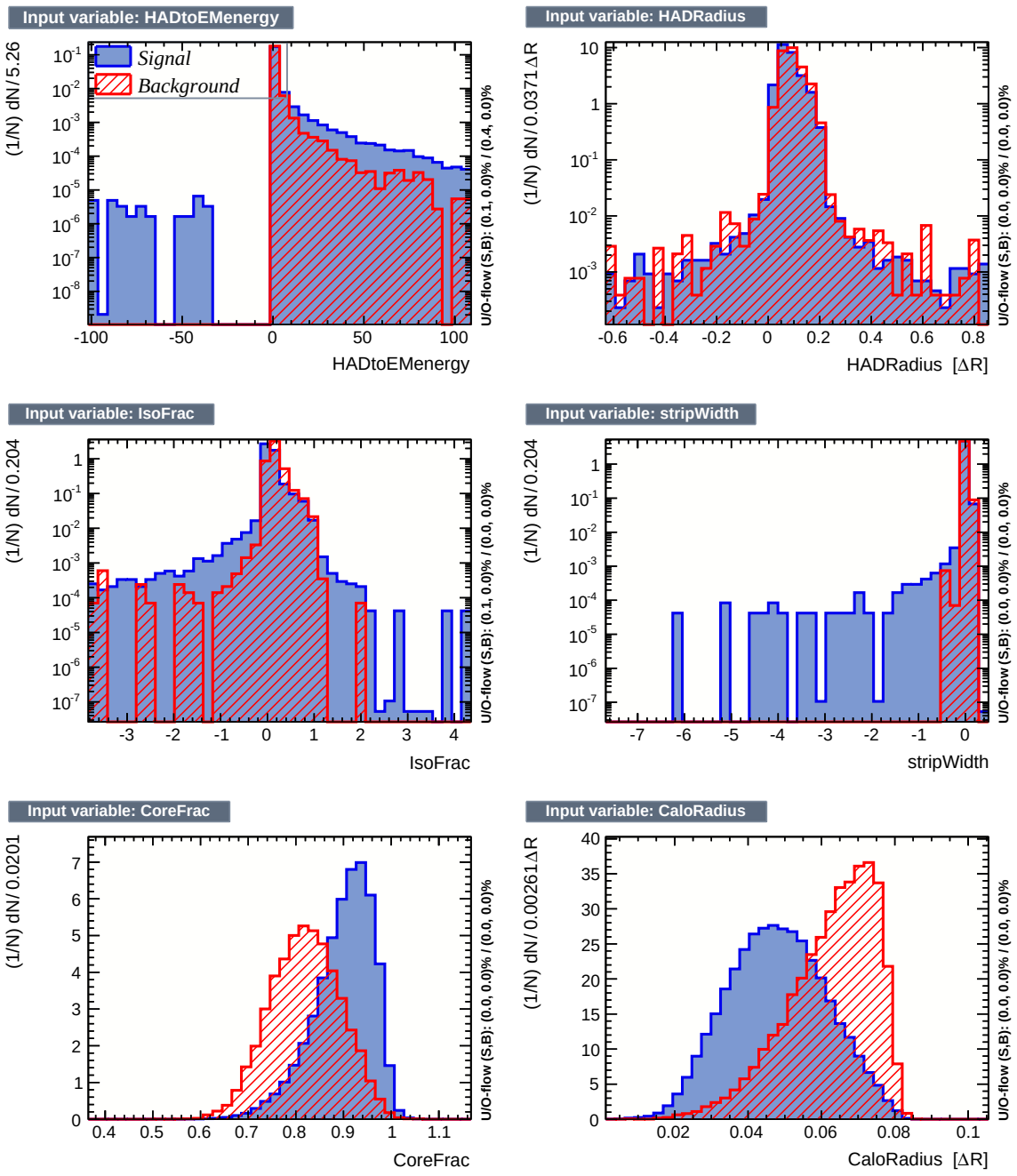


Figure 6.7: TMVA generated histograms of the values of the first six variables used for BDT-2 for the sample of events that are passed to its training procedure, when none of the values are error codes.

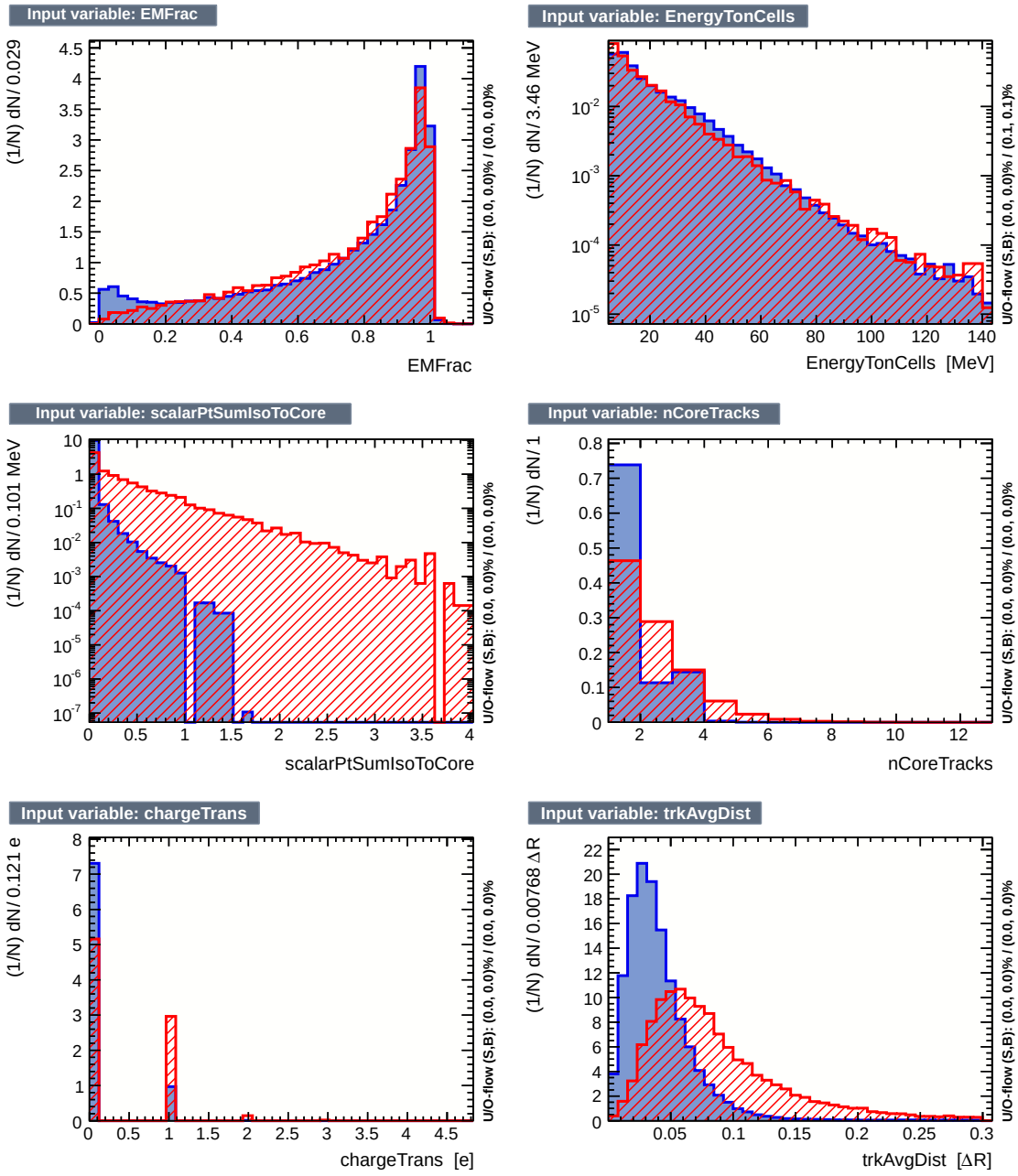


Figure 6.8: TMVA generated histograms of the values of the second six variables used for BDT-2 for the sample of events that are passed to its training procedure, when none of the values are error codes.

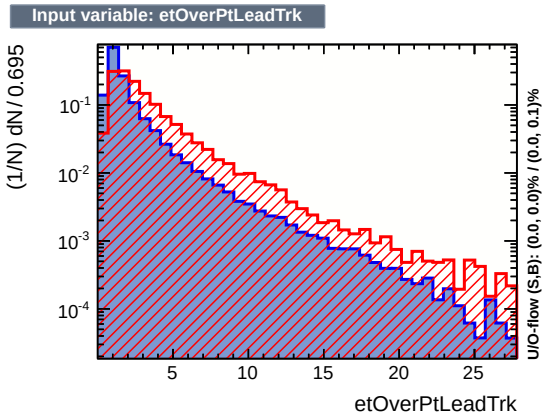
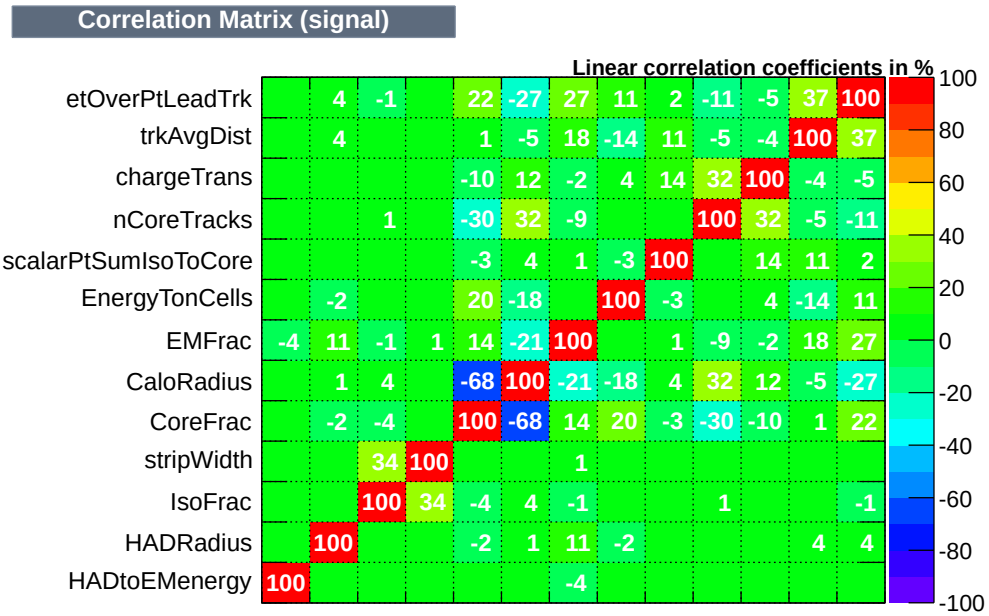
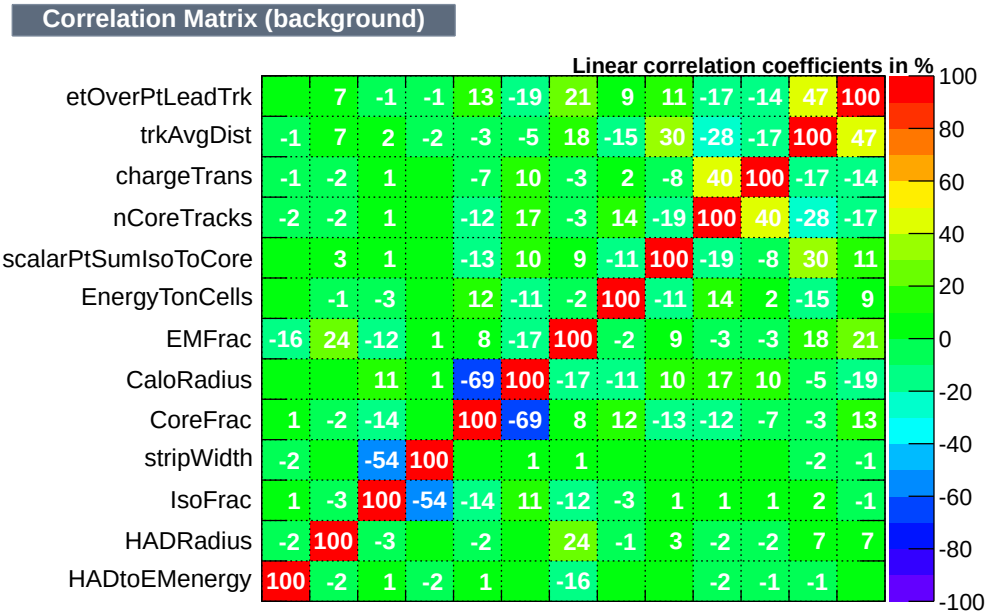


Figure 6.9: A TMVA generated histogram of the values of the last variable used for BDT-2 for the sample of events that are passed to its training procedure, when none of the values are error codes.



(a)



(b)

Figure 6.10: The TMVA generated correlation matrices for the variables used for BDT-2 for the sample of events that are passed to its training procedure. Each entry is the sample correlation coefficient (scaled by 100) for the variables associated with the corresponding row and column. Implicitly, the columns correspond to the same ordered set of variables as the rows. An entry without a shown value means that the value is 0. (a) shows the signal correlation matrix and (b) shows the background correlation matrix.

6.5 Error Codes

For BDT-1, the variables with defined error codes are `CoreFrac` and `CaloRadius`. The representative sample contains 53 RoIs with an error code for `CoreFrac` and 40 RoIs with an error code for `CaloRadius`. However, after Pre-BDT-1, this sample contains no RoIs with an error code for both `CoreFrac` and `CaloRadius`. Therefore, BDT-1 consists of only one error code case: the class with no error codes.

For BDT-2, the situation is less trivial. Table 6.2 shows the number of events per error code class for the representative sample after Pre-BDT-2 has been applied, where we only show the error code classes that are not automatically rejected by the algorithm.

Error code class	**Number of events**
None	173950
`HADRadius`	23387
`trkAvgDist`	7290
`HADRadius` & `trkAvgDist`	3691
`etOverPtLeadTrk`	3396
`HADRadius` & `etOverPtLeadTrk`	2194
`trkAvgDist` & `etOverPtLeadTrk`	482
`stripWidth`	110
Total	211280

Table 6.2: Number of events per error code class for the representative sample after Pre-BDT-2 has been applied. The classes are labeled by the variables with error code values. We only show the error code classes that are not automatically rejected by the algorithm.

For all cases, the reason a class was rejected was that TMVA was not able to train a BDT for the class due to a low number of statistics.

Note: For the error code classes with error codes for `etOverPtLeadTrk`, we removed all tracking based variables from the BDT training program. This is because they are all 0 or error code valued in this case, and hence, cannot be used for training.

6.6 Overtraining Check

In general, a machine learning algorithm is said to be overtrained if too many model parameters are adjusted to too few data points. This can be detected by comparing the performance of the algorithm on independent training and testing samples. If the performance is significantly greater for the training sample, then overtraining has occured. [54]

Our training procedure was first applied with the maximum tree depth and number of trees fixed to the same value for each BDT classifier (for each stage and each error code class). In order to ensure reasonable overtraining conditions, without significantly sacrificing discrimination power, we manually reoptimized these parameters for each trained BDT classifier. Table 6.3 shows the optimized values of these parameters for each case.

Error code class	**Max. tree depth**	**No. of trees**
BDT-1 None	6	50
BDT-2 None	6	50
BDT-2 `HADRadius`	6	20
BDT-2 `trkAvgDist`	6	15
BDT-2 `HADRadius` & `trkAvgDist`	4	15
BDT-2 `etOverPtLeadTrk`	3	35
BDT-2 `HADRadius` & `etOverPtLeadTrk`	3	20
BDT-2 `trkAvgDist` & `etOverPtLeadTrk`	3	15
BDT-2 `stripWidth`	2	25

Table 6.3: The maximum tree depth and number of trees for each stage and each error code class. These values were manually optimized in order to ensure reasonable overtraining conditions, without significantly sacrificing discrimination power. The classes are labeled by the variables with error code values.

Figure 6.11 shows the overtraining check for BDT-1 and Figures 6.12 to 6.19 show the overtraining check for each of the error code classes of BDT-2.

The Kolmogorov-Smirnov (KS) probability is calculated using ROOT's `TH1->KolmogorovTest()` method, which gives a value for the amount of compatibility between two histograms; in this case, the training and testing samples. In theory, if we reject all pairs of histograms with KS probability less then α, then the probability of rejecting a pair of truly compatible histogram is bounded above by α. When optimizing the above training parameters, we

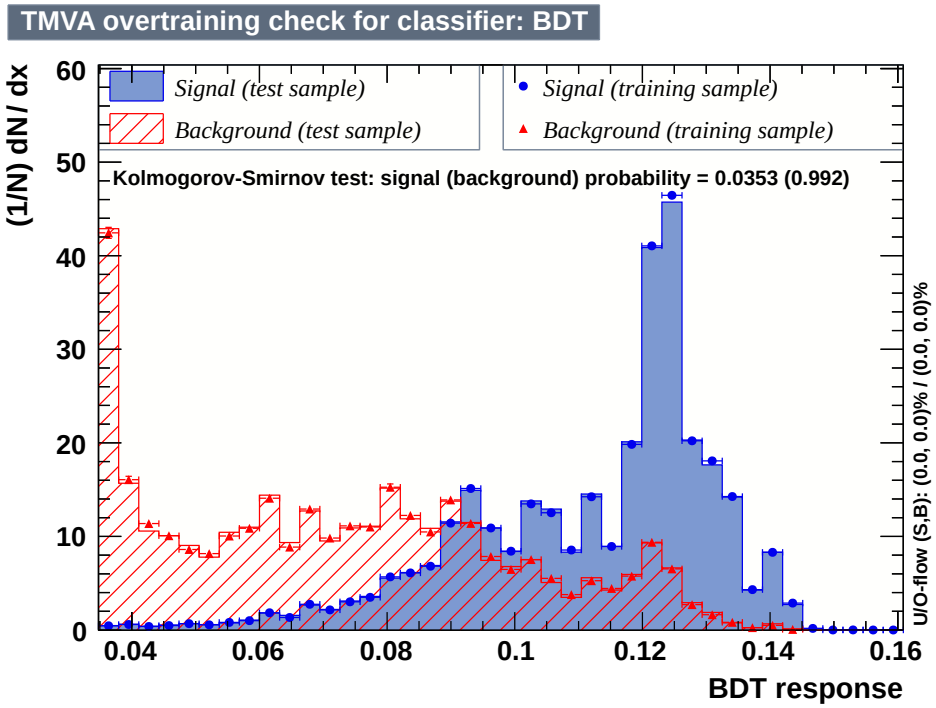


Figure 6.11: TMVA generated overtraining check for BDT-1, with no error codes for all the variables used for selection at this stage.

aimed for a KS probability of at least 0.01 for both signal and background. This results in reasonably looking distributions, and is the greatest value, up to order of magnitude, that we were able to attain for some of the error code cases.

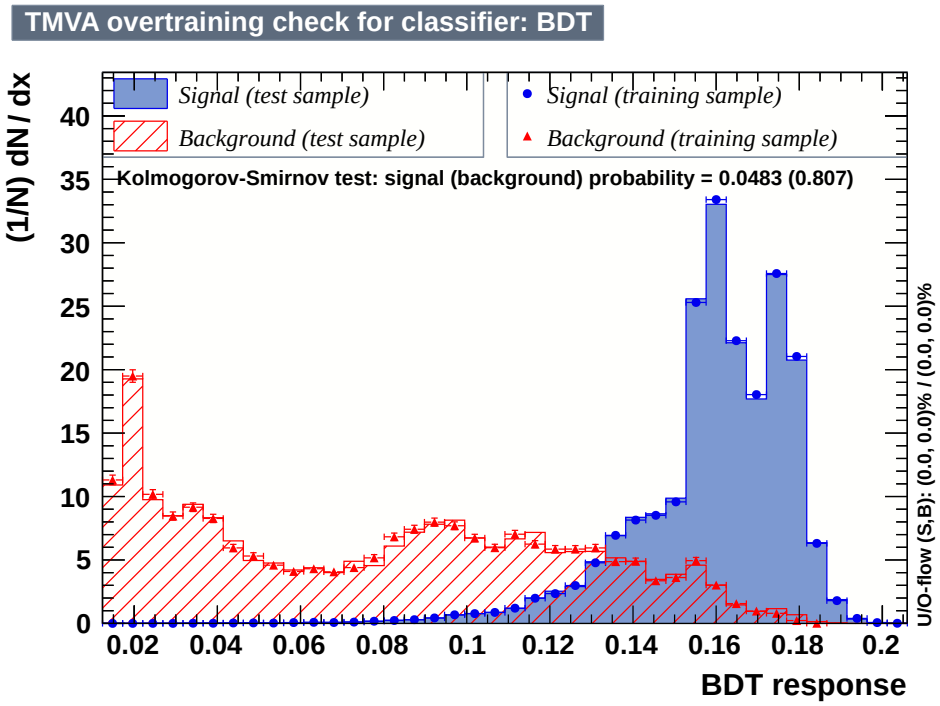


Figure 6.12: TMVA generated overtraining check for BDT-2, with no error codes for all variables.

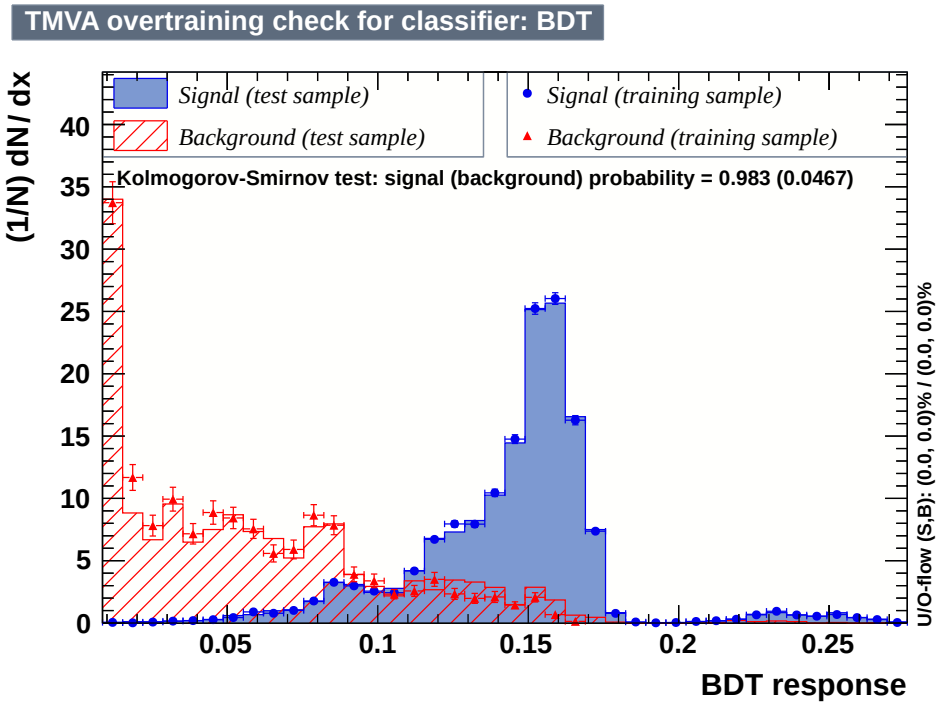


Figure 6.13: TMVA generated overtraining check for BDT-2, with error codes for `HADRadius`.

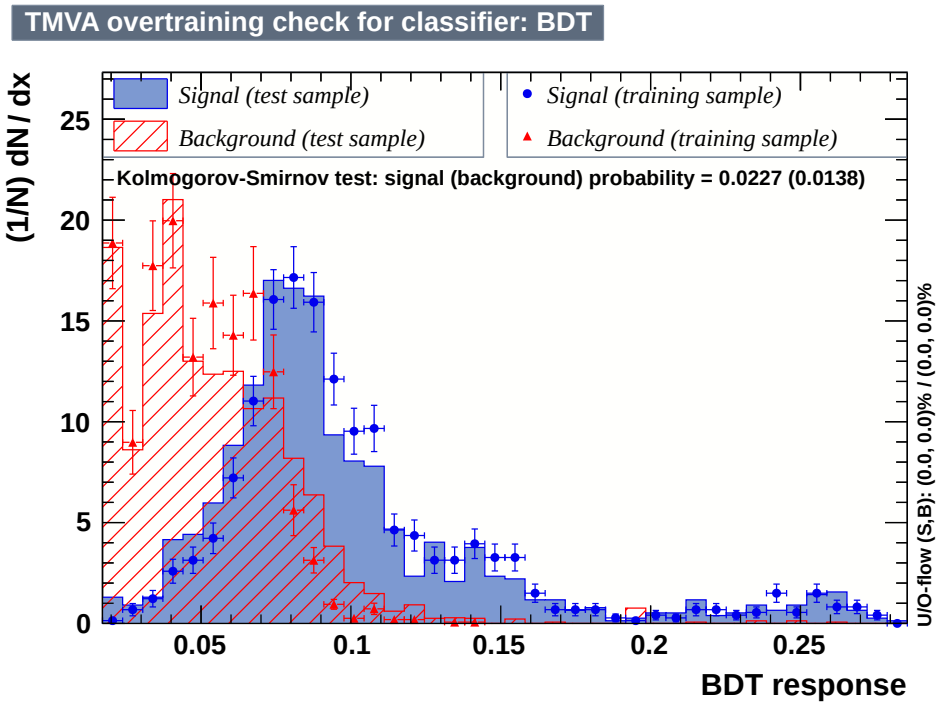


Figure 6.14: TMVA generated overtraining check for BDT-2, with error codes for `trkAvgDist`.

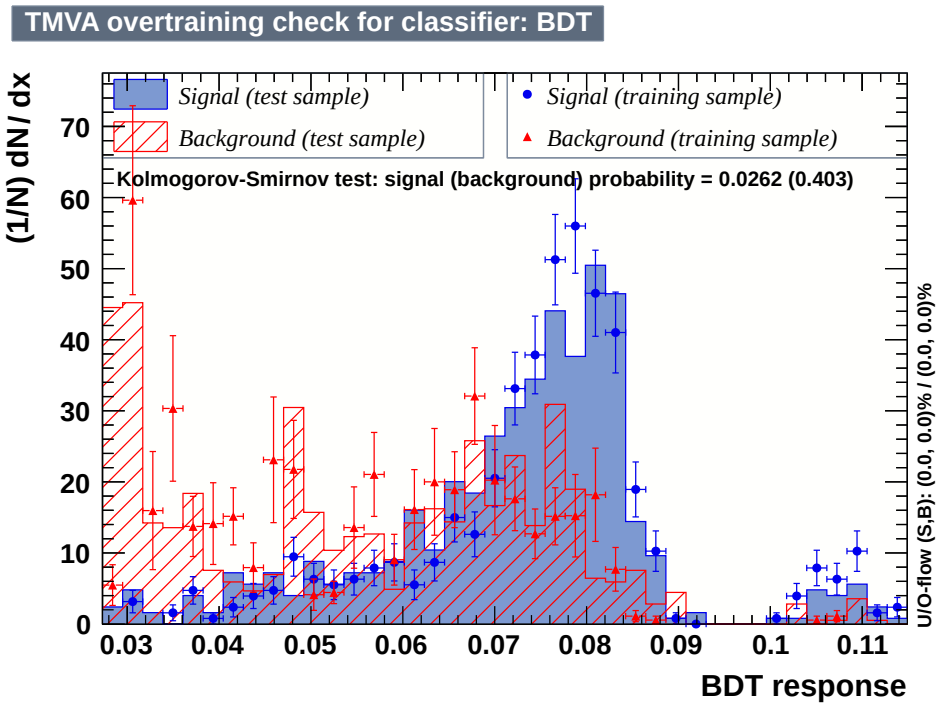


Figure 6.15: TMVA generated overtraining check for BDT-2, with error codes for `HADRadius` and `trkAvgDist`.

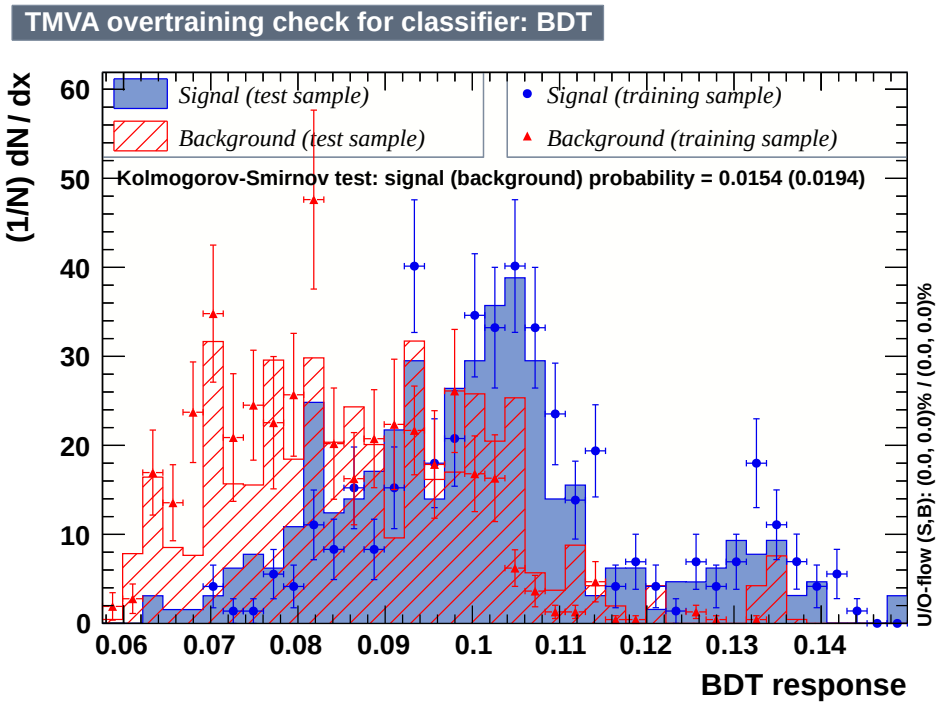


Figure 6.16: TMVA generated overtraining check for BDT-2, with error codes for `etOverPtLeadTrk`.

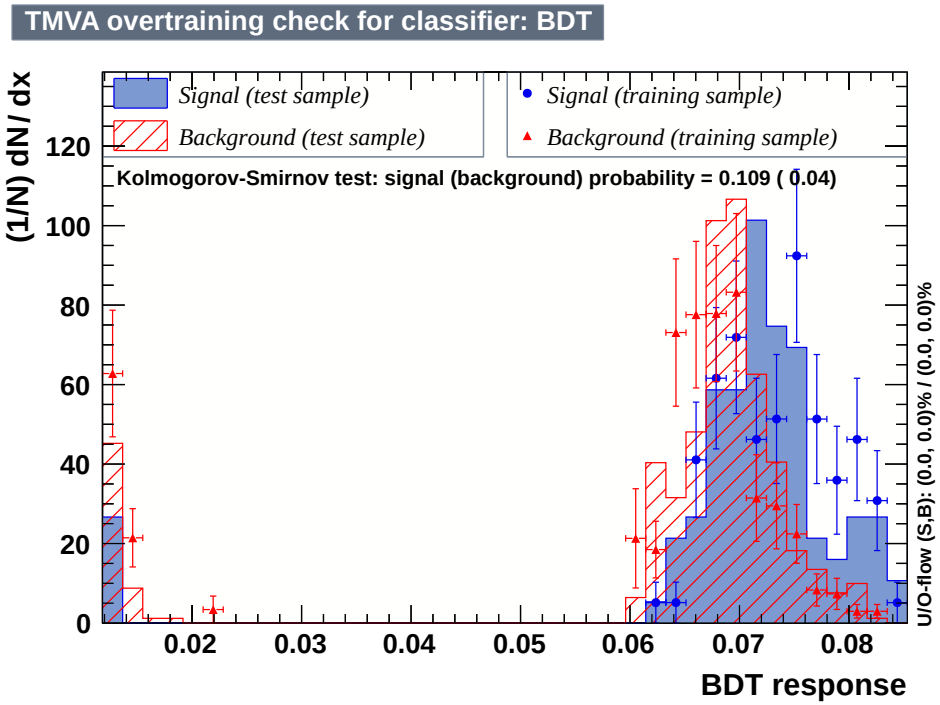


Figure 6.17: TMVA generated overtraining check for BDT-2, with error codes for `HADRadius` and `etOverPtLeadTrk`.

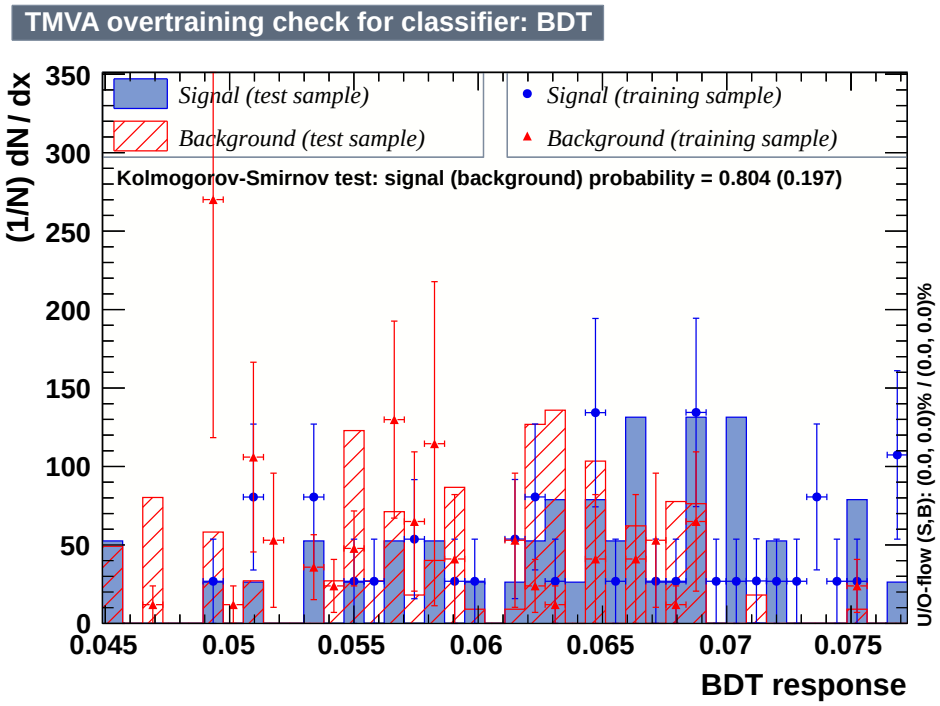


Figure 6.18: TMVA generated overtraining check for BDT-2, with error codes for `trkAvgDist` and `etOverPtLeadTrk`.

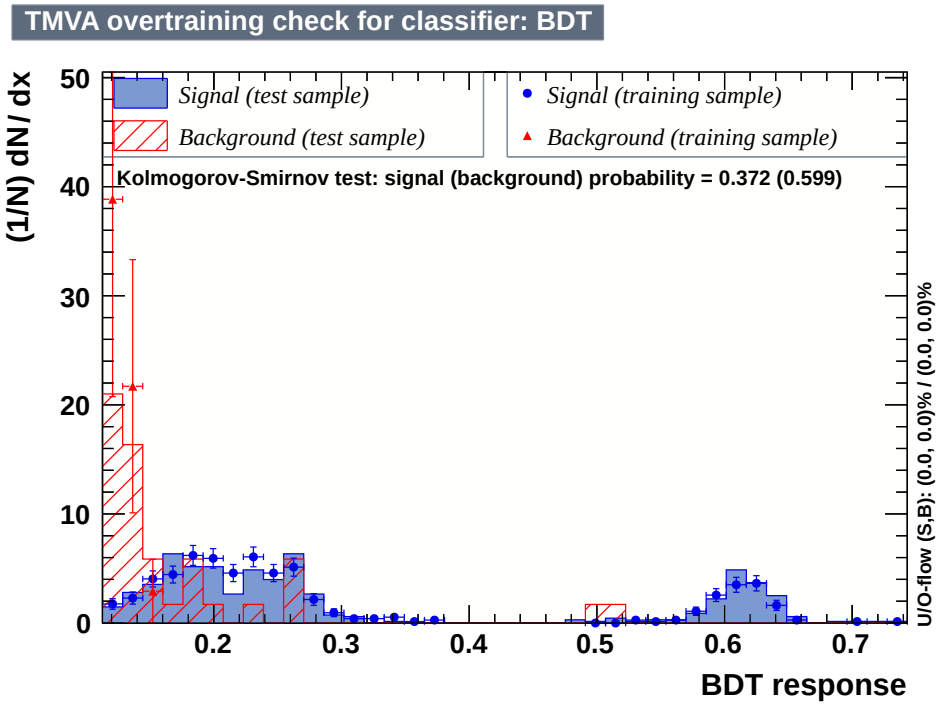


Figure 6.19: TMVA generated overtraining check for BDT-2, with error codes for `stripWidth`.

6.7 Decision Trees

The output of the training procedure is a set of boosted decision trees. Figure 6.20 shows a diagram of one of the trees for BDT-1, and Figure 6.21 shows a diagram of one of the trees for BDT-2.

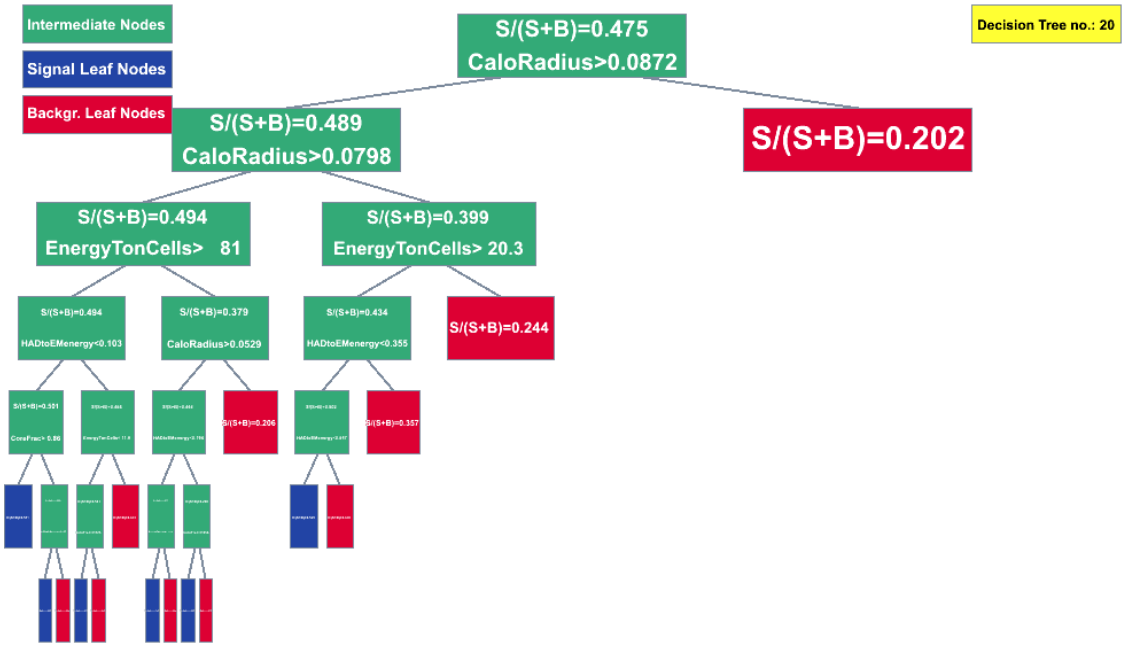


Figure 6.20: A TMVA generated diagram of decision tree no. 20 for BDT-1, with no error codes for all the variables used for selection at this stage.

For BDT-1, the importance index of each variable is listed in Table 6.4, and for BDT-2, the importance indices are listed in Table 6.5.

Variable	**Importance**
`CoreFrac`	4.271×10^{-1}
`CaloRadius`	4.140×10^{-1}
`HADtoEMenergy`	1.065×10^{-1}
`EnergyTonCells`	5.235×10^{-2}

Table 6.4: The TMVA generated importance indices of the variables for BDT-1, with no error codes for all the variables used for selection at this stage.

We can see from Table 6.5 that `HADRadius` and `HADtoEMenergy` are good potential candidates for removal from BDT-2.

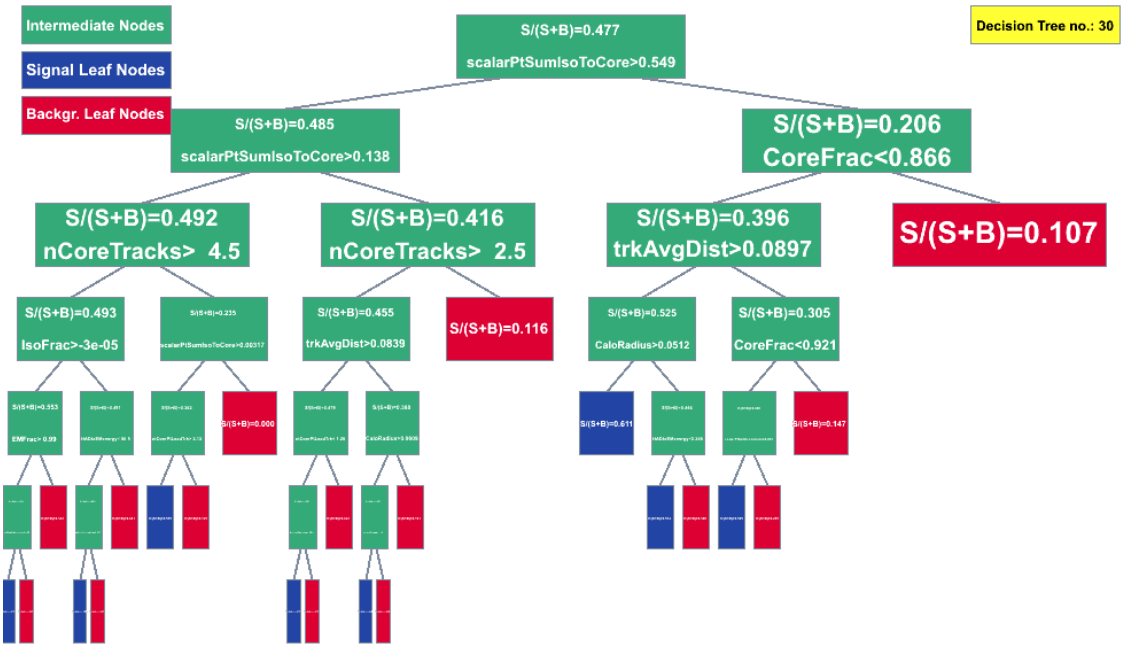


Figure 6.21: A TMVA generated diagram of decision tree no. 30 for BDT-2, with no error codes for all variables.

Variable	Importance
`scalarPtSumIsoToCore`	2.485×10^{-1}
`EMFrac`	1.626×10^{-1}
`CaloRadius`	1.341×10^{-1}
`CoreFrac`	1.115×10^{-1}
`etOverPtLeadTrk`	8.980×10^{-2}
`trkAvgDist`	7.095×10^{-2}
`nCoreTracks`	7.008×10^{-2}
`IsoFrac`	2.846×10^{-2}
`chargeTrans`	2.191×10^{-2}
`EnergyTonCells`	2.084×10^{-2}
`stripWidth`	1.852×10^{-2}
`HADRadius`	1.320×10^{-2}
`HADtoEMenergy`	9.475×10^{-3}

Table 6.5: The TMVA generated importance indices of the variables for BDT-2, with no error codes for all variables.

6.8 Calorimeter Stage

It is important to measure the acceptance rates exclusively for the calorimeter stage of the trigger, to ensure that they are below the predetermined limits. In order to satisfy the requirements stated in Section 5.6, the cutoff scores for BDT-1 have been set so that the event-wise calorimeter-stage acceptance rate for the representative sample is below 0.29, with a 3σ certainty threshold. Figure 6.22 shows the calorimeter-stage acceptance rate as a function of pile-up for the representative sample, with no error codes for all the variables used for selection at this stage, and Figure 6.23 shows these rates for the data taking period E. The overall event-wise rates are listed in Table 6.6.

Sample	**L2_TAU20_MEDIUM1**	**BDT Trigger**
Representative	$(2.9369 \pm 0.0057) \times 10^{-1}$	$(2.88136 \pm 0.00603 \pm 0.00064) \times 10^{-1}$
Period E	$(2.9164 \pm 0.0080) \times 10^{-1}$	$(2.8685 \pm 0.0080 \pm 0.0012) \times 10^{-1}$

Table 6.6: The calorimeter stage event-wise acceptance rates, with no error codes for all the variables used for selection at this stage.

We only show the error code class with no error codes, since this is the only error code class that gets accepted by BDT-1, and thus the acceptance rates for this class give us upper bounds on the overall acceptance rates. As we can see, for each case presented and for each value of μ (with the exception of a few outlier values of μ where statistics are low), the rate of the BDT trigger generally falls below the rate of the reference trigger. Moreover, the overall event-wise rates of the BDT trigger are each confidently less than the required limit and less than the corresponding rates for the reference trigger.

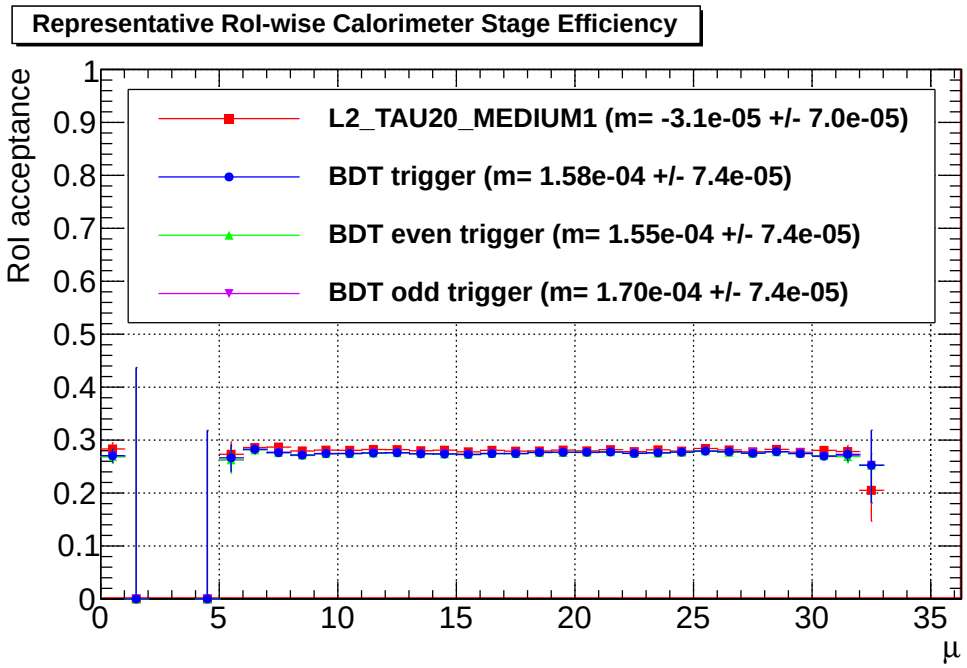
As for the amount of pile-up dependence, we apply standard error propagation to obtain the difference between the absolute value of the slope of the linear fit for the BDT trigger with that for the reference trigger, for each of the four cases presented. The results are summarized in Table 6.7.

As we can see, the BDT trigger slopes are each greater than the corresponding slopes for the reference trigger, with 1σ certainty. However, they are not greater with 2σ certainty, and for the representative sample, they are each less than 3.32 % efficiency per 50 interactions

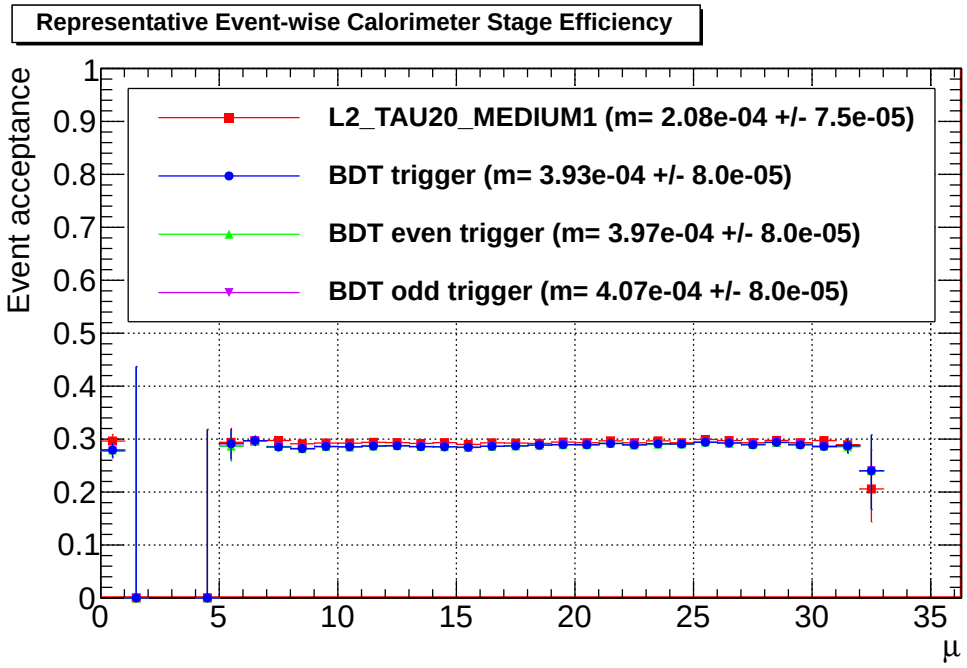
Sample	**RoI-wise**	**Event-wise**
Representative	$(1.27 \pm 1.02 \pm 0.15) \times 10^{-4}$	$(1.85 \pm 1.10 \pm 0.10) \times 10^{-4}$
Period E	$(2.8 \pm 1.7 \pm 0.2) \times 10^{-4}$	$(2.7 \pm 1.8 \pm 0.0) \times 10^{-4}$

Table 6.7: The difference between the absolute value of the slope of the linear fit for the BDT trigger with that for the reference trigger, for each of the four cases presented in this section. Each case has no error codes for all the variables used for selection at this stage.

per bunch crossing (an approximate upper bound on pile-up), with 3σ certainty. Also, visually, all of the slopes look indistinguishably flat, and it is possible that the outlier values of μ may be misrepresenting the results. It is therefore reasonable to conclude that pile-up dependence is not a problem for the BDT trigger, and its pile-up dependence is not significantly worse than that of the reference trigger.

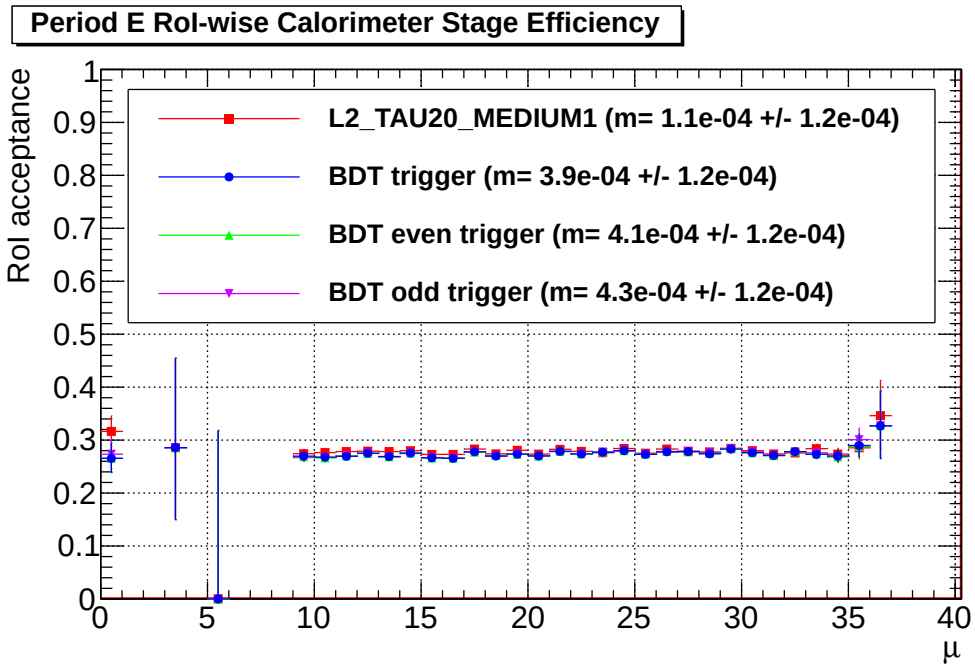


(a)

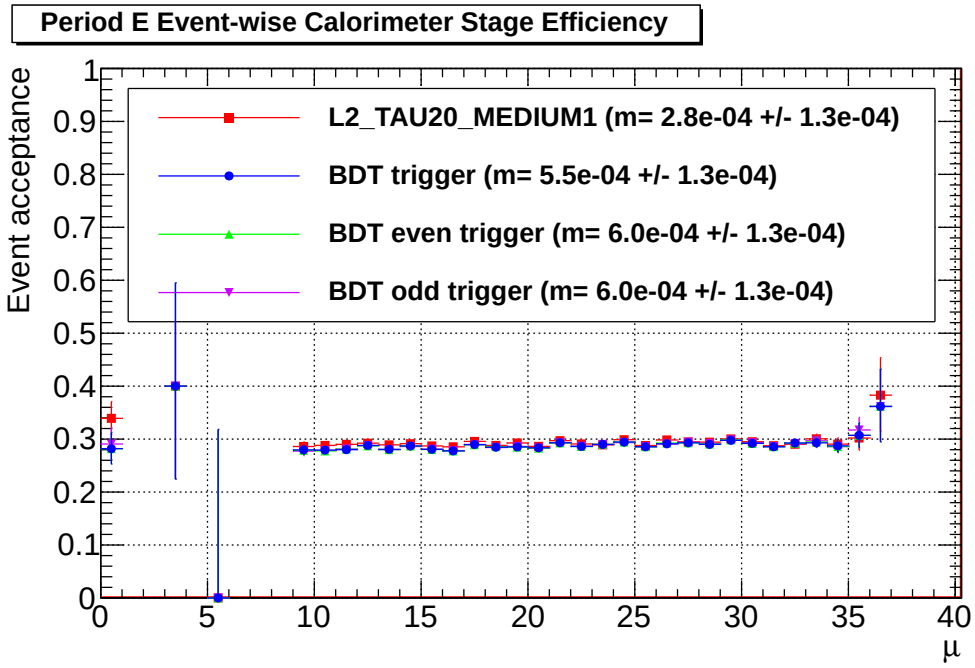


(b)

Figure 6.22: The calorimeter stage acceptance rate as a function of pile-up for the representative sample, with no error codes for all the variables used for selection at this stage. (a) shows the RoI-wise rate and (b) shows the event-wise rate. m is the slope of a linear fit for the corresponding graph.



(a)



(b)

Figure 6.23: The calorimeter stage acceptance rate as a function of pile-up for the data taking period E, with no error codes for all the variables used for selection at this stage. (a) shows the RoI-wise rate and (b) shows the event-wise rate. m is the slope of a linear fit for the corresponding graph.

6.9 Signal Acceptance versus Overall Rejection

To show the performance of the overall algorithm (calorimeter stage followed by tracking stage), we apply it to the $Z \rightarrow \tau\tau$ signal training sample and the representative sample for seven different cutoff scores for BDT-2, including the cutoff score used to optimize the overall event-wise acceptance rate to make it is as close as possible to the stated limit (0.085). Each cutoff score is determined by a target event-wise acceptance rate, with 0.085 as the central rate, surrounded by the three evenly spaced out rates between 0 and 0.085 (0.02125, 0.0425, 0.06375) and the three evenly spaced out rates between 0 and 2×0.085 (0.10625, 0.1275, 0.14875). Figures 6.24 to 6.30 show the $Z \rightarrow \tau\tau$ signal acceptance rate plotted against the overall (representative) rejection rate (1 minus acceptance rate) for each of the principal error code classes.

The acceptance rates for the full sample of events (no restriction to an error code class) are shown in Figure 6.32.

Note: For the class with error codes for `trkAvgDist` and `etOverPtLeadTrk`, both the even and odd BDT triggers contain the trivial (reject everything) BDT classifier for this class. This is due to the low number of statistics, which resulted in maximally pruned trees for these classifiers. Futhermore, for the class with error codes for `stripWidth`, both the even and odd BDT triggers have no corresponding BDT classifier for this class, due again to the insufficient number of statistics needed to train the classifiers. Hence, all events from this class are automatically rejected by these triggers as well.

For each case, except the one shown in Figure 6.31, it is clear that the BDT trigger performs better than the reference trigger in the sense that it has a cutoff score that results in both a significantly higher overall rejection rate and signal acceptance rate. For the class with error codes for `HADRadius` (Figure 6.25), a point showing better performance is not visible, but generally, the curve for signal acceptance versus overall rejection is concave to the origin, due to the typically diminishing returns of signal acceptance as background rejection is decreased, which can be visualized by looking at the overtraining plots. Therefore, any point between the 6th point and the 7th point in Figure 6.25 should lie outside a straight

line drawn between the two points. For the classes with error codes for `etOverPtLeadTrk` and possibly other variables (Figures 6.28 to 6.30), the acceptance rate for the reference trigger is exactly 0 (see the definition of the reference trigger). Hence, the question of whether the BDT trigger performs better for these classes is not well defined. But in terms of signal acceptance, we can say that some signal is better than no signal (as long as the background acceptance is sufficiently limited) and conclude that the BDT trigger performs better for these classes. Finally, for the class with error codes for `stripWidth`, the reference trigger seems to perform slightly better than the BDT trigger. However, if we ignore the systematic uncertainty and take 0.085 as our event-wise rate limitation, then we can attain a slightly greater signal acceptance rate with the BDT trigger while still remaining below the rate limitation with 3σ certainty. Due to the low number of statistics in this class, a more conservative solution for improving the BDT trigger's performance for this class can be to set the trigger to use the reference trigger's criteria when considering events from this class.

Henceforth, in this and the following sections, the results will be implicitly assumed to be for the overall algorithm, with a target rate of 0.085 and with no restriction to an error code class. Tables 6.8 and 6.9 summarize the results for the RoI-wise and event-wise cases, respectively.

Sample	**L2_TAU20_MEDIUM1**	**BDT Trigger**
Signal	$(6.8349 \pm 0.0049) \times 10^{-1}$	$(7.6591 \pm 0.0047 \pm 0.0321) \times 10^{-1}$
Representative	$(7.791 \pm 0.032) \times 10^{-2}$	$(7.47239 \pm 0.03996 \pm 0.00047) \times 10^{-2}$

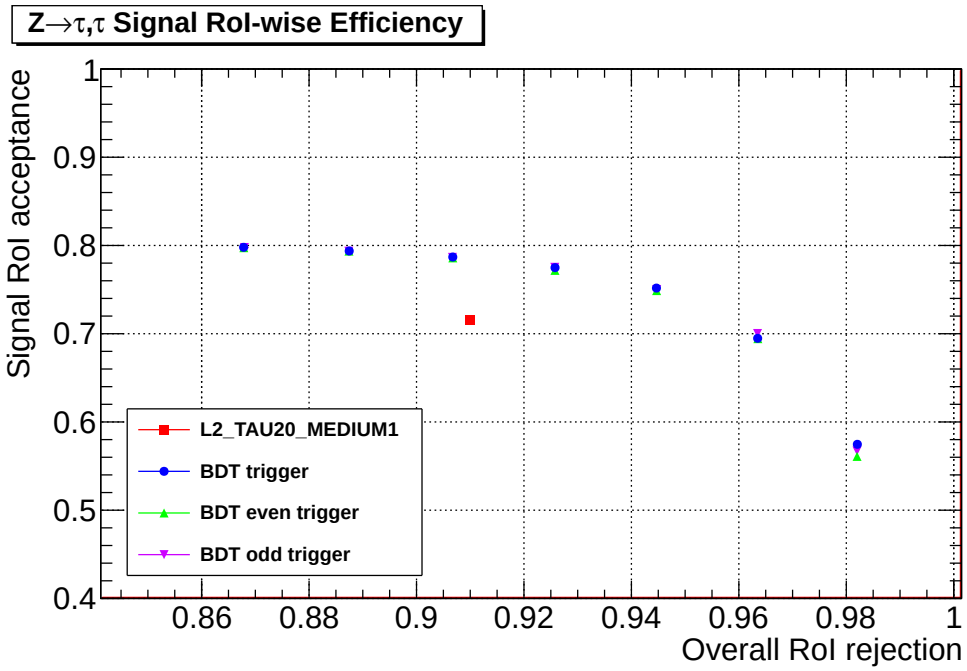
Table 6.8: The RoI-wise acceptance rates for both signal and the representative sample, with the $Z \to \tau\tau$ training sample as the signal sample.

Sample	**L2_TAU20_MEDIUM1**	**BDT Trigger**
Signal	$(7.0698 \pm 0.0050) \times 10^{-1}$	$(8.1326 \pm 0.0047 \pm 0.0276) \times 10^{-1}$
Representative	$(8.673 \pm 0.035) \times 10^{-2}$	$(8.3661 \pm 0.0442 \pm 0.0017) \times 10^{-2}$

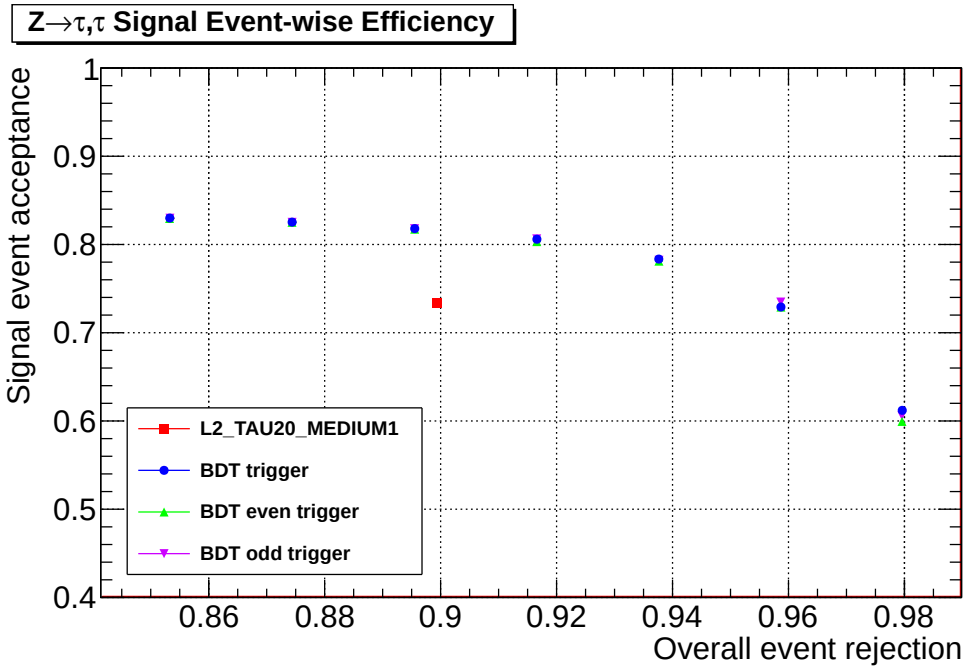
Table 6.9: The event-wise acceptance rates for both signal and the representative sample, with the $Z \to \tau\tau$ training sample as the signal sample.

By applying standard error propagation to the results listed in the tables above, we can

obtain the rate differences between the BDT trigger and the reference trigger. The BDT trigger adds $(0.319 \pm 0.051 \pm 0.000) \times 10^{-2}$ (nearly 0) to the RoI-wise overall rejection rate of the reference trigger and adds $(0.8242 \pm 0.0068 \pm 0.0321) \times 10^{-1}$ to the RoI-wise $Z \to \tau\tau$ signal acceptance rate of the reference trigger. The BDT trigger also adds $(0.307 \pm 0.056 \pm 0.002) \times 10^{-2}$ (nearly 0) to the event-wise overall rejection rate of the reference trigger and adds $(1.0628 \pm 0.0069 \pm 0.0276) \times 10^{-1}$ to the event-wise $Z \to \tau\tau$ signal acceptance rate of the reference trigger. For both the RoI-wise and event-wise cases, the addition to the overall rejection rate is greater than 0 with more than 3σ certainty. For the RoI-wise case, the addition to the signal acceptance rate is more than 7.075% to 68.349%, with 3σ certainty. For the event-wise case, the addition to the signal acceptance rate is more than 9.593% to 70.698%, with 3σ certainty.

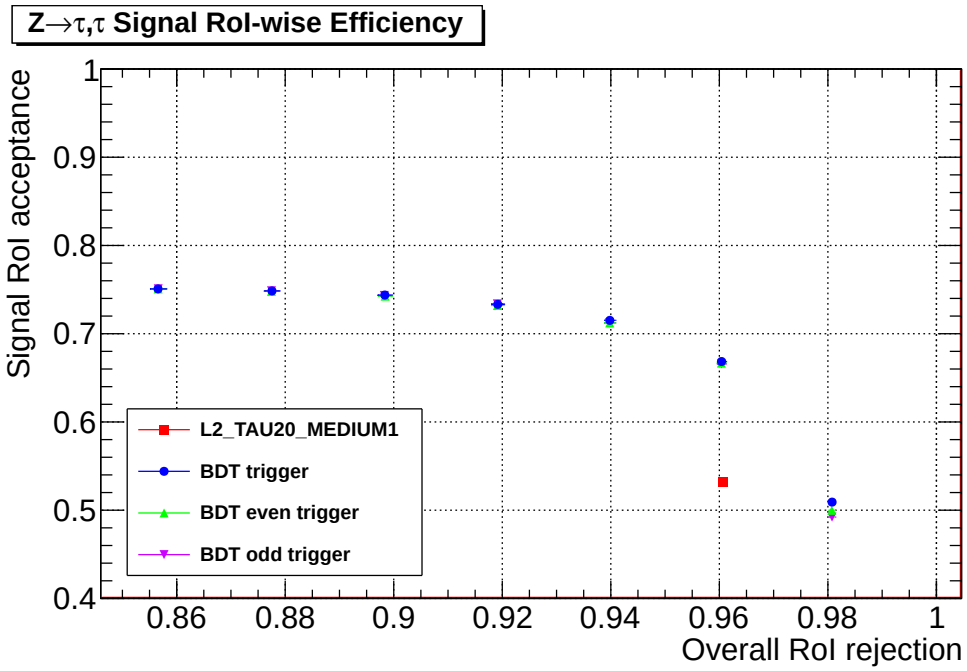


(a)

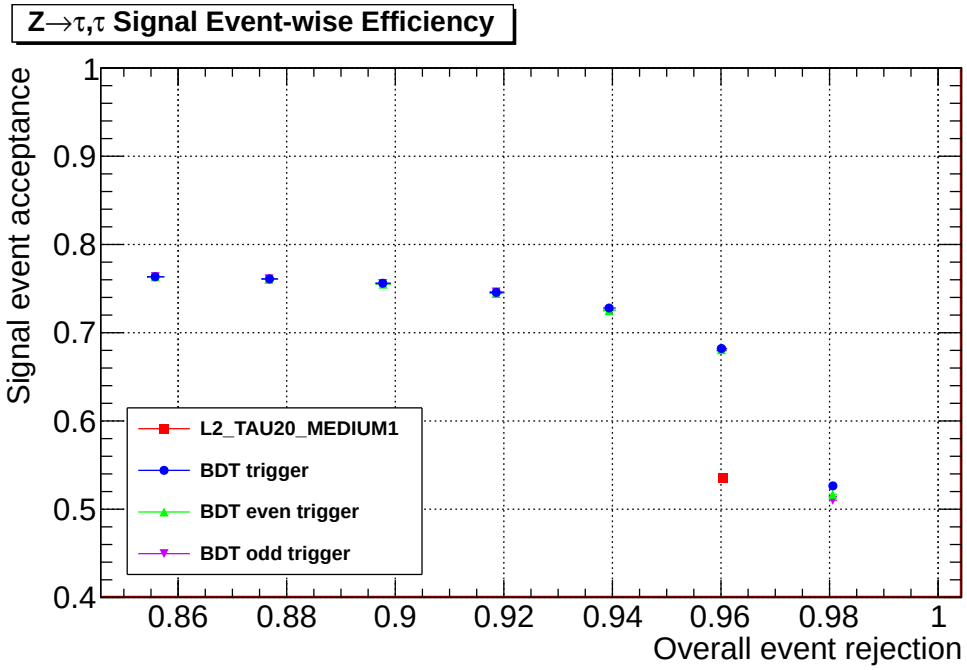


(b)

Figure 6.24: The acceptance rate as a function of the overall acceptance rate for $Z \to \tau\tau$ signal, with no error codes for all the variables. (a) shows the RoI-wise rate and (b) shows the event-wise rate.

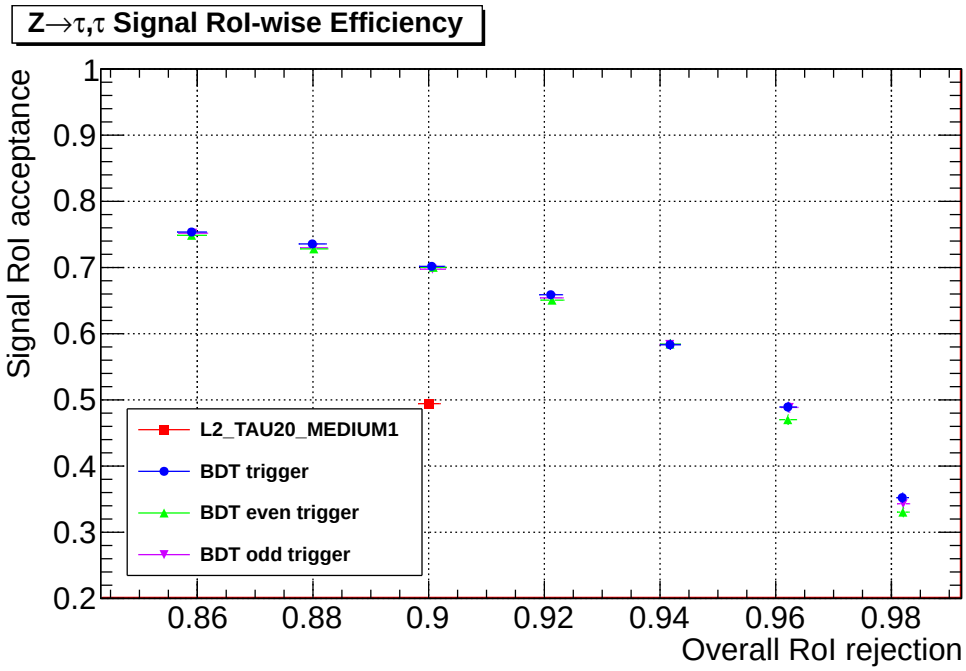


(a)

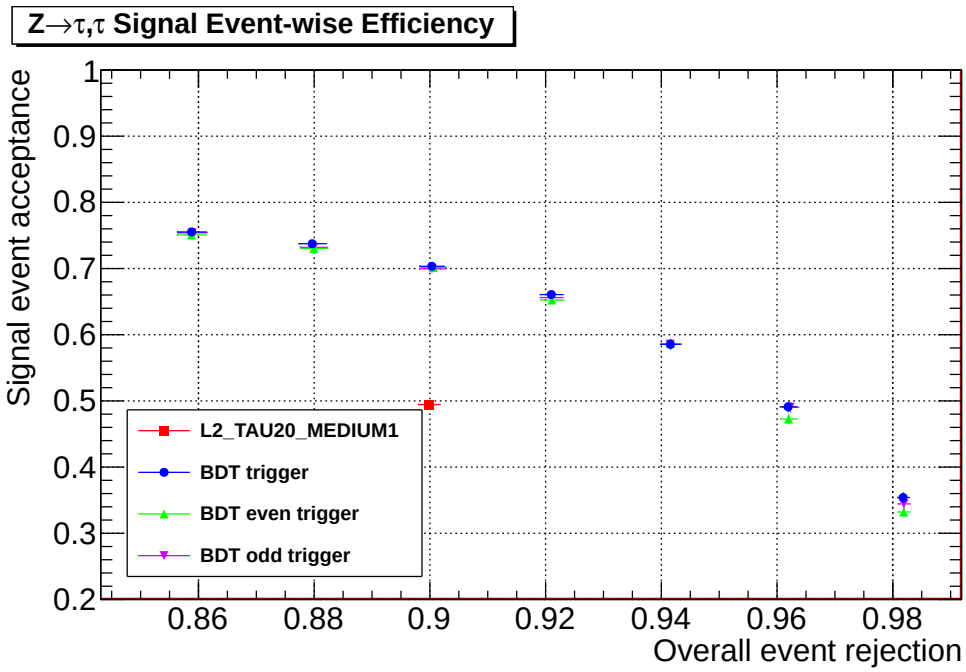


(b)

Figure 6.25: The acceptance rate as a function of the overall acceptance rate for $Z \to \tau\tau$ signal, with error codes for `HADRadius`. (a) shows the RoI-wise rate and (b) shows the event-wise rate.

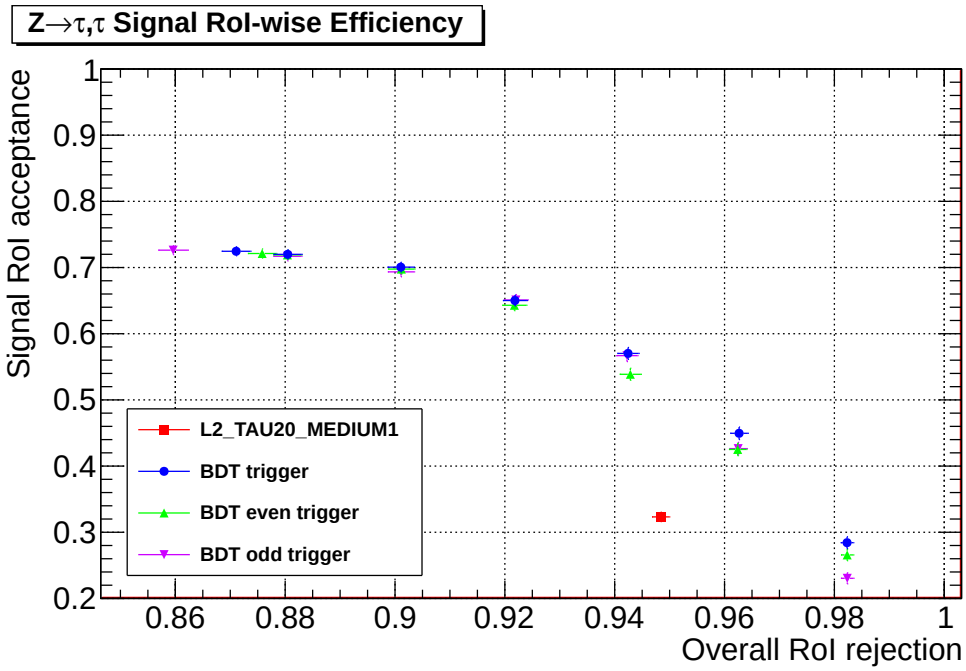


(a)

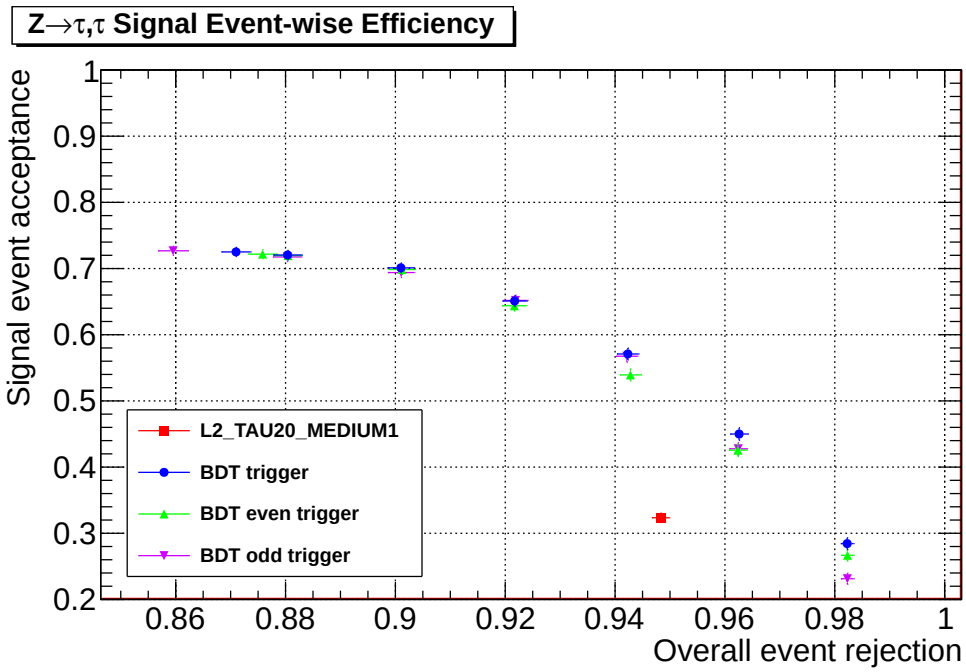


(b)

Figure 6.26: The acceptance rate as a function of the overall acceptance rate for $Z \to \tau\tau$ signal, with error codes for `trkAvgDist`. (a) shows the RoI-wise rate and (b) shows the event-wise rate.

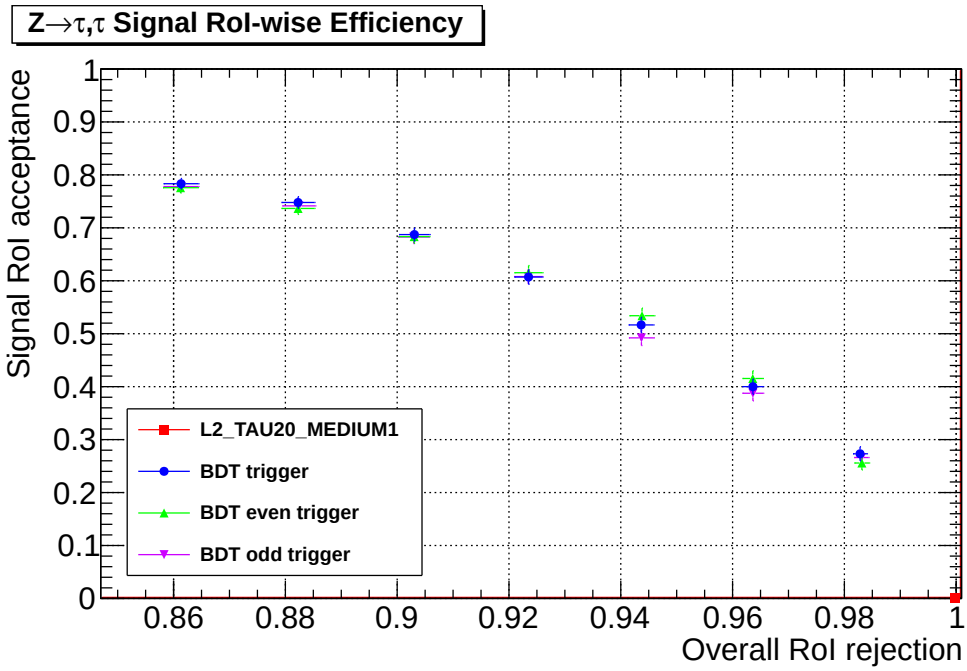


(a)

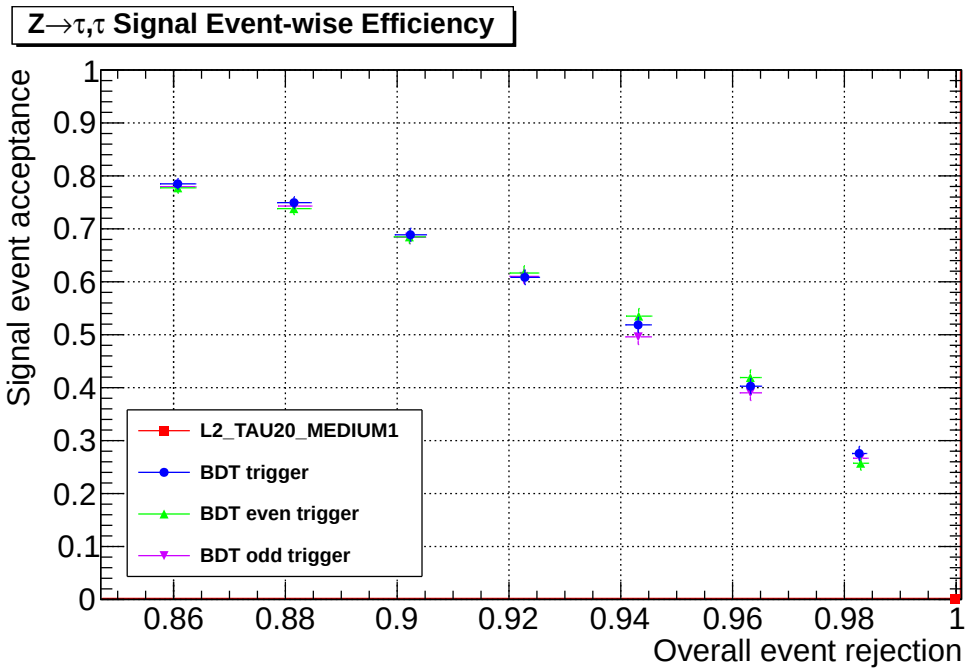


(b)

Figure 6.27: The acceptance rate as a function of the overall acceptance rate for $Z \to \tau\tau$ signal, with error codes for `HADRadius` and `trkAvgDist`. (a) shows the RoI-wise rate and (b) shows the event-wise rate.

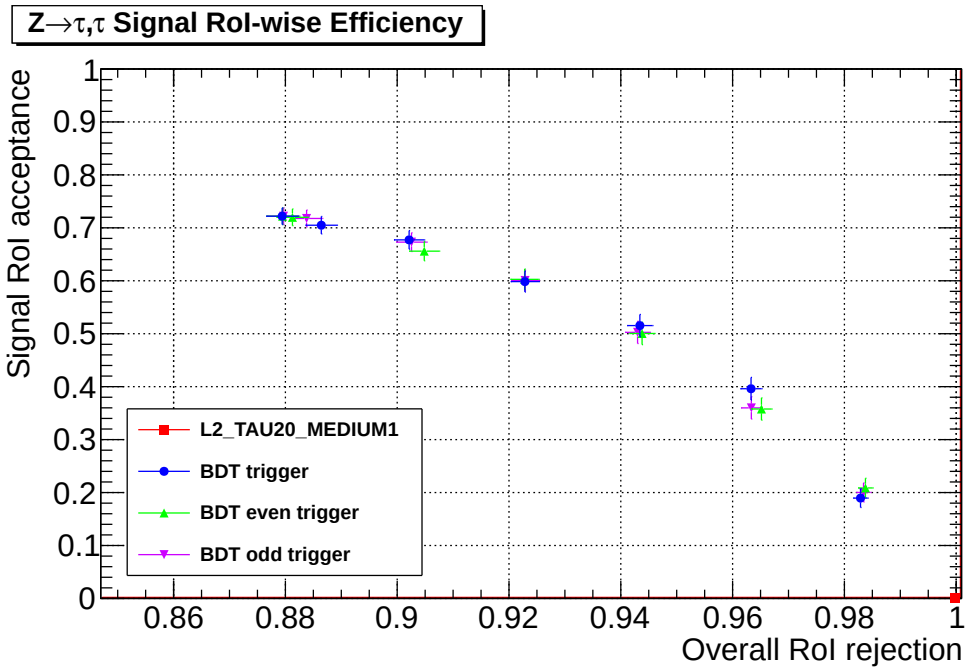


(a)

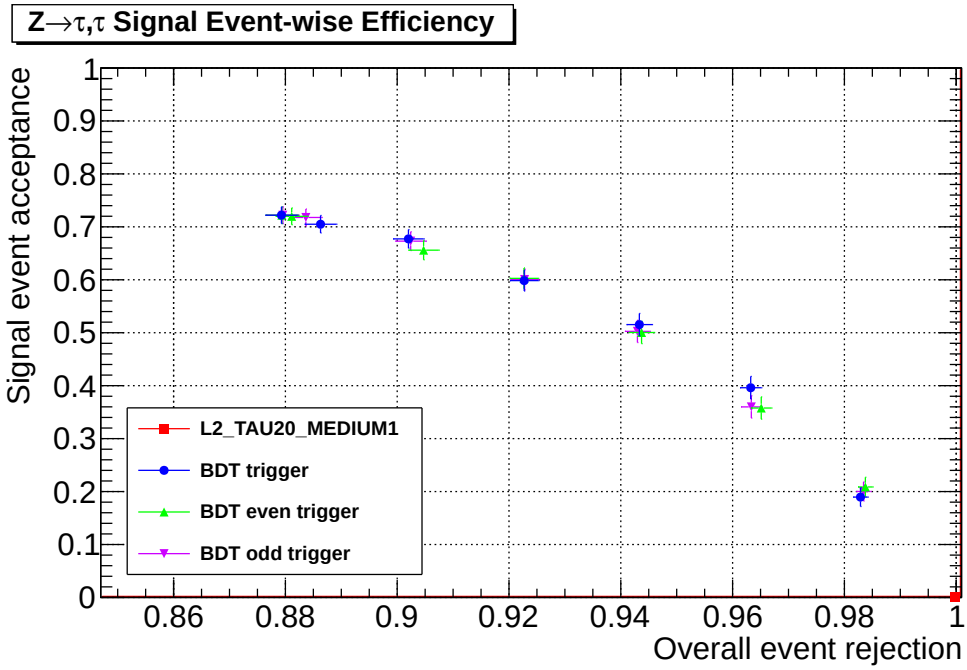


(b)

Figure 6.28: The acceptance rate as a function of the overall acceptance rate for $Z \to \tau\tau$ signal, with error codes for `etOverPtLeadTrk`. (a) shows the RoI-wise rate and (b) shows the event-wise rate.

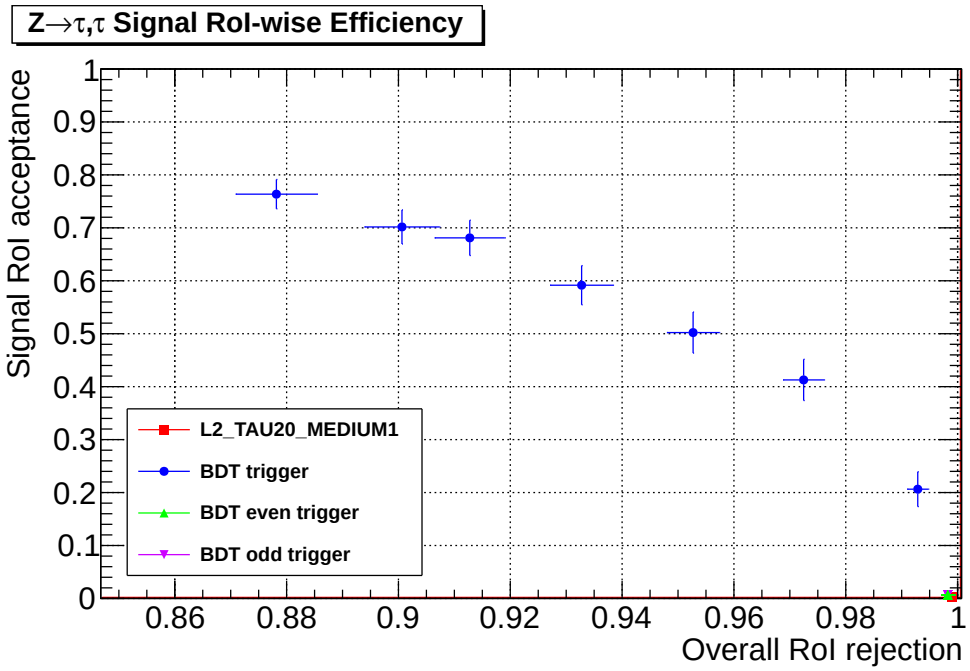


(a)

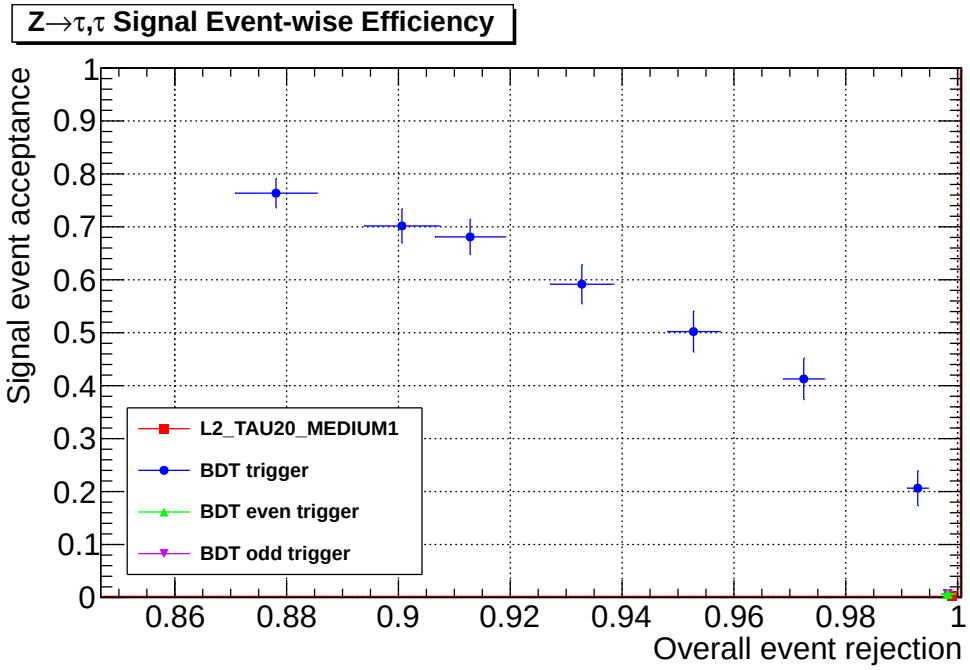


(b)

Figure 6.29: The acceptance rate as a function of the overall acceptance rate for $Z \to \tau\tau$ signal, with error codes for `HADRadius` and `etOverPtLeadTrk`. (a) shows the RoI-wise rate and (b) shows the event-wise rate.

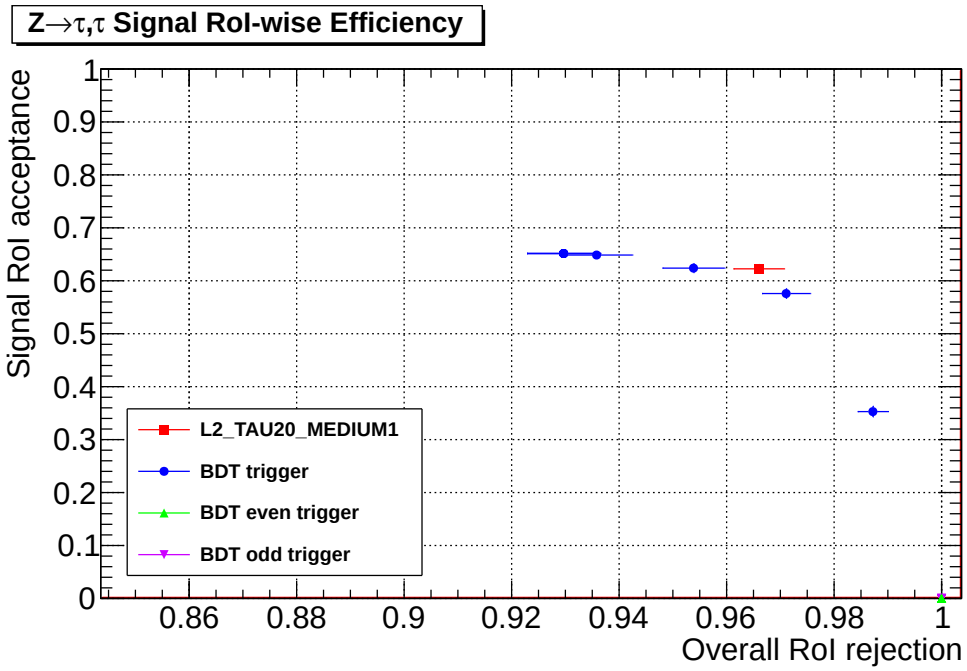


(a)

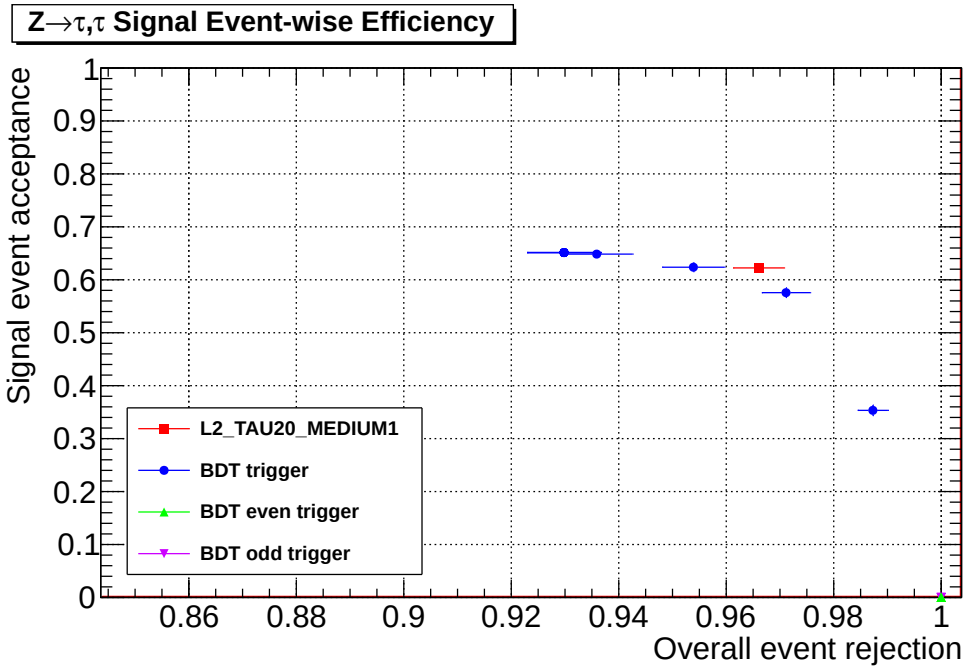


(b)

Figure 6.30: The acceptance rate as a function of the overall acceptance rate for $Z \to \tau\tau$ signal, with error codes for `trkAvgDist` and `etOverPtLeadTrk`. (a) shows the RoI-wise rate and (b) shows the event-wise rate.

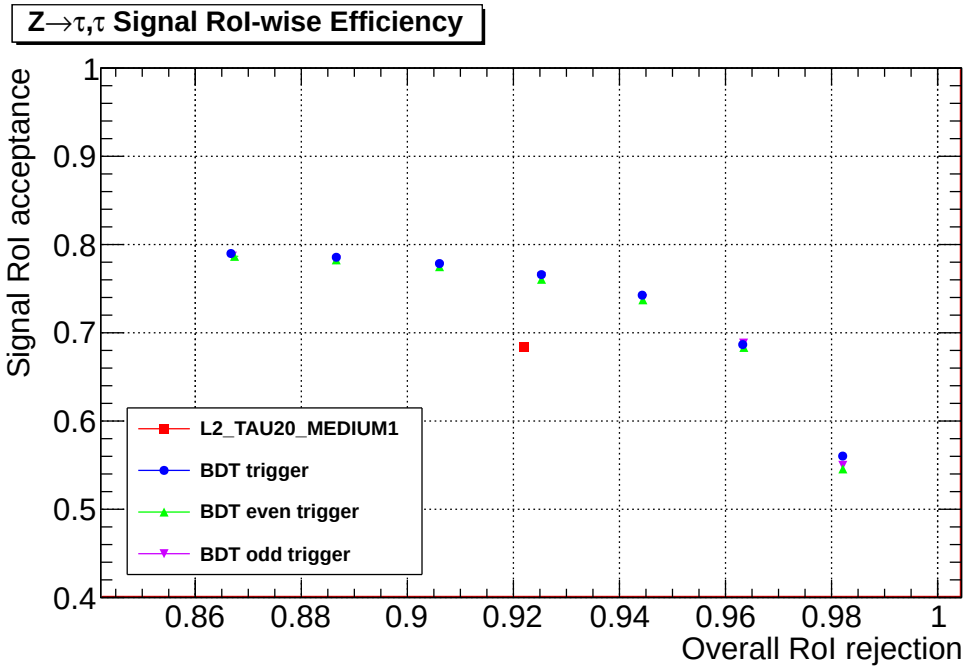


(a)

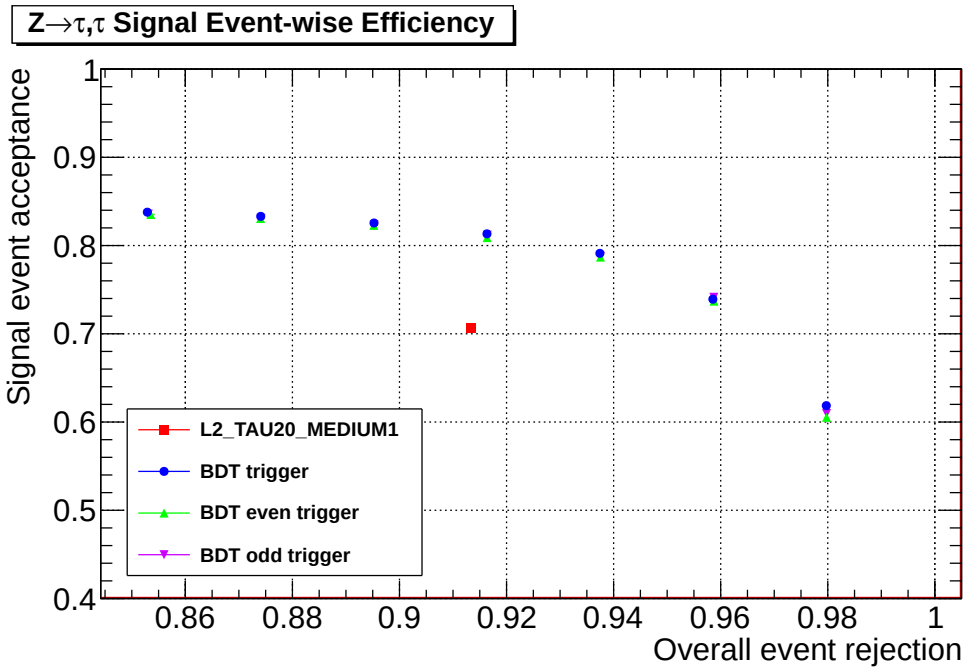


(b)

Figure 6.31: The acceptance rate as a function of the overall acceptance rate for $Z \to \tau\tau$ signal, with error codes for `stripWidth`. (a) shows the RoI-wise rate and (b) shows the event-wise rate.



(a)



(b)

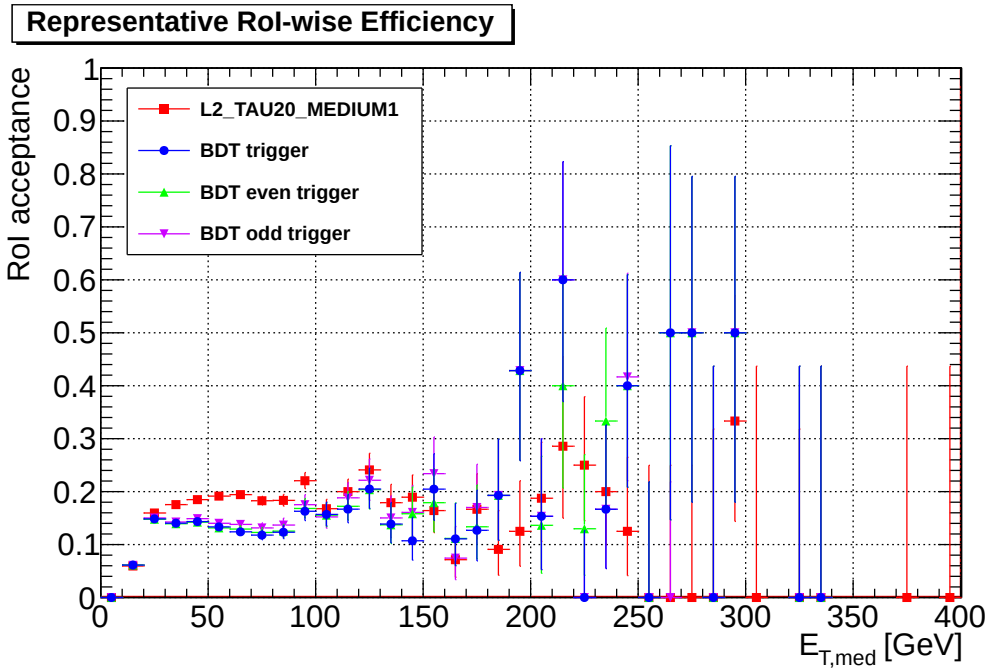
Figure 6.32: The acceptance rate as a function of the overall acceptance rate for $Z \to \tau\tau$ signal. (a) shows the RoI-wise rate and (b) shows the event-wise rate.

6.10 Energy Dependence

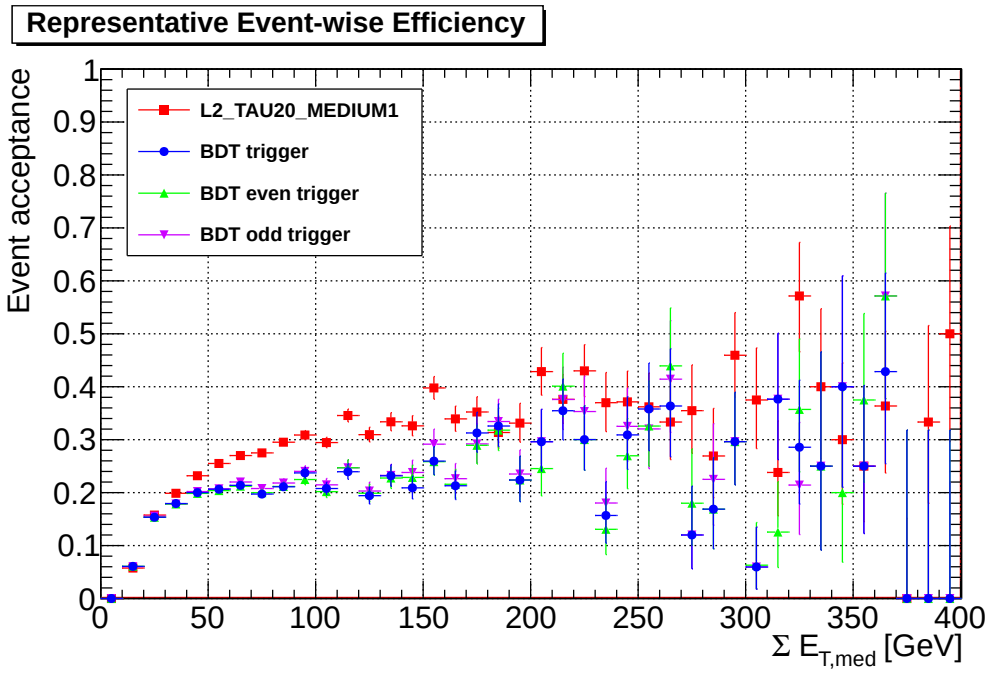
Our next results focus on the energy dependence of the acceptance rates. It is important to ensure that the algorithm performs well on a wide range of energy levels. In particular, the higher levels of energy are expected to be more accessible, and hence, more important in the future. Figures 6.33 to 6.36 show the E_T dependent rates for the training samples.

For the representative sample, we can see that, for each level of energy, the BDT trigger's acceptance rate is clearly lower or about equal to that of the reference trigger. As the energy level increases, the difference between the rates becomes greater, but eventually becomes less clear due to the low number of statistics. However, this is not the case for dijet background. While the BDT trigger's rate for this sample eventually crosses and falls below the reference trigger's rate, the BDT trigger has a clearly higher acceptance rate for low energy levels than that for the reference trigger. Also, the difference for the higher energy levels quickly becomes unclear due again to the lower number of statistics. This underperformance may be a problem depending on how important the rejection of dijet background is. By design, we used the representative sample to measure the acceptance rate limit, and optimized our algorithm to accept only slightly less than the reference trigger for this sample. One modification to fix this problem may be to add to the optimization procedure an additional constraint for the acceptance rate of dijet background. For the signal samples, the dominance of the BDT trigger over the entire spectrum of energy is very clear. Interestingly, for high levels of energy, the reference trigger RoI-wise rate for $Z' \to \tau\tau$ signal slopes down steeply with E_T. An analysis of the E_T dependence of the discriminating variables for the reference trigger suggests that this is largely due to the trigger's cuts on `etOverPtLeadTrk`, a variable that is found to have a significant amount of positive correlatation with $E_{T,\text{med}}$, especially for the single-prong case. This highlights the unique ability of multivariate triggers to perform well on a wide variety of energy levels, and to scale appropriately when energy levels increase in the future. This can have a significant impact for searches for particles such as the supersymmetric Higgs bosons, that are hypothesized to have a large tau lepton branching fraction, but a greater mass than the

SM Higgs boson. Also notice that the higher energy levels of the Z' as compared to the Z are due to its much greater mass. As mentioned before, a greater mass results in taus that are more collimated, and thus easier to identify, as is apparent by the trend of an increasing acceptance rate of signal with increasing energy.

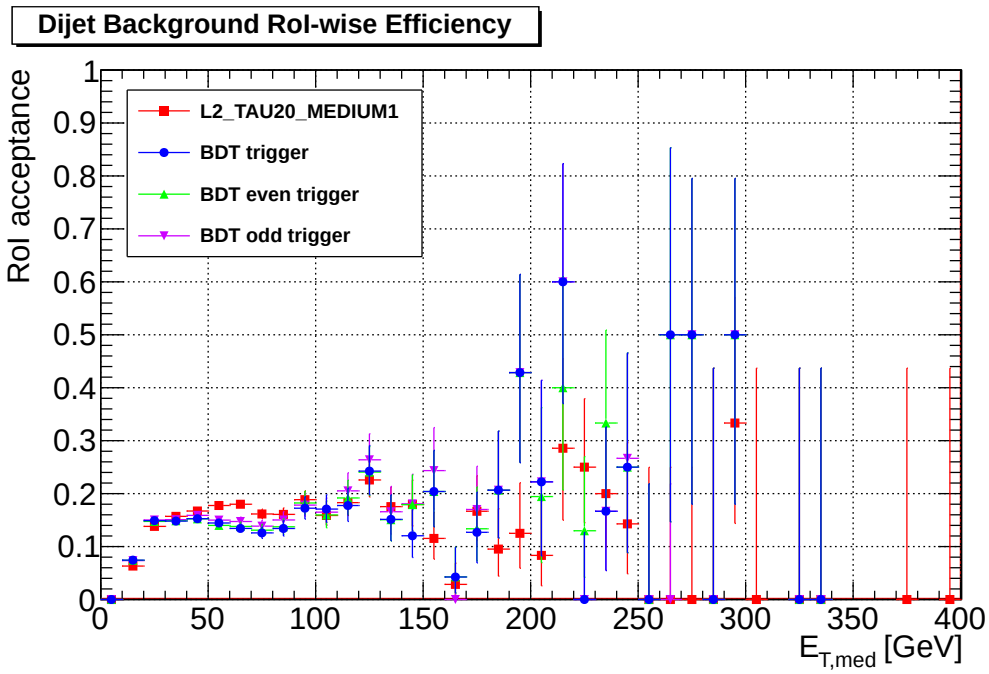


(a)

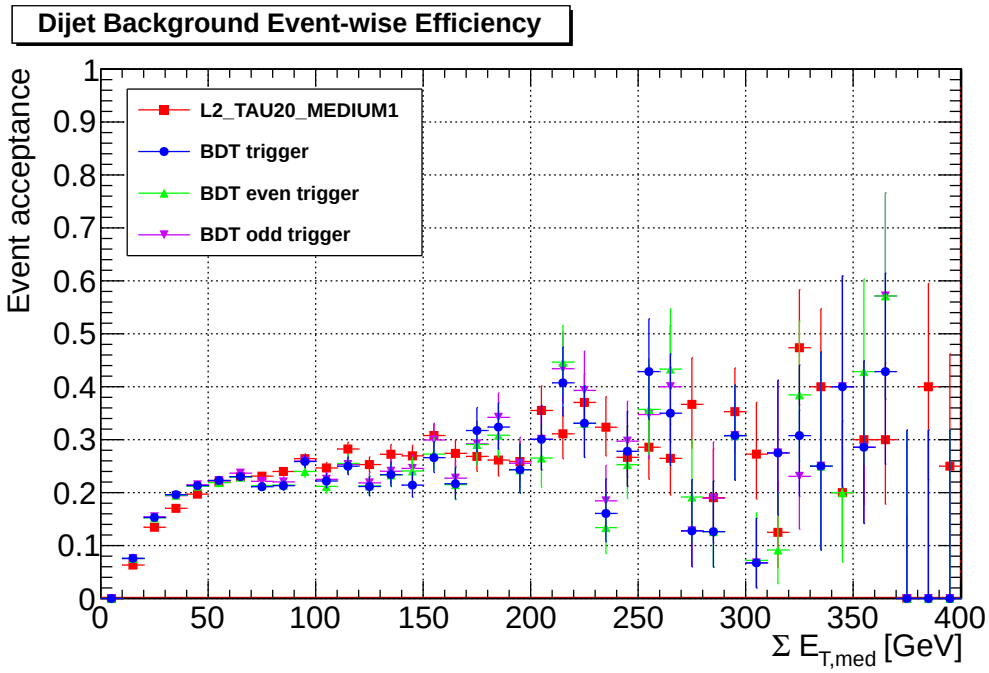


(b)

Figure 6.33: The acceptance rate as a function of transverse energy for the representative sample. (a) shows the RoI-wise case and (b) shows the event-wise case.

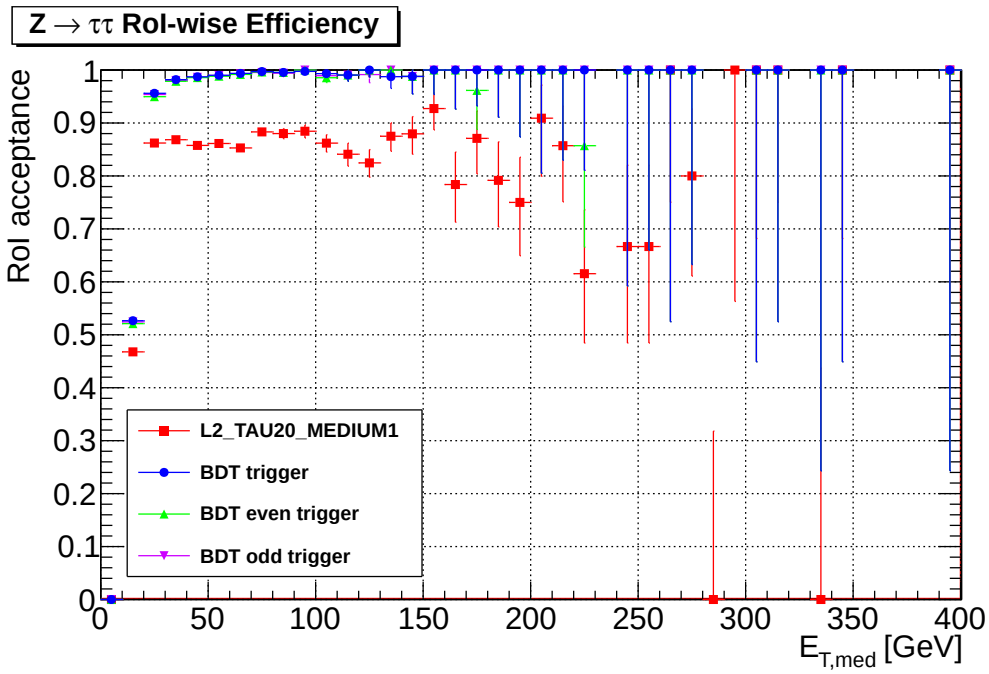


(a)

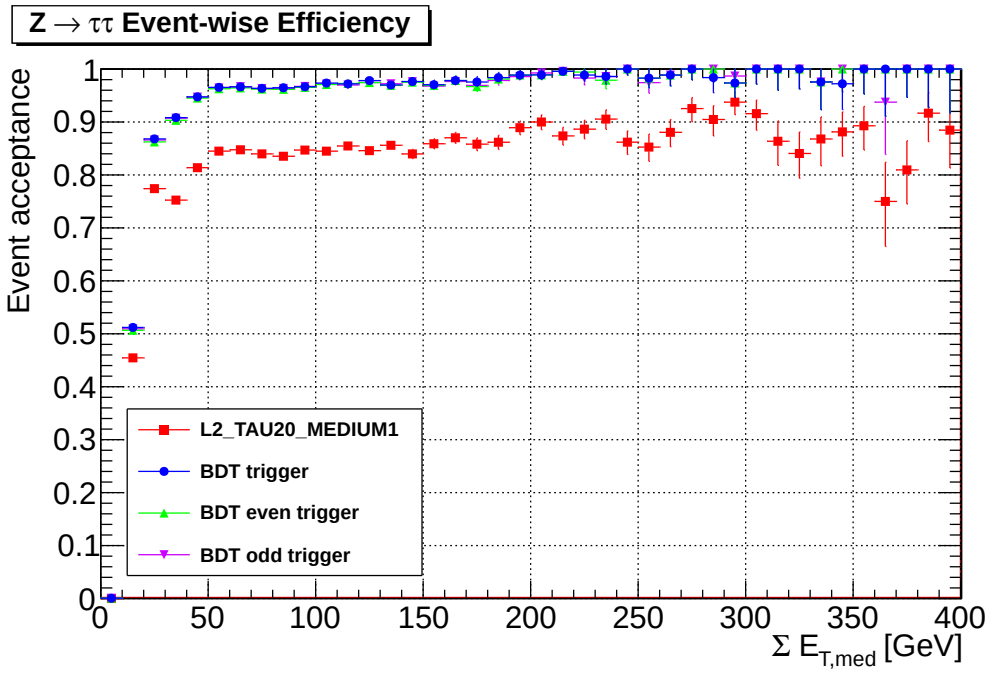


(b)

Figure 6.34: The acceptance rate as a function of transverse energy for the dijet background training sample. (a) shows the RoI-wise case and (b) shows the event-wise case.

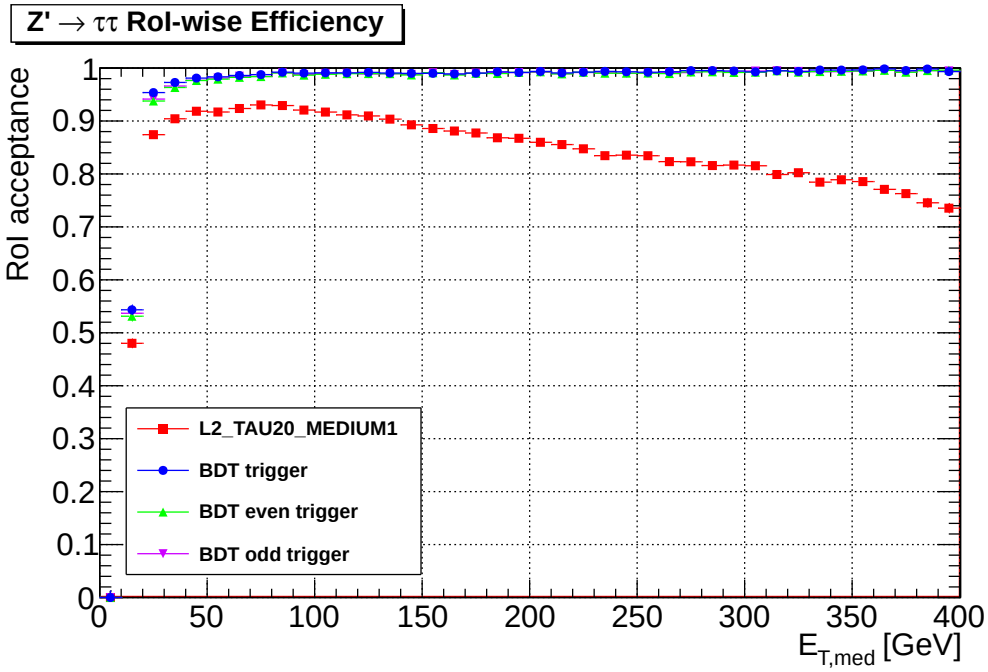


(a)

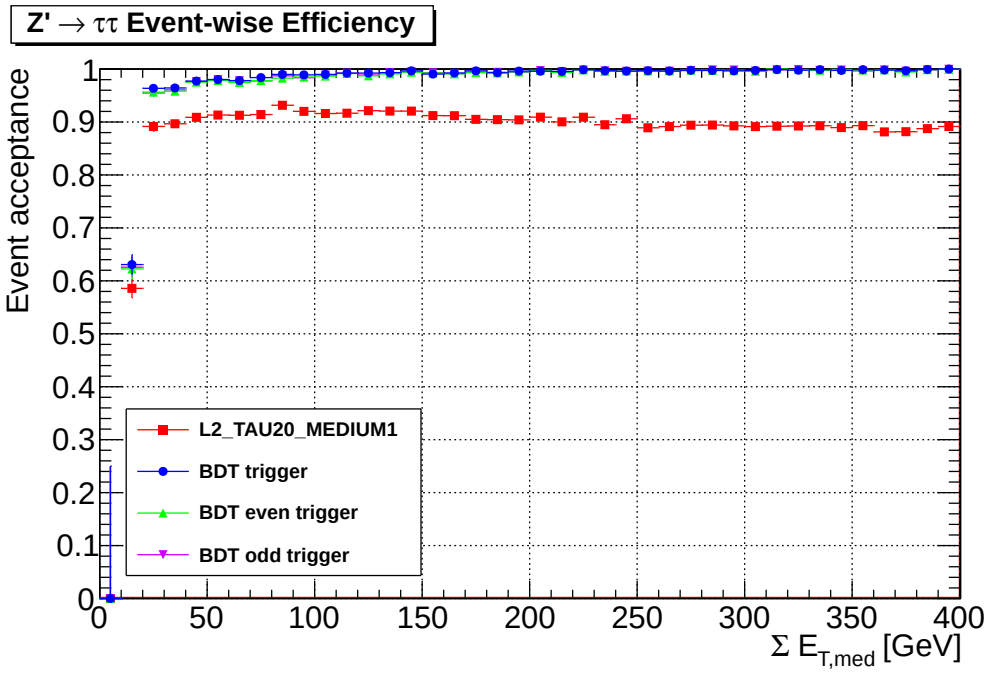


(b)

Figure 6.35: The acceptance rate as a function of transverse energy for the $Z \to \tau\tau$ signal training sample. (a) shows the RoI-wise case and (b) shows the event-wise case.



(a)



(b)

Figure 6.36: The acceptance rate as a function of transverse energy for the $Z' \to \tau\tau$ signal training sample. (a) shows the RoI-wise case and (b) shows the event-wise case.

6.11 Pile-up Dependence

We also want to study the pile-up dependence of the efficiencies. Not only do we wish to have our trigger perform better than the reference trigger for each level of pile-up, but we also want to ensure that the dependence of our trigger's efficiency on pile-up is below a reasonable limit. As in Section 6.8, we measure this dependence by using the linear fits to each efficiency versus pile-up graph, and comparing the slopes of these fits. Figures 6.37 to 6.40 show the pile-up dependent efficiencies for the training samples, and Table 6.10 summarizes the pile-up dependence difference for each case shown in the figures.

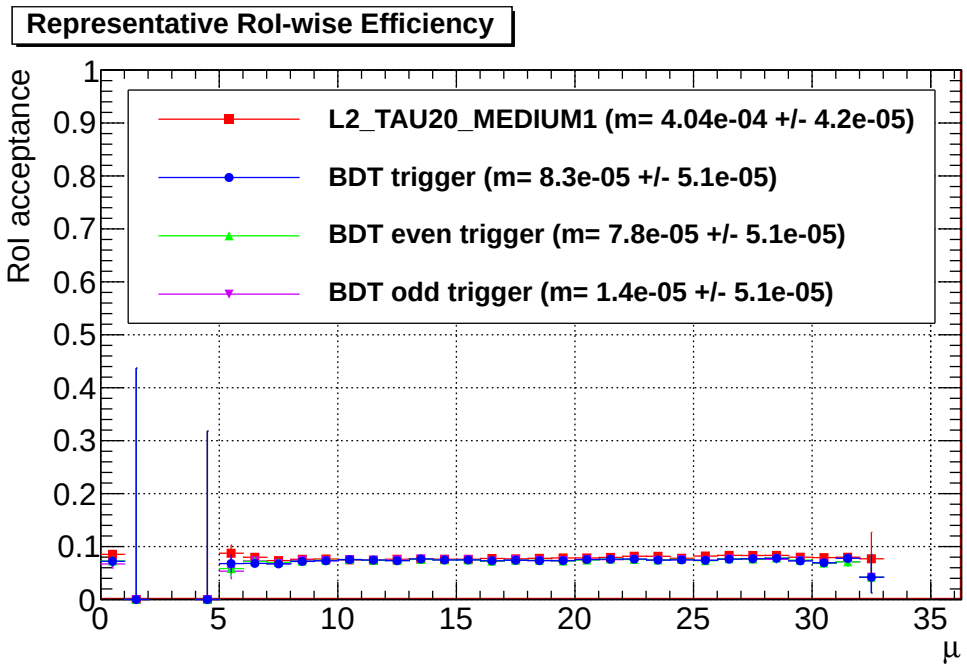
Sample	**RoI-wise**	**Event-wise**
Representative	$(-3.21 \pm 0.66 \pm 0.64) \times 10^{-4}$	$(-3.64 \pm 0.74 \pm 0.68) \times 10^{-4}$
Dijet Background	$(-1.52 \pm 1.23 \pm 0.84) \times 10^{-4}$	$(-0.96 \pm 1.49 \pm 1.00) \times 10^{-4}$
$Z \to \tau\tau$ Signal	$(7.82 \pm 0.89 \pm 1.02) \times 10^{-4}$	$(5.21 \pm 0.92 \pm 1.04) \times 10^{-4}$
$Z' \to \tau\tau$ Signal	$(-6.74 \pm 0.88 \pm 0.15) \times 10^{-4}$	$(-6.14 \pm 0.83 \pm 0.01) \times 10^{-4}$

Table 6.10: The difference between the absolute value of the slope of the linear fit for the BDT trigger with that for the reference trigger, for each of the 8 cases presented in this section.

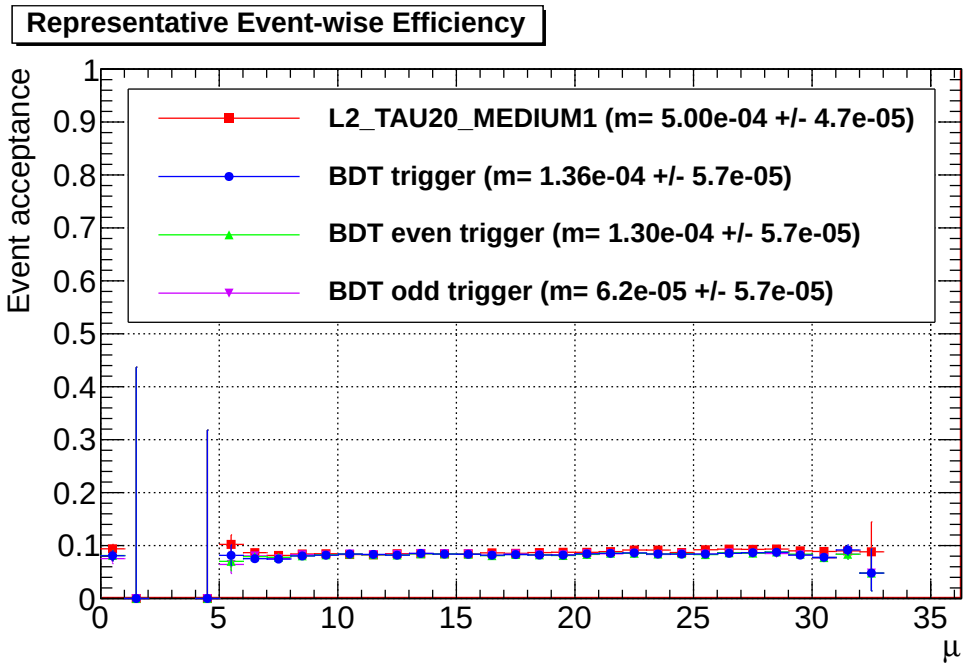
From the graphs we see that the BDT trigger, applied to the representative sample, generally has a slightly greater or equal performance for each value of μ. As for the energy dependence, the situation is not the same for dijet background. The BDT trigger generally has a slightly worse performance for each value of μ, when applied to the dijet background sample. As expected, the pile-up dependent performance of the BDT trigger on the signal samples is consistently much greater than that for the reference trigger.

As for the amount of pile-up dependence, we see that, for the representative sample, the slopes for the BDT trigger are less than the corresponding slopes for reference trigger, with 2σ certainty. For dijet background, the slopes are equivalent within 1σ deviation. For $Z \to \tau\tau$ signal, the BDT trigger's slopes are greater than the corresponding slopes for the reference trigger, with 4σ certainty for the RoI-wise case and 2σ certainty for the event-wise case. The RoI-wise slope for the BDT trigger for this sample is below 8.6250% efficiency per 50 interactions per bunch crossing, with 3σ certainty. For the reference trigger, this

number is 3.1550%. For the event-wise rates, this number is 7.6450% for the BDT trigger and 3.4800% for the reference trigger. For $Z' \to \tau\tau$ signal, the BDT trigger's slopes are less than the corresponding slopes for the reference trigger, with 6σ certainty for the RoI-wise case and 7σ certainty for the event-wise case. Whether the BDT trigger's greater pile-up dependence for the $Z \to \tau\tau$ sample is acceptable depends on the specific requirements of the tau trigger group. An upward sloping dependence for signal, such as the one we have, may be more favourable than one that is negatively sloped, and a greater measured upward sloping dependence provides more assurance that the dependence is not really downward sloping. Also, we found that removing some variables, such as `CaloRadius`, from the BDT training program does have the effect of lowering the pile-up dependence, with only a small effect on the signal acceptance rate.

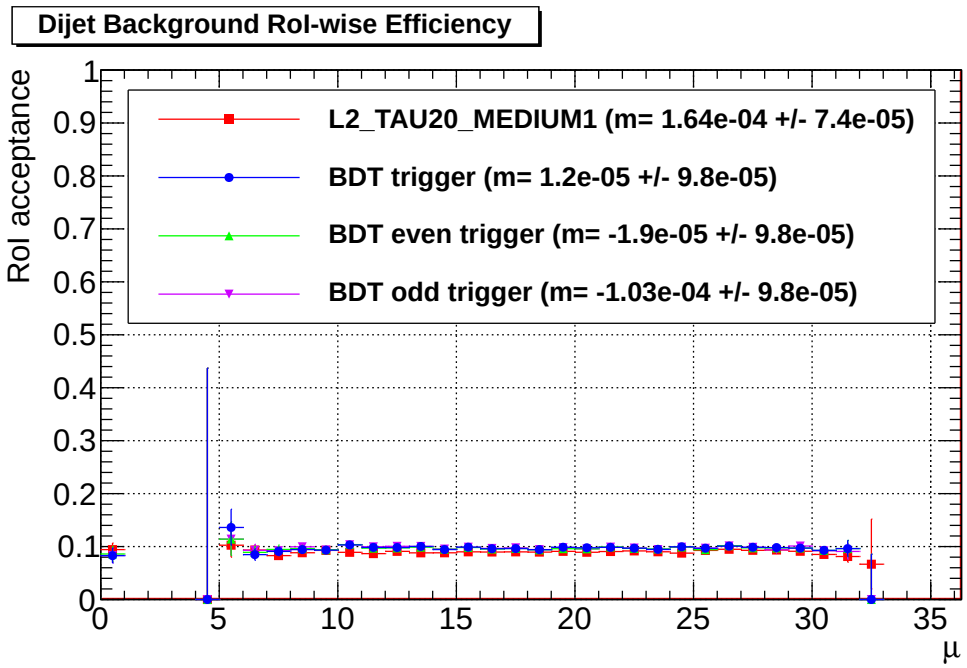


(a)

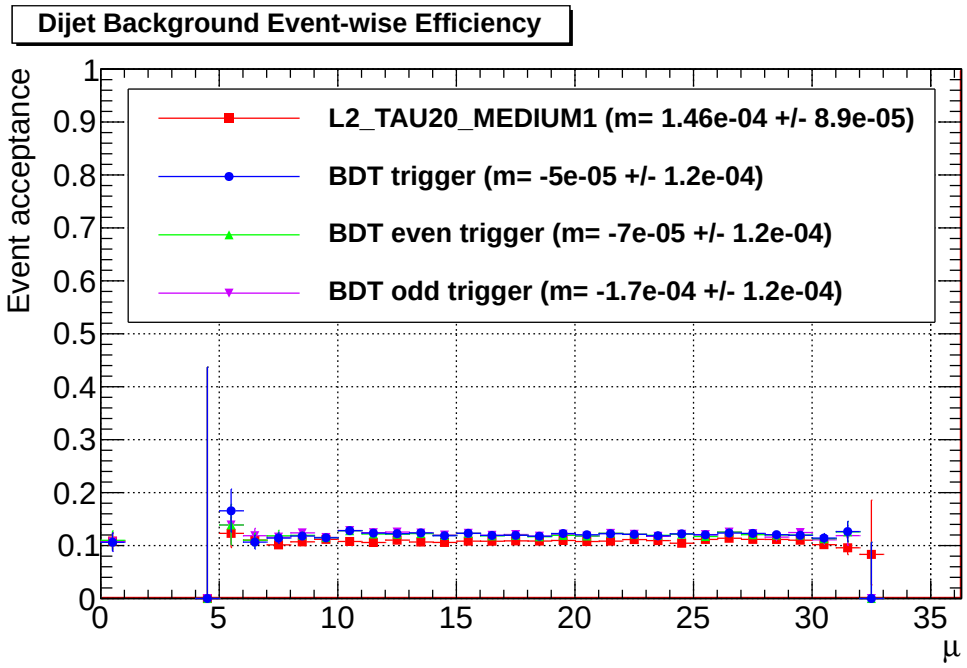


(b)

Figure 6.37: The acceptance rate as a function of pile-up for the representative sample. (a) shows the RoI-wise case and (b) shows the event-wise case.

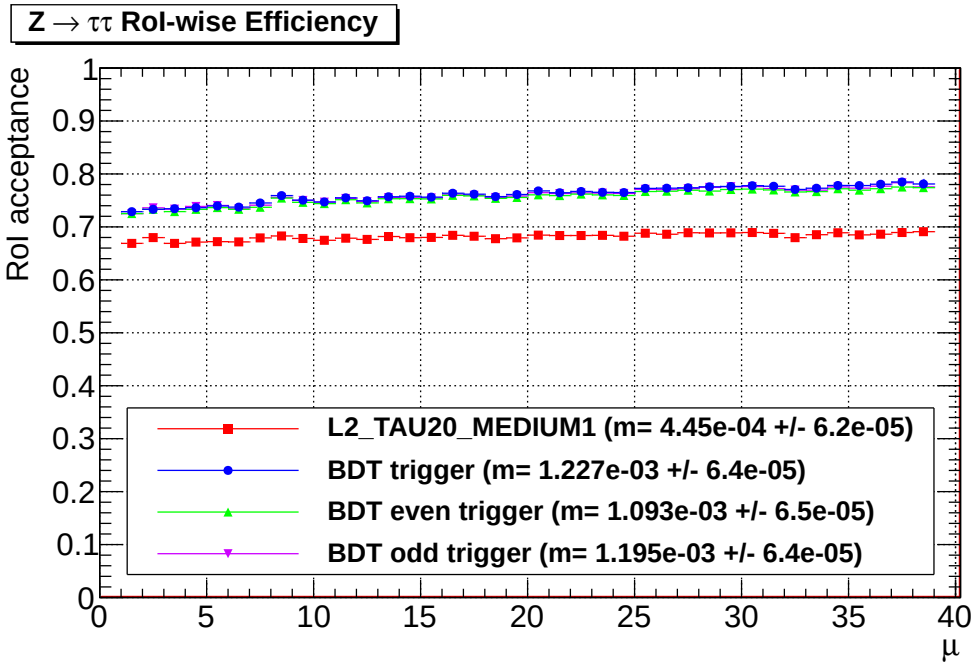


(a)

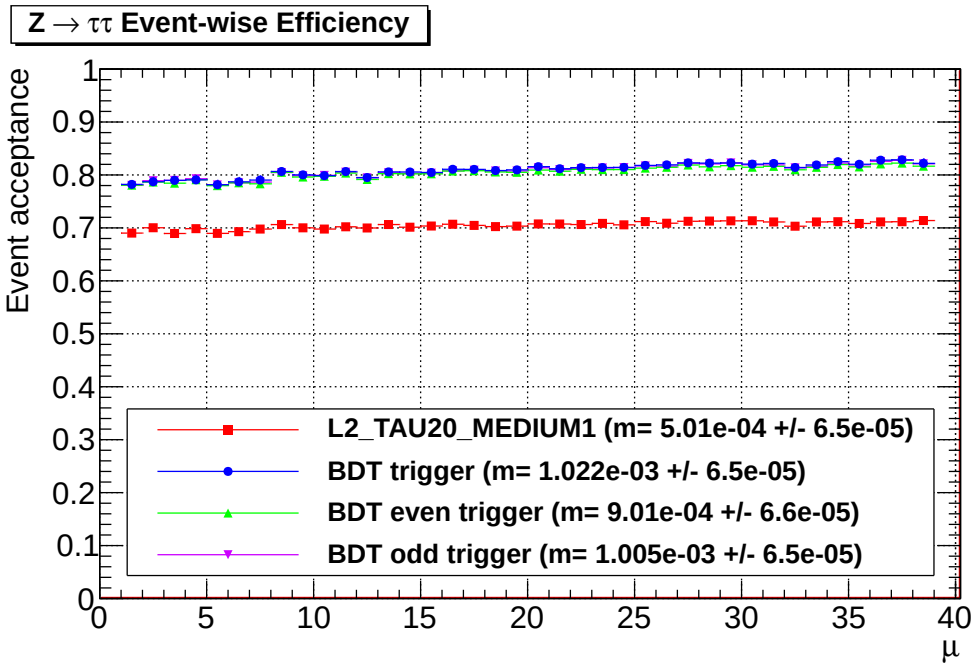


(b)

Figure 6.38: The acceptance rate as a function of transverse energy for the dijet background training sample. (a) shows the RoI-wise case and (b) shows the event-wise case.

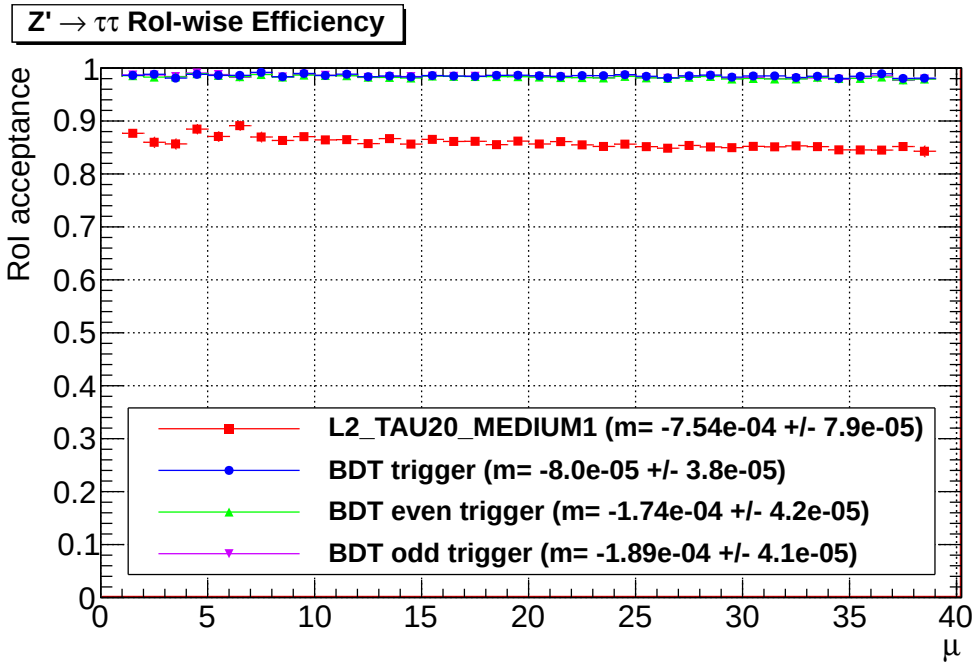


(a)

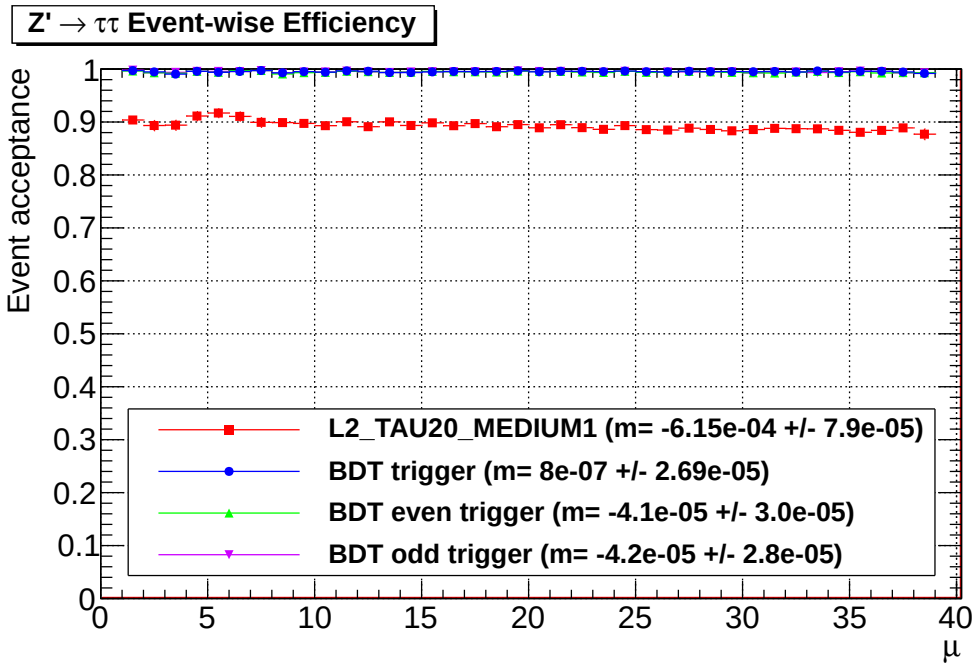


(b)

Figure 6.39: The acceptance rate as a function of transverse energy for the $Z \to \tau\tau$ signal training sample. (a) shows the RoI-wise case and (b) shows the event-wise case.



(a)



(b)

Figure 6.40: The acceptance rate as a function of transverse energy for the $Z' \to \tau\tau$ signal training sample. (a) shows the RoI-wise case and (b) shows the event-wise case.

6.12 Other Samples

We now show an overview of the results for a variety of samples that were not used for training. The goal is to show that our algorithm is robust in finding taus from a wide variety of decay channels and run conditions, without having to explicitly train it for all such cases. The results for Period E, overall and dijet background, are shown in Figures 6.41 and 6.42, respectively.

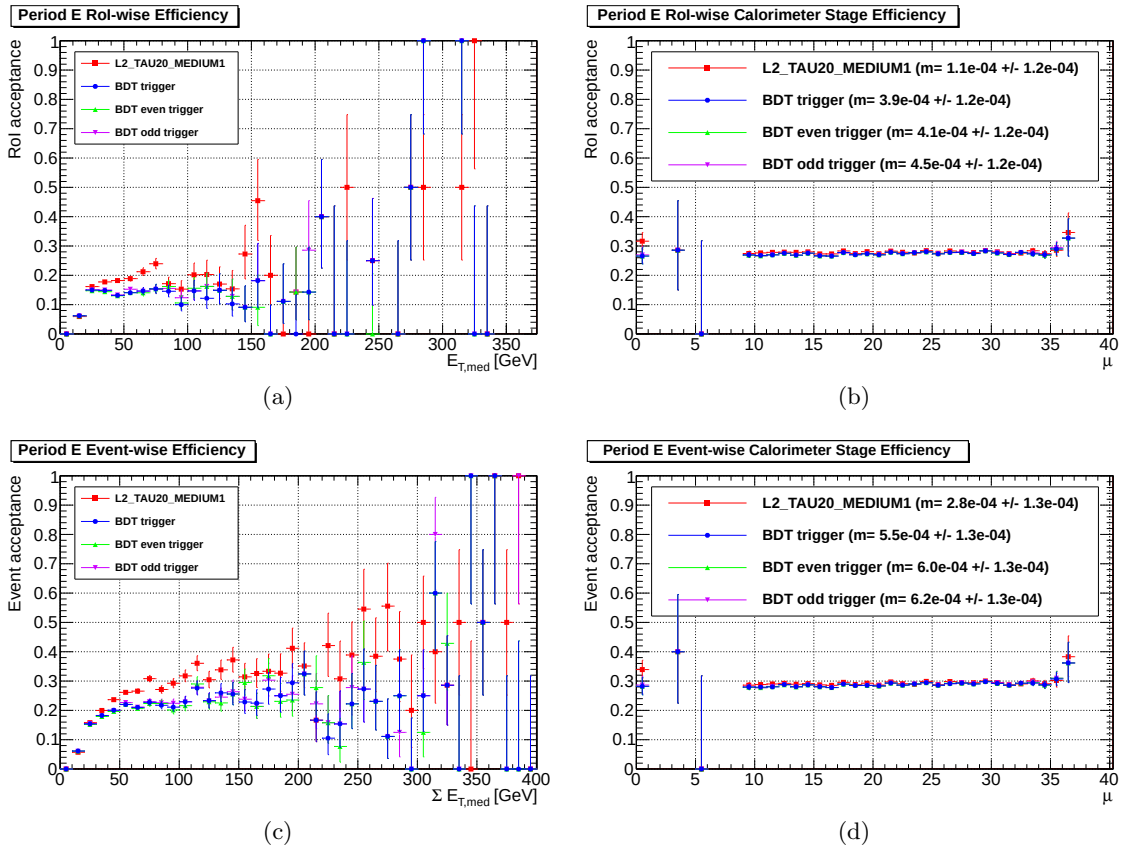


Figure 6.41: The acceptance rate for the data taking period E. (a) shows the RoI-wise acceptance rate w.r.t. L2 E_T, (b) shows the RoI-wise acceptance rate w.r.t. μ, (c) shows the event-wise acceptance rate w.r.t. $\sum$L2 E_T, and (d) shows the event-wise acceptance rate w.r.t. μ.

For the overall period E, the graphs are similar to those shown for the representative sample in Sections 6.10 and 6.11. The energy dependent rates of the BDT trigger are generally less than or equal to the corresponding rates of the reference trigger for each level

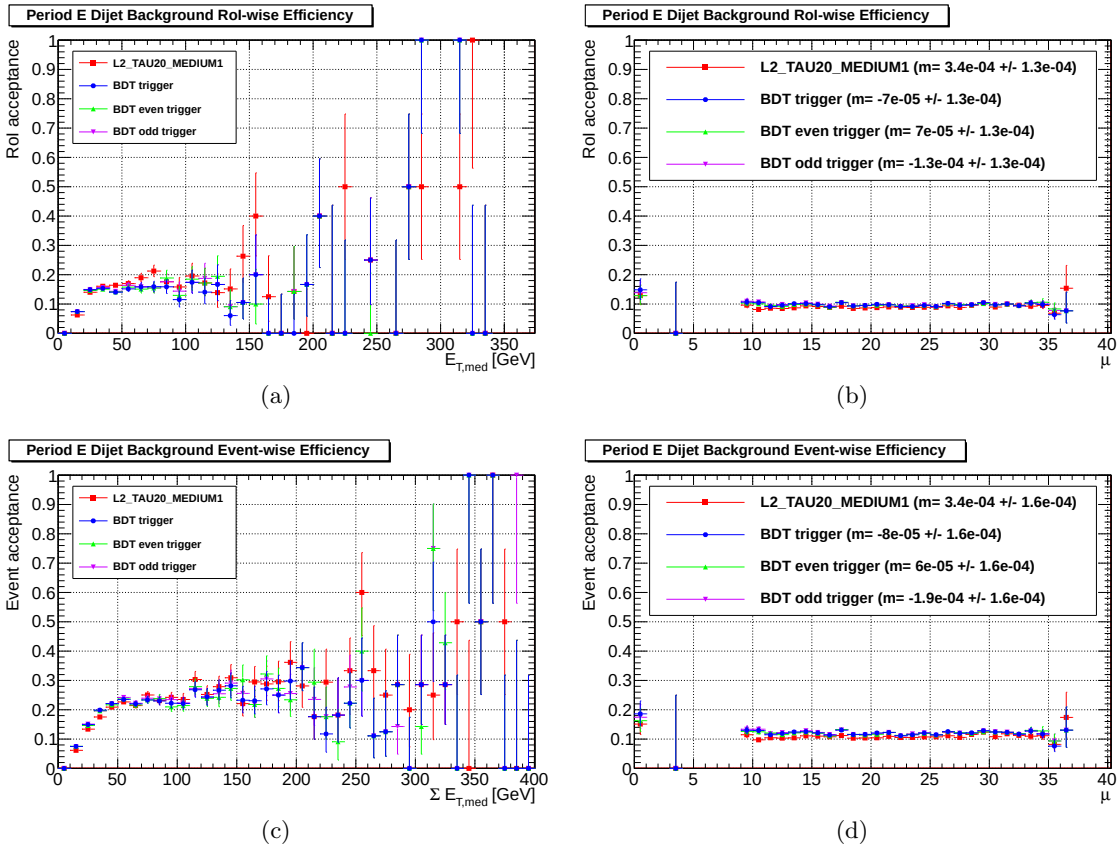


Figure 6.42: The acceptance rate for dijet background from the data taking period E. (a) shows the RoI-wise acceptance rate w.r.t. L2 E_T, (b) shows the RoI-wise acceptance rate w.r.t. μ, (c) shows the event-wise acceptance rate w.r.t. $\sum$L2 E_T, and (d) shows the event-wise acceptance rate w.r.t. μ.

of energy, and the separation becomes wider with increasing levels of energy, until some point where the difference is unclear due to the lower amount of statistics. The pile-up dependent rates of the BDT trigger are slightly less than or equal to the corresponding rates of the reference trigger, and the slopes of the BDT trigger are less than the corresponding slopes of the reference trigger, with 1σ certainty. For the dijet background from this period, the results are similar to those shown for the dijet background from the representative sample (same sections). For low levels of energy, the energy dependent rates of the BDT trigger are generally slightly greater than the corresponding rates of the reference trigger. They eventually cross at some point, with the separation becoming wider with increasing levels

of energy, until another point where the difference is unclear due to the lower amount of statistics. The pile-up dependent rates of the BDT trigger are slightly greater than or equal to the corresponding rates of the reference trigger, and the slopes of the BDT trigger are equivalent to the corresponding slopes of the reference trigger, within 1σ deviation.

The results for the signal MC samples are shown in Figures 6.43 to 6.48.

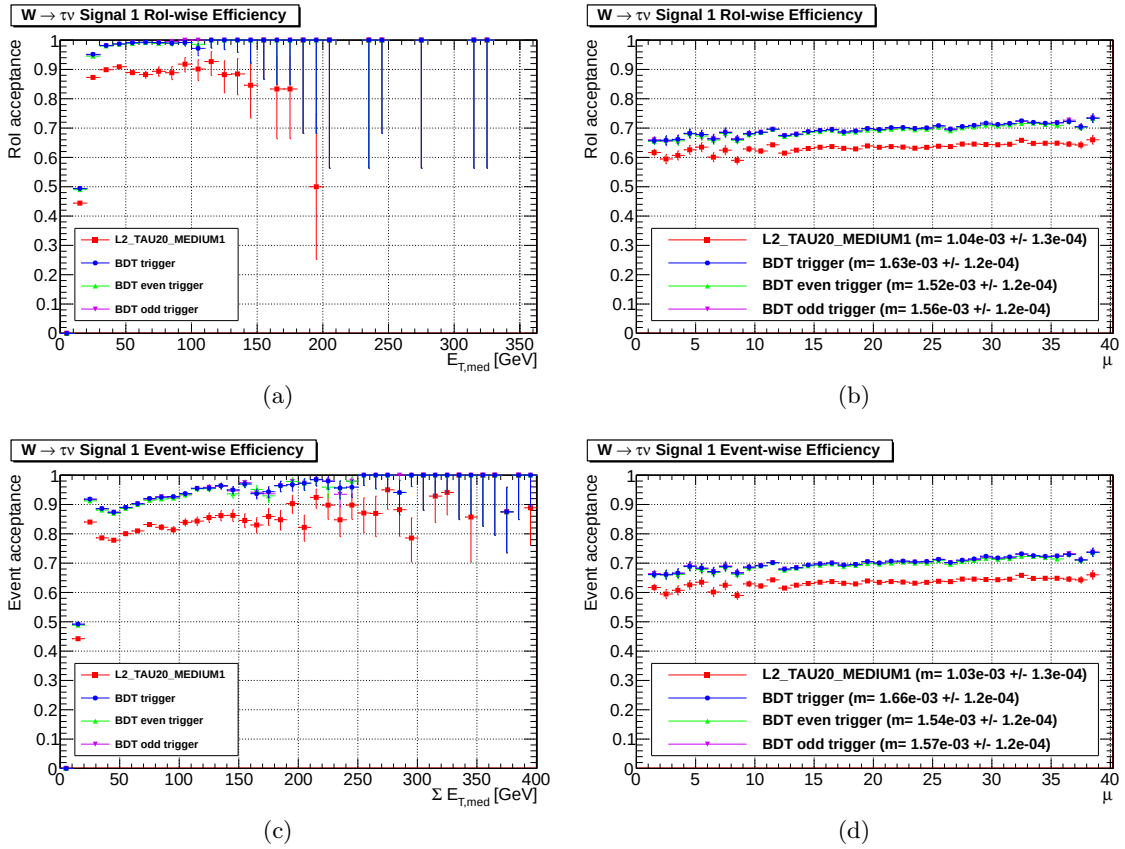


Figure 6.43: The acceptance rate for $W \to \tau\nu$ signal from the ATLAS dataset tagged as `147812.e1176_s1479_s1470_r3553_r3549_p1130`. (a) shows the RoI-wise acceptance rate w.r.t. L2 E_T, (b) shows the RoI-wise acceptance rate w.r.t. μ, (c) shows the event-wise acceptance rate w.r.t. $\sum$L2 E_T, and (d) shows the event-wise acceptance rate w.r.t. μ.

The first two samples consist of $W \to \tau\nu$ signal events, which represent an important channel for tau detection at the LHC. Again, the performance of the BDT trigger generally exceeds the performance of the reference trigger for these types of events. However, the slopes of the pile-up dependence for the BDT trigger are significantly greater than the

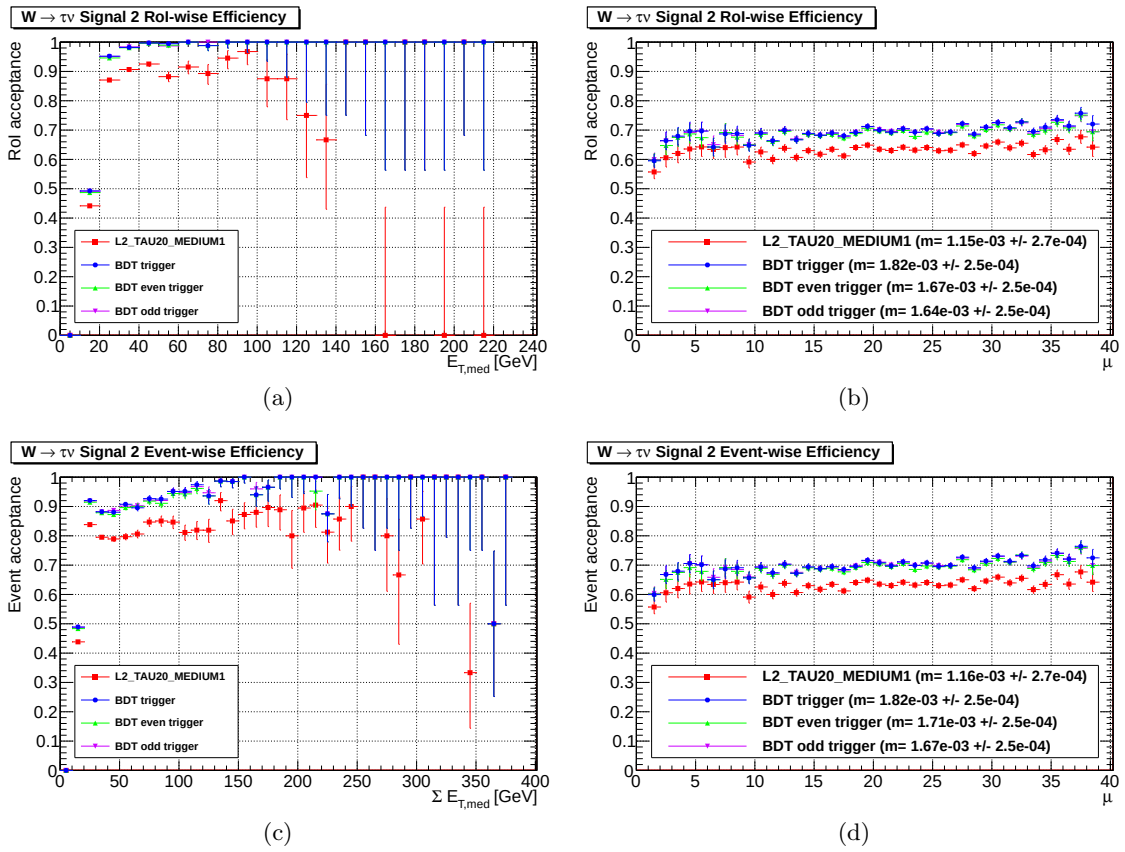


Figure 6.44: The acceptance rate for $W \to \tau\nu$ signal from the ATLAS dataset tagged as `147812.e1176_s1486_s1473_r3553_r3549_p1130`. (a) shows the RoI-wise acceptance rate w.r.t. L2 E_T, (b) shows the RoI-wise acceptance rate w.r.t. μ, (c) shows the event-wise acceptance rate w.r.t. $\sum$L2 E_T, and (d) shows the event-wise acceptance rate w.r.t. μ.

corresponding slopes for the reference trigger. They have the same order of magnitude as the corresponding slopes of the pile-up dependence for the $Z \to \tau\tau$ signal training sample, shown in Section 6.11, but the level of certainty is lower due to the lower number of statistics. Next, we have a $Z \to \tau\tau$ signal sample. The results are similar as for the $Z \to \tau\tau$ signal training sample: the BDT trigger's performance exceeds that of the reference trigger, except for the pile-up dependence; and by a similar amount as for the training sample. The next two samples consist of $H \to \tau\tau$ events where both taus are hadronic, with the first sample using a Higgs mass of $125\,\mathrm{GeV}/c^2$ and the other using a Higgs mass of $130\,\mathrm{GeV}/c^2$. Similar to the Z' boson, the Higgs boson has a higher mass than the Z boson, and hence the resulting

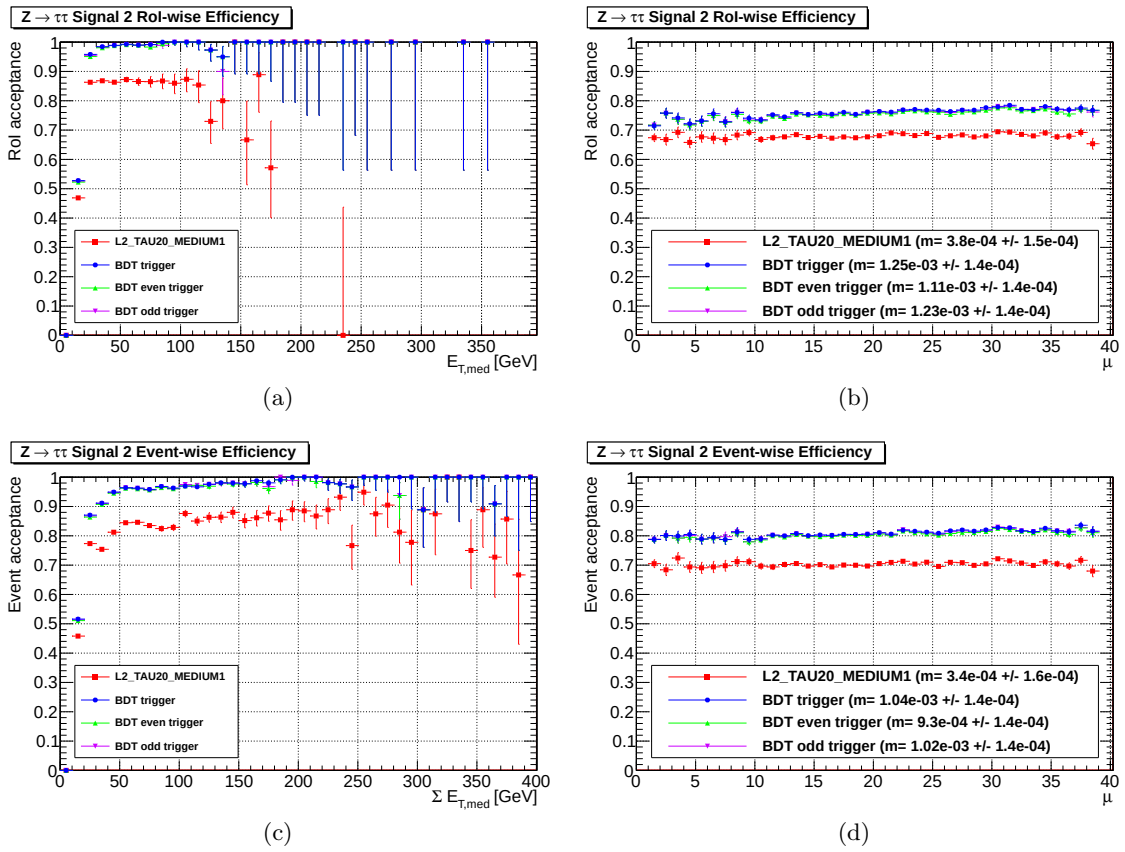


Figure 6.45: The acceptance rate for $Z \to \tau\tau$ signal from the ATLAS dataset tagged as `147818.e1176_s1486_s1473_r3553_r3549_p1130`. (a) shows the RoI-wise acceptance rate w.r.t. L2 E_T, (b) shows the RoI-wise acceptance rate w.r.t. μ, (c) shows the event-wise acceptance rate w.r.t. $\sum$L2 E_T, and (d) shows the event-wise acceptance rate w.r.t. μ.

taus tend to be more energetic and easier to identify. We can see that the performance results for both Higgs boson samples are similar to those for the $Z' \to \tau\tau$ training sample: the BDT trigger's performance exceeds that of the reference trigger, except for the amount of pile-up dependence where the two triggers are equivalent for all but one case (H(125) event-wise) where the BDT trigger has a smaller slope with 2σ certainty. The next sample is a $Z' \to \tau\tau$ sample with 1250 GeV/c^2 as the mass of the Z' boson. Again, the performance results for this sample are similar to those for the $Z' \to \tau\tau$ training sample.

Lastly, Figures 6.49 to 6.50 show the results for collision data signal taken from the representative sample and the data taking period E, respectively.

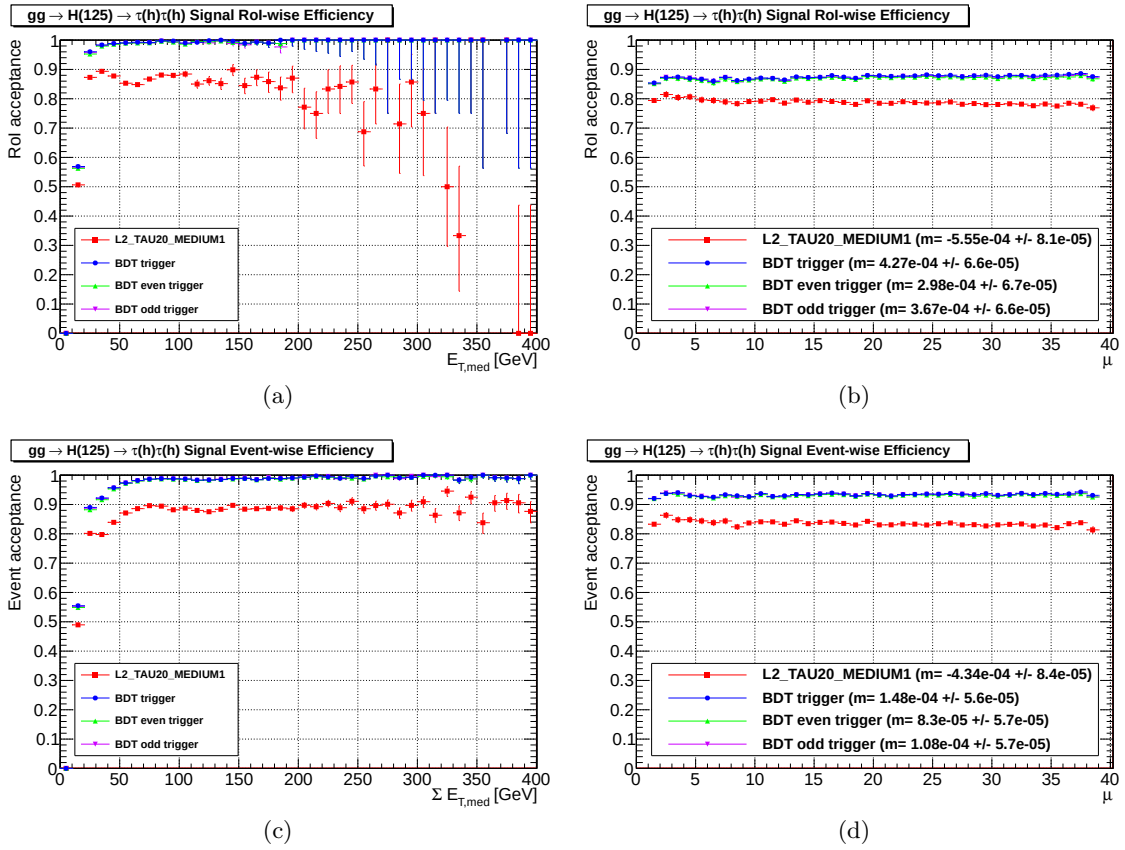


Figure 6.46: The acceptance rate for $gg \rightarrow H(125) \rightarrow \tau(h)\tau(h)$ signal from the ATLAS dataset tagged as `161577.e1217_s1469_s1470_r3542_r3549_p1130`. (a) shows the RoI-wise acceptance rate w.r.t. L2 E_T, (b) shows the RoI-wise acceptance rate w.r.t. μ, (c) shows the event-wise acceptance rate w.r.t. $\sum$L2 E_T, and (d) shows the event-wise acceptance rate w.r.t. μ.

For both cases, the BDT trigger generally outperforms the reference trigger, except for the amount of pile-up dependence, where the performance is equivalent, within 1σ deviation, for all but one case (representative event-wise), where the BDT trigger has a greater slope with 1σ certainty.

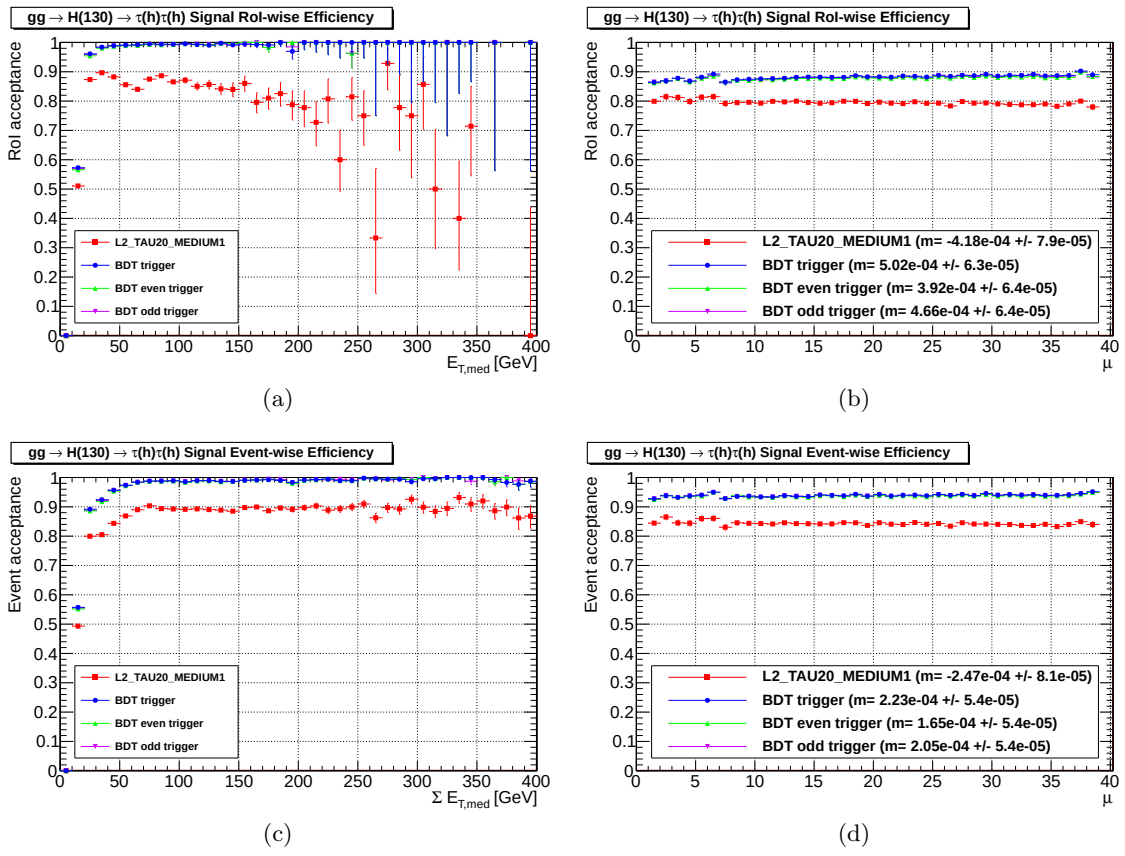


Figure 6.47: The acceptance rate for $gg \rightarrow H(130) \rightarrow \tau(h)\tau(h)$ signal from the ATLAS dataset tagged as `161578.e1217_s1469_s1470_r3542_r3549_p1130`. (a) shows the RoI-wise acceptance rate w.r.t. L2 E_T, (b) shows the RoI-wise acceptance rate w.r.t. μ, (c) shows the event-wise acceptance rate w.r.t. $\sum$L2 E_T, and (d) shows the event-wise acceptance rate w.r.t. μ.

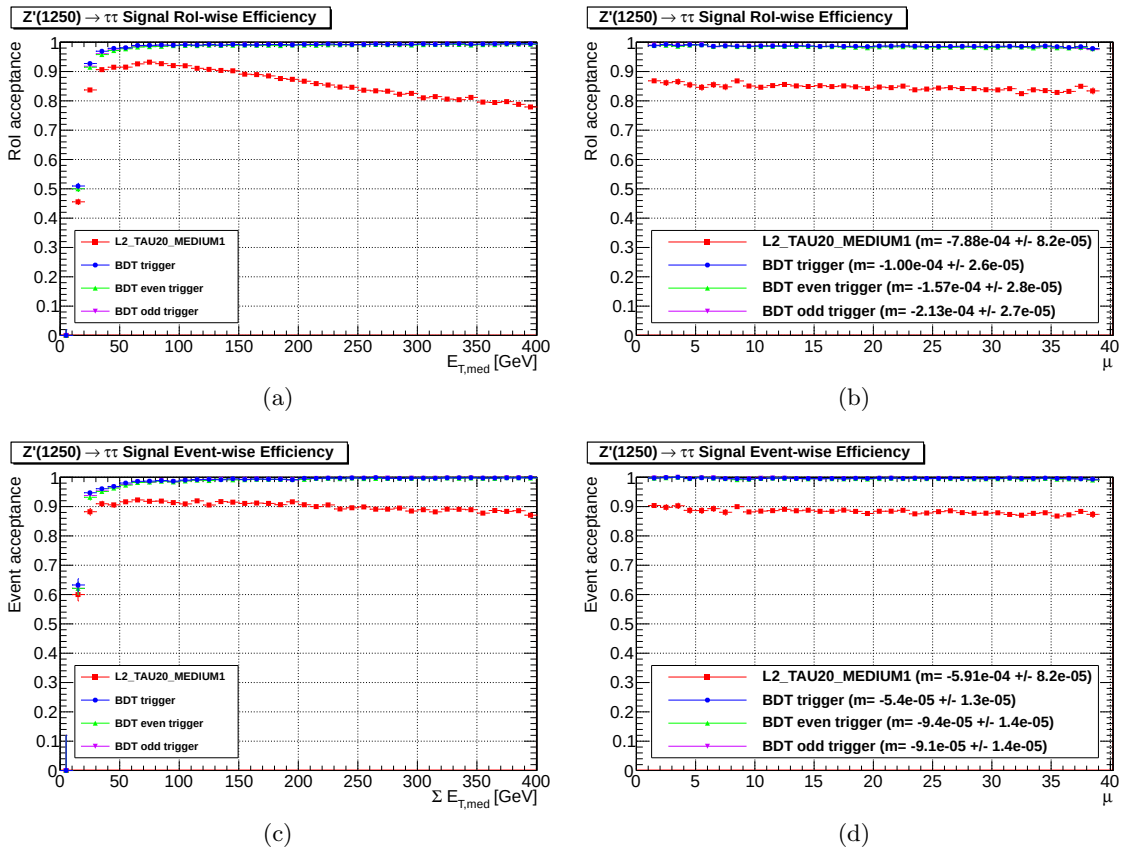


Figure 6.48: The acceptance rate for $Z'(1250) \to \tau\tau$ signal from the ATLAS dataset tagged as `170205.e1176_s1479_s1470_r3553_r3549_p1130`. (a) shows the RoI-wise acceptance rate w.r.t. L2 E_T, (b) shows the RoI-wise acceptance rate w.r.t. μ, (c) shows the event-wise acceptance rate w.r.t. $\sum$L2 E_T, and (d) shows the event-wise acceptance rate w.r.t. μ.

Sample	**RoI-wise**	**Event-wise**
Period E	$(-3.83 \pm 1.04 \pm 1.08) \times 10^{-4}$	$(-4.10 \pm 1.17 \pm 1.19) \times 10^{-4}$
Period E Dijet Bkg	$(-2.7 \pm 1.8 \pm 2.0) \times 10^{-4}$	$(-2.6 \pm 2.3 \pm 2.5) \times 10^{-4}$
$W \to \tau\nu$ Sig 1	$(5.9 \pm 1.8 \pm 0.4) \times 10^{-4}$	$(6.3 \pm 1.8 \pm 0.3) \times 10^{-4}$
$W \to \tau\nu$ Sig 2	$(6.7 \pm 3.7 \pm 0.3) \times 10^{-4}$	$(6.6 \pm 3.7 \pm 0.4) \times 10^{-4}$
$Z \to \tau\tau$ Sig 2	$(8.7 \pm 2.1 \pm 1.2) \times 10^{-4}$	$(7.0 \pm 2.1 \pm 0.9) \times 10^{-4}$
$gg \to H(125)$ Sig	$(-1.28 \pm 0.82 \pm 0.69) \times 10^{-4}$	$(-2.86 \pm 1.01 \pm 0.25) \times 10^{-4}$
$gg \to H(130)$ Sig	$(0.84 \pm 1.01 \pm 0.74) \times 10^{-4}$	$(-0.24 \pm 0.97 \pm 0.40) \times 10^{-4}$
$Z'(1250) \to \tau\tau$ Sig	$(-6.88 \pm 0.86 \pm 0.56) \times 10^{-4}$	$(-5.37 \pm 0.83 \pm 0.03) \times 10^{-4}$
Representative Sig	$(3.7 \pm 3.0 \pm 1.3) \times 10^{-4}$	$(4.4 \pm 3.1 \pm 0.9) \times 10^{-4}$
Period E Sig	$(4.2 \pm 5.0 \pm 0.0) \times 10^{-4}$	$(3.5 \pm 5.1 \pm 0.1) \times 10^{-4}$

Table 6.11: The difference between the absolute value of the slope of the linear fit for the BDT trigger with that for the reference trigger, for each of the 20 cases presented in this section.

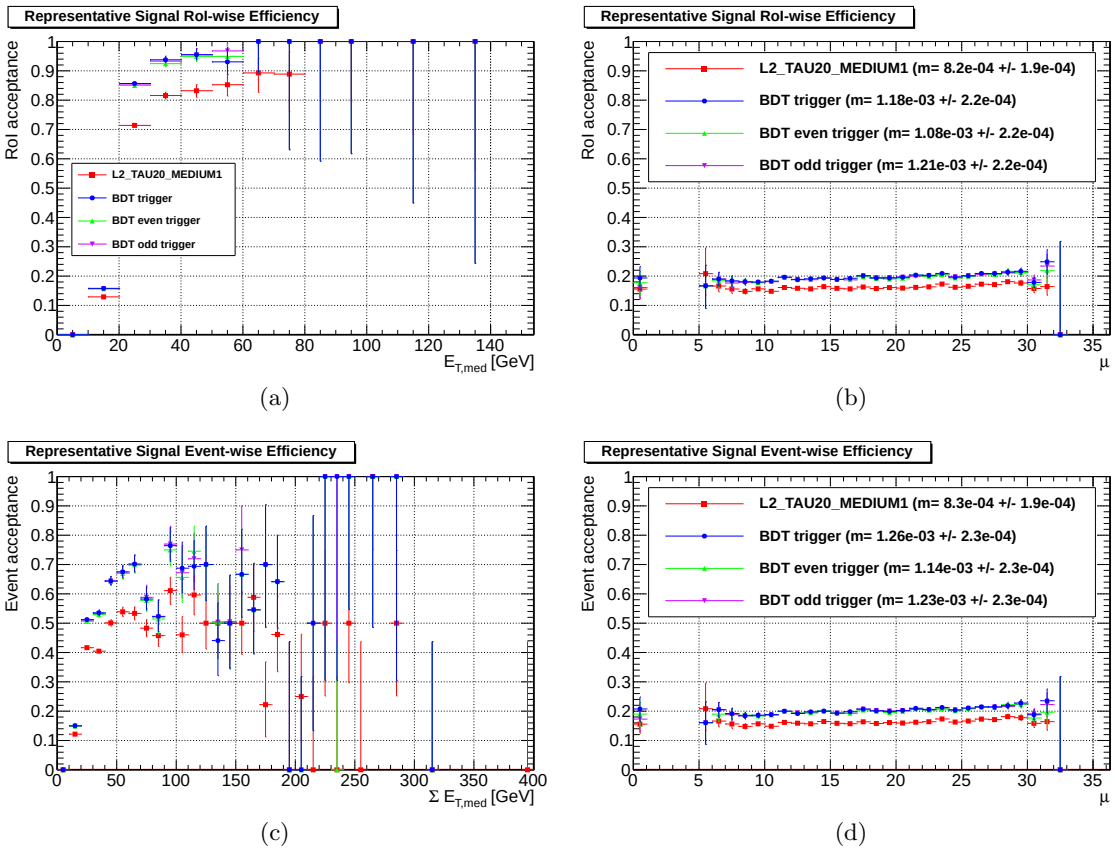


Figure 6.49: The acceptance rate for signal from the representative sample. (a) shows the RoI-wise acceptance rate w.r.t. L2 E_T, (b) shows the RoI-wise acceptance rate w.r.t. μ, (c) shows the event-wise acceptance rate w.r.t. $\sum$L2 E_T, and (d) shows the event-wise acceptance rate w.r.t. μ.

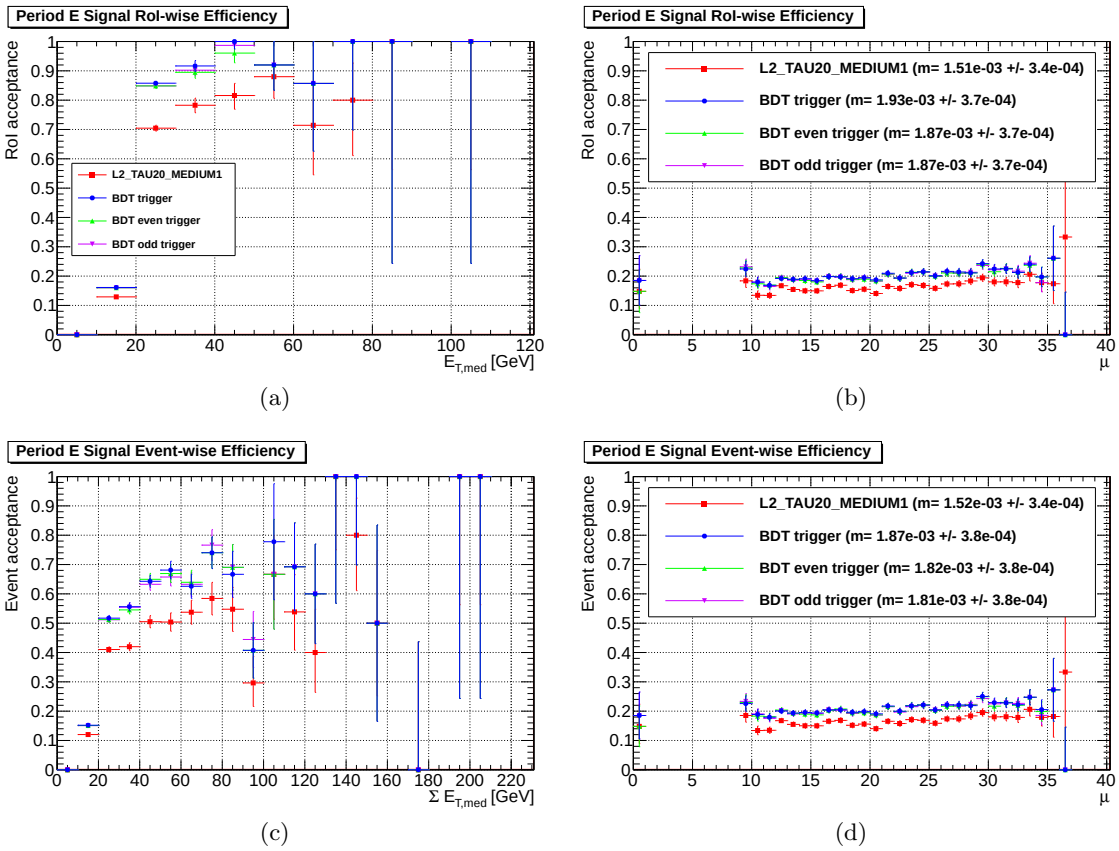


Figure 6.50: The acceptance rate for signal from the data taking period E. (a) shows the RoI-wise acceptance rate w.r.t. L2 E_T, (b) shows the RoI-wise acceptance rate w.r.t. μ, (c) shows the event-wise acceptance rate w.r.t. $\sum$L2 E_T, and (d) shows the event-wise acceptance rate w.r.t. μ.

6.13 Summary of Overall Rates

Here we list the overall acceptance rates (BDT trigger versus reference trigger) for each sample that was shown in the previous sections. Table 6.12 contains the RoI-wise rates and Table 6.13 contains the event-wise rates.

Sample	**L2_TAU20_MEDIUM1**	**BDT Trigger**
Representative	$(7.791 \pm 0.032) \times 10^{-2}$	$(7.47239 \pm 0.03996 \pm 0.00047) \times 10^{-2}$
Dijet Bkg	$(8.947 \pm 0.058) \times 10^{-2}$	$(9.750 \pm 0.079 \pm 0.057) \times 10^{-2}$
Period E	$(7.827 \pm 0.045) \times 10^{-2}$	$(7.564 \pm 0.044 \pm 0.051) \times 10^{-2}$
Period E Dijet Bkg	$(8.986 \pm 0.079) \times 10^{-2}$	$(9.714 \pm 0.082 \pm 0.094) \times 10^{-2}$
$Z \to \tau\tau$ Sig	$(6.8349 \pm 0.0049) \times 10^{-1}$	$(7.6591 \pm 0.0047 \pm 0.0321) \times 10^{-1}$
$Z' \to \tau\tau$ Sig	$(8.5623 \pm 0.0062) \times 10^{-1}$	$(9.8498 \pm 0.0026 \pm 0.0062) \times 10^{-1}$
$W \to \tau\nu$ Sig 1	$(6.367 \pm 0.010) \times 10^{-1}$	$(6.9995 \pm 0.0097 \pm 0.0301) \times 10^{-1}$
$W \to \tau\nu$ Sig 2	$(6.335 \pm 0.021) \times 10^{-1}$	$(6.973 \pm 0.020 \pm 0.043) \times 10^{-1}$
$Z \to \tau\tau$ Sig 2	$(6.809 \pm 0.012) \times 10^{-1}$	$(7.636 \pm 0.011 \pm 0.030) \times 10^{-1}$
$gg \to H(125)$ Sig	$(7.8623 \pm 0.0063) \times 10^{-1}$	$(8.7650 \pm 0.0051 \pm 0.0280) \times 10^{-1}$
$gg \to H(130)$ Sig	$(7.9377 \pm 0.0062) \times 10^{-1}$	$(8.8418 \pm 0.0049 \pm 0.0246) \times 10^{-1}$
$Z'(1250) \to \tau\tau$ Sig	$(8.4433 \pm 0.0064) \times 10^{-1}$	$(9.8622 \pm 0.0021 \pm 0.0035) \times 10^{-1}$
Representative Sig	$(1.621 \pm 0.015) \times 10^{-1}$	$(1.9634 \pm 0.0197 \pm 0.0064) \times 10^{-1}$
Period E Sig	$(1.623 \pm 0.020) \times 10^{-1}$	$(2.00491 \pm 0.02193 \pm 0.00060) \times 10^{-1}$

Table 6.12: The RoI-wise acceptance rates for all the tested samples. The Representative, Dijet Background, $Z \to \tau\tau$ Signal, $Z' \to \tau\tau$ Signal samples were used for training the BDT trigger. *Bkg* stands for Background, and *Sig* stands for Signal.

Sample	**L2_TAU20_MEDIUM1**	**BDT Trigger**
Representative	$(8.673 \pm 0.035) \times 10^{-2}$	$(8.3661 \pm 0.0442 \pm 0.0017) \times 10^{-2}$
Dijet Bkg	$(1.0770 \pm 0.0070) \times 10^{-1}$	$(1.2100 \pm 0.0097 \pm 0.0077) \times 10^{-1}$
Period E	$(8.709 \pm 0.050) \times 10^{-2}$	$(8.471 \pm 0.049 \pm 0.055) \times 10^{-2}$
Period E Dijet Bkg	$(1.0799 \pm 0.0095) \times 10^{-1}$	$(1.2042 \pm 0.0100 \pm 0.0121) \times 10^{-1}$
$Z \to \tau\tau$ Sig	$(7.0698 \pm 0.0050) \times 10^{-1}$	$(8.1326 \pm 0.0047 \pm 0.0276) \times 10^{-1}$
$Z' \to \tau\tau$ Sig	$(8.9065 \pm 0.0062) \times 10^{-1}$	$(9.9523 \pm 0.0017 \pm 0.0026) \times 10^{-1}$
$W \to \tau\nu$ Sig 1	$(6.369 \pm 0.010) \times 10^{-1}$	$(7.0550 \pm 0.0096 \pm 0.0317) \times 10^{-1}$
$W \to \tau\nu$ Sig 2	$(6.337 \pm 0.021) \times 10^{-1}$	$(7.026 \pm 0.020 \pm 0.046) \times 10^{-1}$
$Z \to \tau\tau$ Sig 2	$(7.042 \pm 0.012) \times 10^{-1}$	$(8.113 \pm 0.011 \pm 0.026) \times 10^{-1}$
$gg \to H(125)$ Sig	$(8.3380 \pm 0.0065) \times 10^{-1}$	$(9.3428 \pm 0.0043 \pm 0.0197) \times 10^{-1}$
$gg \to H(130)$ Sig	$(8.4181 \pm 0.0063) \times 10^{-1}$	$(9.3947 \pm 0.0041 \pm 0.0146) \times 10^{-1}$
$Z'(1250) \to \tau\tau$ Sig	$(8.8165 \pm 0.0064) \times 10^{-1}$	$(9.9619 \pm 0.0012 \pm 0.0015) \times 10^{-1}$
Representative Sig	$(1.626 \pm 0.016) \times 10^{-1}$	$(2.0193 \pm 0.0203 \pm 0.0053) \times 10^{-1}$
Period E Sig	$(1.627 \pm 0.020) \times 10^{-1}$	$(2.0665 \pm 0.0222 \pm 0.0021) \times 10^{-1}$

Table 6.13: The Event-wise acceptance rates for all the tested samples. The Representative, Dijet Background, $Z \to \tau\tau$ Signal, $Z' \to \tau\tau$ Signal samples were used for training the BDT trigger. *Bkg* stands for Background, and *Sig* stands for Signal.

Chapter 7

Conclusion

Studying ATLAS events with taus will contribute greatly to both the understanding of SM processes as well as the discovery potential of the ATLAS detector. In particular, both the SM and supersymmetric Higgs bosons are expected to have large branching ratios for tau decays. [Section 4.1] The ATLAS tau trigger has performed remarkably well in 2011 [Section 4.4] and a performance report on 2012 is likewise expected to be positive [Section 4.6]. Increasing luminosities required adjustments to account for the more challenging pile-up conditions. Multivariate techniques for tau selection, first used at the offline level, were largely successful at discriminating signal from background. This led to the adoption of multivariate techniques at the EF, namely BDTs and LLH. We presented a BDT-based single tau selection algorithm for use at the L2 tau trigger. [Chapter 5]

The results indicate that our algorithm outperforms the current algorithm on a wide variety of measures, for both the event-wise and the RoI-wise cases. With 3σ certainty, the calorimeter-stage acceptance rate of our trigger (the BDT trigger) for the representative sample (2012 data taking periods B-D) is less than that of the currently implemented trigger (the reference trigger). More precisely, it is $29.369\% \pm 0.057\%$ for the reference trigger, and $28.8136\% \pm 0.0603\% \pm 0.0064\%$ for the BDT trigger. The pile-up dependence (slope) of the calorimeter stage of the BDT trigger for the representative sample, is equivalent to that of the reference trigger, within 2σ deviation. Furthermore, both the RoI-wise and event-wise

calorimeter-stage slopes for the reference trigger and the BDT trigger are less than 3.32 % efficiency per 50 interactions per bunch crossing, with 3σ certainty. [Section 6.8] For the overall algorithm (calorimeter stage followed by tracking stage), the reference trigger has an RoI-wise acceptance rate of $(7.791 \pm 0.032)\%$ for the representative sample and $(68.349 \pm 0.049)\%$ for $Z \to \tau\tau$ signal. In comparison, the BDT trigger has an RoI-wise acceptance rate of $(7.47239 \pm 0.03996 \pm 0.00047)\%$ for the representative sample and $(76.591 \pm 0.047 \pm 0.321)\%$ for $Z \to \tau\tau$ signal. Furthermore, the reference trigger has an event-wise acceptance rate of $(8.673 \pm 0.035)\%$ for the representative sample and $(70.698 \pm 0.050)\%$ for $Z \to \tau\tau$ signal, while the BDT trigger has an event-wise acceptance rate of $(8.3661 \pm 0.0442 \pm 0.0017)\%$ for the representative sample and $(81.326 \pm 0.047 \pm 0.276)\%$ for $Z \to \tau\tau$ signal. Using standard error propagation, this means that the BDT trigger adds $(0.319 \pm 0.051 \pm 0.000)\%$ to the representative RoI-wise rejection rate of the reference trigger and adds $(8.242 \pm 0.068 \pm 0.321)\%$ to the $Z \to \tau\tau$ signal RoI-wise acceptance rate of the reference trigger. Futhermore, the BDT trigger adds $(0.307 \pm 0.056 \pm 0.002)\%$ to the representative event-wise rejection rate of the reference trigger and adds $(10.628 \pm 0.069 \pm 0.276)\%$ to the $Z \to \tau\tau$ signal event-wise acceptance rate of the reference trigger. By design, the additions to the rejection rates are nearly zero, and are greater than 0 with 3σ certainty. The addition to the RoI-wise $Z \to \tau\tau$ signal acceptance rate is greater than 7.075% with 3σ certainty, and the addition to the event-wise $Z \to \tau\tau$ signal acceptance rate is greater than 9.593% with 3σ certainty. The BDT trigger has also been shown to outperform the reference trigger for each of the principal error code classes that each correspond to a separately trained BDT classifier. [Section 6.9] As for energy dependence, the BDT trigger generally has a greater or equal rejection rate for the representative sample than that for the reference trigger, for each level of energy, with the difference widening with increasing levels of energy. The BDT trigger also has a greater acceptance rate for $Z \to \tau\tau$ signal than that for the reference trigger, for each level of energy, with the difference also widening with increasing levels of energy. In particular, the reference trigger's energy dependent RoI-wise acceptance rate for all the tested signal samples slopes down steeply for high levels of energy. This effect is not

so significant for SM signals such as $Z \to \tau\tau$, but becomes more important when searching for more massive particles such as the supersymmetric Higgs bosons. The BDT trigger's rate continues to slope upwards for the highest levels of energy, and is therefore much more effective at contributing towards the discovery of these types of high mass resonances that will likely be searched for after the focus shifts away from the SM Higgs boson. [Section 6.10] As for pile-up dependence, the BDT trigger generally has a greater or equal rejection rate for the representative sample than that for the reference trigger, for each value of μ. The BDT trigger also has a greater acceptance rate for $Z \to \tau\tau$ signal than that for the reference trigger, for each value of μ. In terms of the amount of pile-up dependence, the (absolute) slope of the BDT trigger for the representative sample is less than that for the reference trigger, with 2σ certainty. For $Z' \to \tau\tau$ signal (high mass resonance) the slope of the BDT trigger is less than that for the reference reference trigger, with 6σ certainty for the RoI-wise case and 7σ certainty for the event-wise case. However, for $Z \to \tau\tau$ signal, the BDT trigger's slope is greater than the reference trigger's slope, with 4σ certainty for the RoI-wise case and 2σ certainty for the event-wise case. Depending on the specific requirements for the tau trigger, some variables may need to be removed from the BDT training program in order to reduce this dependence. However, the underperformance in pile-up dependence is still within reasonable limits, and the dependence actually slopes upwards, which means that the efficiencies are not expected to drop with the expected higher luminosities in the future. [Section 6.11] In order to demonstrate the robustness of our algorithm with respect to changing run conditions, we measured both the energy and pile-up dependent rates for the 2012 data taking period E (overall and dijet background). We also measured the robustness of our algorithm with respect to different channels of tau production: $W \to \tau\nu$, $H(125) \to \tau\tau$, $H(130) \to \tau\tau$, and $Z'(1250) \to \tau\tau$. Finally, we measured the performance of our algorithm on signal taken from real data. In all cases, our algorithm generally outperforms the current algorithm with respect to both energy dependence and pile-up dependence. [Section 6.12]

The main systematic uncertainty included in the study was the dependence of our train-

ing procedure on the training sample. We split our training sample into two independent (even and odd) samples, and obtained the results for training on both cases: just even and just odd. A more rigourous study should include more sources of systematic uncertainty for the training procedure. Some suggestions include: the dependence on the cross sections for weighting the training samples, the dependence on the physics channels used for the training samples, the dependence on the parameters for the MC simulation of a given physics channel. Furthermore, one should also study the systematic uncertainty of the *testing* procedure. In a sense, the dependence with respect to changing run conditions and different signal channels measures this. However, a more detailed and quantitative analysis of these effects should be implemented in the future.

We believe that we have shown strong evidence for implementing our algorithm to the tau trigger. Not only does the BDT trigger outperform the current trigger for SM signal channels, but it also dominates the current trigger for yet undiscovered high mass resonances (e.g. supersymmetric Higgs bosons) that will be more accessible in the upcoming (higher energy) running period. We also believe that a more detailed study, including peer review, is needed before the implementation can occur. It is possible that if the current cut-based trigger is reoptimized for the samples we used for training, then its performance may be much more comparable to ours. It is also possible that while our trigger results in a more efficient L2 stage, the overall L2+EF stage of the trigger is worse with our trigger at L2; hence, a study of the effect our trigger has on the EF is necessary. Moreover, the possibility of a one-step L2 trigger in the future demands a performance analysis of our algorithm when there is no restriction to a calorimeter stage. It is expected that such a study will be completed before the next running period, and if implemented, we hope that our algorithm can significantly contribute towards the discovery of new physics.

Appendix A

Quantum Field Theory

A.1 State Space

In quantum mechanics, the state space for a particle is the projective Hilbert space $\mathbb{P}\mathcal{H}$ of a separable (complex) Hilbert space $\mathcal{H}$, commonly denoted as simply $\mathcal{H}$. The state space for a system of k particles with state spaces $\mathcal{H}_1, \ldots, \mathcal{H}_k$, is a subspace of (the metric space completion of) the algebraic tensor product $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$. If $\{e_n^j : n \in \mathbb{Z}_+\}$ is an orthonormal basis for $\mathcal{H}_j$, then $\{e_{n_1}^1 \otimes \cdots \otimes e_{n_k}^k : n_1, \ldots, n_k \in \mathbb{Z}_+\}$ is an orthonormal basis for $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$. If S_k is the permutation group of k letters, $\sigma \in S_k$, and $u_1 \otimes \cdots \otimes u_k \in \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$, we define

$$\sigma(u_1 \otimes \cdots \otimes u_k) = u_{\sigma(1)} \otimes \cdots \otimes u_{\sigma(k)} \,. \tag{A.1}$$

If the particles are each bosons of the same species with state space $\mathcal{H}$, then the subspace is

$$\textcircled{S}^k\mathcal{H} \coloneqq \{\boldsymbol{v} \in \otimes^k\mathcal{H} : \sigma(\boldsymbol{v}) = \boldsymbol{v} \; \forall \sigma \in S_k\} \,. \tag{A.2}$$

For the case of identical fermions, the subspace is

$$\wedge^k \mathcal{H} \coloneqq \{\boldsymbol{v} \in \otimes^k\mathcal{H} : \sigma(\boldsymbol{v}) = \operatorname{sgn}(\sigma)\boldsymbol{v} \; \forall \sigma \in S_k\} \,. \tag{A.3}$$

Let $P_s : \otimes^k\mathcal{H} \to \textcircled{S}^k\mathcal{H}$ such that

$$P_s(u_1 \otimes \cdots \otimes u_k) = \frac{1}{k!} \sum_{\sigma \in S_k} u_{\sigma(1)} \otimes \cdots \otimes u_{\sigma(k)} \,. \tag{A.4}$$

For any sequence $\{n_j\} \in \mathbb{N}$ with $\sum_{j=1}^{\infty} n_j = k$, let

$$|n_1, n_2, \ldots\rangle = \sqrt{\frac{n_1! n_2! \cdots}{k!}} P_s((\otimes^{n_1} e_1) \otimes (\otimes^{n_2} e_2) \otimes \cdots), \tag{A.5}$$

where $\{e_j\}_{j \in \mathbb{Z}_+}$ is an orthonormal basis for $\mathcal{H}$. It can be shown that

$$\{|n_1, n_2, \ldots\rangle : \sum n_j = k\} \tag{A.6}$$

is an orthonormal basis for $\textcircled{S}^k\mathcal{H}$ [7]. For fermions, let $P_a : \otimes^k\mathcal{H} \to \wedge^k\mathcal{H}$ such that

$$P_a(u_1 \otimes \cdots \otimes u_k) = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) u_{\sigma(1)} \otimes \cdots \otimes u_{\sigma(k)}. \tag{A.7}$$

For any sequence $i_1, \ldots, i_k \in \mathbb{Z}_+$ with $i_1 < \cdots < i_k$, let

$$e_{i_1} \wedge \cdots \wedge e_{i_k} = P_a(e_{i_1} \otimes \cdots \otimes e_{i_k}), \tag{A.8}$$

where $\{e_j\}_{j \in \mathbb{Z}_+}$ is an orthonormal basis for $\mathcal{H}$. It can be shown that

$$\{e_{i_1} \wedge \cdots \wedge e_{i_k} : i_1, \ldots, i_k \in \mathbb{Z}_+ \text{ and } i_1 < \cdots < i_k\} \tag{A.9}$$

is an orthogonal basis for $\wedge^k\mathcal{H}$ [7].

In order to form a state space that can accomodate any number of identical bosons, we combine all the spaces $\textcircled{S}^k\mathcal{H}$ to form the *Boson Fock Space* over $\mathcal{H}$, defined as the orthogonal direct sum

$$\mathcal{F}_s(\mathcal{H}) = \bigoplus_{k=0}^{\infty} \textcircled{S}^k\mathcal{H}, \tag{A.10}$$

where $\textcircled{S}^0\mathcal{H} \coloneqq \mathbb{C}$. The set

$$\{|n_1, n_2, \ldots\rangle : n_j \in \mathbb{N} \text{ and } \sum n_j < \infty\}, \tag{A.11}$$

with $|0, 0, \ldots\rangle \coloneqq 1 \in \textcircled{S}^0\mathcal{H}$, is an orthonormal basis for $\mathcal{F}_s(\mathcal{H})$. For any sequence $\{n_j\} \in \mathbb{N}$ with $\sum n_j < \infty$, the basis vector $|n_1, n_2, \ldots\rangle$ is the state for the set of all bosons of a given species, where, for each $j \in \mathbb{Z}_+$, there are n_j particles in the eigenstate e_j. The annihilation and creation operators on $\mathcal{F}_s(\mathcal{H})$, A_j and $A_j^\dagger$, respectively, are defined as:

$$A_j |n_1, \ldots, n_j, \ldots\rangle = \sqrt{n_j}\, |n_1, \ldots, n_j - 1, \ldots\rangle, \tag{A.12}$$

$$A_j^\dagger |n_1, \ldots, n_j, \ldots\rangle = \sqrt{n_j + 1}\, |n_1, \ldots, n_j + 1, \ldots\rangle. \tag{A.13}$$

Similarly, for fermions, the *Fermion Fock Space* is defined as

$$\mathcal{F}_a(\mathcal{H}) = \bigoplus_{k=0}^{\infty} \wedge^k \mathcal{H}\,, \tag{A.14}$$

where $\wedge^0\mathcal{H} = \mathbb{C}$. For any sequence $\{n_j\} \in \{0,1\}$ and $i_1 < \cdots < i_k$ the indices i for which $n_i = 1$, let

$$|n_1, n_2, \ldots\rangle : \sqrt{k!}e_{i_1} \wedge \cdots \wedge e_{i_k}\,. \tag{A.15}$$

Then it follows that

$$\{|n_1, n_2, \ldots\rangle : n_j \in \{0,1\} \text{ and } \sum n_j < \infty\}\,, \tag{A.16}$$

with $|0, 0, \ldots\rangle \coloneqq 1 \in \wedge^0\mathcal{H}$, is an orthonormal basis for $\mathcal{F}_a(\mathcal{H})$ [7]. For any sequence $\{n_j\} \in \{0,1\}$ with $\sum n_j < \infty$, the basis vector $|n_1, n_2, \ldots\rangle$ is the state for the set of all fermions of a given species, where, for each $j \in \mathbb{Z}_+$, there are n_j particles in the eigenstate e_j. The restriction to $\{0,1\}$ is a precise statement of the Pauli Exclusion Principle. The annihilation and creation operators on $\mathcal{F}_a(\mathcal{H})$, A_j and $A_j^\dagger$, respectively, are defined as:

$$A_j\,|n_1, \ldots, n_j, \ldots\rangle = \begin{cases} (-1)^m\sqrt{n_j}\,|n_1, \ldots, n_j - 1, \ldots\rangle \text{ if } n_j = 1 \\ 0 \text{ if } n_j = 0 \end{cases}\,, \tag{A.17}$$

$$A_j^\dagger\,|n_1, \ldots, n_j, \ldots\rangle = \begin{cases} (-1)^m\sqrt{n_j}\,|n_1, \ldots, n_j + 1, \ldots\rangle \text{ if } n_j = 0 \\ 0 \text{ if } n_j = 1 \end{cases}\,, \tag{A.18}$$

where $n_i = 1$ for m values of i with $i < j$.

More generally, we want a state space that can accomodate any number of particles, where each particle is a member of one of k different species. If species i has single particle state space $\mathcal{H}_i$, let $\mathcal{F}_i = \mathcal{F}_s(\mathcal{H}_i)$ if the species is bosonic and let $\mathcal{F}_i = \mathcal{F}_a(\mathcal{H}_i)$ otherwise. Then the state space we want is the *Fock space*

$$\mathcal{F} \coloneqq \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k\,. \tag{A.19}$$

Now, let $\mathcal{N}_i = \mathbb{N}$ if species i is bosonic, and let $\mathcal{N}_i = \{0,1\}$ otherwise. If $\{|n_1^i, n_2^i, \ldots\rangle : n_j^i \in$

$\mathcal{N}_i, \sum_{j=1}^{\infty} n_j^i < \infty\}$ is an orthonormal basis for $\mathcal{F}_i$, then

$$\{|n_1^1, n_2^1, \ldots\rangle \otimes \cdots \otimes |n_1^k, n_2^k, \ldots\rangle : n_j^1 \in \mathcal{N}_1, \ldots, n_j^k \in \mathcal{N}_k, \text{ and } \sum_{j=1}^{\infty} n_j^1, \ldots, \sum_{j=1}^{\infty} n_j^k < \infty\} \tag{A.20}$$

is an orthonormal basis for $\mathcal{F}$.

A.2 Harmonic Oscillator Correspondence

The Hamiltonian for a system of K one-dimensional quantum harmonic oscillators of equal mass is

$$H = \sum_{j=1}^{K} \frac{1}{2m} P_j^2 + \frac{1}{2} m\omega_j^2 X_j^2 \,, \tag{A.21}$$

where m is the mass of each oscillator, and, P_j, X_j, and w_j is the jth oscillator's momentum operator, position operator, and angular frequency, respectively. This is just the sum over the K Hamiltonians $\{H_j : j \in \mathbb{Z}_{+K}\}$ for each one-dimensional oscillator indexed by j. For each $j \in \mathbb{Z}_{+K}$, if we let

$$A_j = \sqrt{\frac{m\omega}{2}} \left(X_j + \frac{i}{m\omega} P_j \right) \tag{A.22}$$

and

$$A_j^\dagger = \sqrt{\frac{m\omega}{2}} \left(X_j - \frac{i}{m\omega} P_j \right) , \tag{A.23}$$

then the eigenfunctions of H_j are given by:

$$\psi_0^j(x_j) = \left(\frac{m\omega_j}{\pi} \right)^{1/4} e^{-\frac{m}{2}\omega_j x_j^2} \,, \tag{A.24}$$

$$\psi_n^j(x_j) = \frac{(A_j^\dagger)^n}{\sqrt{n!}} \psi_0(x_j) \,, \tag{A.25}$$

and the eigenfunctions of H are simply given by the products:

$$\psi_{00\ldots0}(\boldsymbol{x}) = \left(\frac{m}{\pi} \right)^{K/4} \prod_{j=1}^{K} \omega_j^{1/4} e^{-\frac{m}{2}\sum_{j=1}^{K} \omega_j x_j^2} \,, \tag{A.26}$$

$$\psi_{n_1 \ldots n_K}(\boldsymbol{x}) = \frac{(A_1^\dagger)^{n_1} \ldots (A_K^\dagger)^{n_K}}{\sqrt{n_1! \ldots n_K!}} \psi_{00\ldots0}(\boldsymbol{x}) \,. \tag{A.27}$$

Now, for each $j \in \mathbb{Z}_{+K}$,

$$A_j \psi_{n_1 \ldots n_K} = \sqrt{n_j}\psi_{n_1 \ldots n_{j-1} \ldots n_K} \text{ for } n \geq 1\,, \tag{A.28}$$

$$A_j^\dagger \psi_{n_1 \ldots n_K} = \sqrt{n_j + 1}\psi_{n_1 \ldots n_{j+1} \ldots n_K} \text{ for } n \geq 0\,, \tag{A.29}$$

and clearly, the correspondence

$$\psi_{n_1 \ldots n_K} \mapsto |n_1, \ldots, n_K\rangle \tag{A.30}$$

defines a unitary map from $L^2(\mathbb{R}^K)$ to $\mathcal{F}_s(\mathbb{C}^K)$, with each $A_j, A_j^\dagger$ corresponding to the annihilation and creation operations on $\mathcal{F}_s(\mathbb{C}^K)$. In the limit that $K \to \infty$, $\mathcal{F}_s(\mathbb{C}^K)$ becomes $\mathcal{F}_s(\mathcal{H})$ for a separable infinite-dimensional Hilbert space $\mathcal{H}$ and $A_j, A_j^\dagger$ become the annihilation and creation operations on $\mathcal{F}_s(\mathcal{H})$. By (A.22) and (A.23), the position operator for each oscillator can be expressed in terms of A_j and $A_j^\dagger$ as follows:

$$X_j = \frac{1}{\sqrt{2m\omega_j}} \left(A_j + A_j^\dagger \right) . \tag{A.31}$$

By scaling our units for length by $\frac{1}{\sqrt{m}}$, we can simply write this as

$$X_j = \frac{1}{\sqrt{2\omega_j}} \left(A_j + A_j^\dagger \right) . \tag{A.32}$$

Moreover, the Hamiltonian for each oscillator is

$$H_j = \omega_j \left(A_j^\dagger A_j + \frac{1}{2} \right) . \tag{A.33}$$

But, $\sum_{j=1}^{\infty} \omega_j$ may be divergent, so to fix this, we subtract $\omega_j/2$ from each H_j. This can be done since adding a constant to the energy of a system does not affect its dynamics. We thus obtain the renormalized Hamiltonians, given by

$$H_j = \omega_j A_j^\dagger A_j \tag{A.34}$$

for each $j \in \mathbb{Z}_{+K}$.

A.3 Free Neutral Scalar Field

We will now construct the free quantum neutral scalar field for particles of mass $m > 0$. The relevant free field equation is the Klein-Gordon Equation:

$$\left(\partial_t^2 - \boldsymbol{\nabla}^2 + m^2\right)\phi(t, \boldsymbol{x}) = 0\,, \tag{A.35}$$

where ϕ is a classical real-valued field, $\boldsymbol{x}$ is initially restricted to the cube $\mathcal{B} \coloneqq [-\frac{1}{2}L, \frac{1}{2}L]^3 \subset \mathbb{R}^3$, $t \in \mathbb{R}$, and the equation is initially subject to real boundary conditions that make $-\boldsymbol{\nabla}^2$ positive and self-adjoint. The boundedness of $\mathcal{B}$ guarantees that $-\boldsymbol{\nabla}^2$ has only a discrete spectrum (as an operator on $L^2(\mathcal{B})$), so that there is an orthonormal eigenbasis of *real*-valued functions $\{f_j\}_{j=1}^{\infty}$ for $-\boldsymbol{\nabla}^2$ with real eigenvalues $\{\lambda_j^2\}_{j=1}^{\infty}$. Thus, we can expand ϕ as

$$\phi(t, \boldsymbol{x}) = \sum_{j=1}^{\infty} q_j(t) f_j(\boldsymbol{x})\,, \tag{A.36}$$

where each $q_j(t)$ is a real-valued function of t. By applying the Klein-Gordon equation to the expansion, we see that

$$q_j'' + \omega_j^2 q_j = 0 \tag{A.37}$$

for each $j \in \mathbb{Z}_+$, where $\omega_j^2 = \lambda_j^2 + m^2$. But this is the equation for a classical harmonic oscillator with displacement q_j and angular frequency ω_j. We turn the classical field into a quantum field by replacing each classical displacement q_j with the displacement operator for the quantum harmonic oscillator with angular frequency ω_j. By (A.32), $q_j(t)$ is replaced with $\frac{1}{\sqrt{2\omega_j}}(A_j(t) + A_j^\dagger(t))$, where (viewed in the Heisenberg picture) $A_j(t) = e^{itH} A_j e^{-itH}$ and $A_j^\dagger(t) = e^{itH} A_j^\dagger e^{-itH}$, for $H = \sum_{j\in\Lambda} \omega_j A_j^\dagger A_j$ (by (A.34)). It can be shown that $A_j(t) = e^{-i\omega_j t} A_j$ and $A_j^\dagger(t) = e^{i\omega_j t} A_j^\dagger$ [7]. We regard each A_j and $A_j^\dagger$ as operators on $\mathcal{F}_s(\mathcal{H})$, for a separable infinite-dimensional Hilbert space $\mathcal{H}$, as per the correspondence shown in Section A.2. The field is now

$$\phi(x) = \sum_{j=1}^{\infty} \frac{1}{\sqrt{2\omega_j}} f_j(\boldsymbol{x}) \left(A_j(t) + A_j^\dagger(t)\right)\,. \tag{A.38}$$

In this form, for each point x in spacetime, the value of the field at x is an operator on $\mathcal{F}_s(\mathcal{H})$. But when we apply this operator to some vector in $\mathcal{F}_s(\mathcal{H})$, the resulting vector may

have a divergent norm, so to fix this we instead regard ϕ as an operator-valued *distribution* that maps each real-valued Schwartz class function f on $\mathbb{R}$ and g on $\mathcal{B}$ to the operator

$$\int_{\mathbb{R}}\int_{\mathcal{B}} f(t)g(\boldsymbol{x})\phi(t,\boldsymbol{x})\mathrm{d}^3\boldsymbol{x}\mathrm{d}t\,. \tag{A.39}$$

Now, let us simply require $-\boldsymbol{\nabla}^2$ to be Hermitian with an orthonormal eigenbasis of *complex*-valued functions $\{g_j\}_{j=1}^{\infty}$. As shown in [7], if we replace the real-valued functions $\{f_j\}_{j=1}^{\infty}$ in (A.38) with appropriately chosen expansions of $\{g_j\}_{j=1}^{\infty}$, we get

$$\phi(x) = \sum_{j=1}^{\infty} \frac{1}{\sqrt{2\omega_j}} \left(g_j(\boldsymbol{x})A_j(t) + g_j(\boldsymbol{x})^* A_j^\dagger(t) \right) , \tag{A.40}$$

where each A_j and $A_j^\dagger$ is now defined in terms of a new basis of $\mathcal{H}$. In particular, for $\Lambda \coloneqq \left[\frac{2\pi}{L}\mathbb{Z}\right]^3 \subset \mathbb{R}^3$, let $\{g_j\}_{j=1}^{\infty} = \{g_{\boldsymbol{p}} : g_{\boldsymbol{p}}(\boldsymbol{x}) = L^{-3/2}e^{i\boldsymbol{p}\boldsymbol{x}}, \boldsymbol{p} \in \Lambda\}$, so that the eigenvalues are $\{\lambda_{\boldsymbol{p}}^2 : \lambda_{\boldsymbol{p}}^2 = \boldsymbol{p}^2, \boldsymbol{p} \in \Lambda\}$, and

$$\phi(x) = \sum_{\boldsymbol{p}\in\Lambda} \frac{1}{\sqrt{2\omega_{\boldsymbol{p}}L^3}} \left(e^{i\boldsymbol{p}\boldsymbol{x}}A_{\boldsymbol{p}}(t) + e^{-i\boldsymbol{p}\boldsymbol{x}}A_{\boldsymbol{p}}^\dagger(t) \right) \tag{A.41}$$

$$= \sum_{\boldsymbol{p}\in\Lambda} \frac{1}{\sqrt{2\omega_{\boldsymbol{p}}L^3}} \left(e^{i\boldsymbol{p}\boldsymbol{x}}e^{-i\omega_{\boldsymbol{p}}t}A_{\boldsymbol{p}} + e^{-i\boldsymbol{p}\boldsymbol{x}}e^{i\omega_{\boldsymbol{p}}t}A_{\boldsymbol{p}}^\dagger \right) \tag{A.42}$$

$$= \sum_{\boldsymbol{p}\in\Lambda} \frac{1}{\sqrt{2\omega_{\boldsymbol{p}}L^3}} \left(e^{ipx}A_{\boldsymbol{p}} + e^{-ipx}A_{\boldsymbol{p}}^\dagger \right) . \tag{A.43}$$

Interestingly, Lorentz invariance is a natural consequence of the theory. Now, let $\Delta V = (2\pi/L)^3$. Then

$$\phi(x) = \sum_{\boldsymbol{p}\in\Lambda} \frac{1}{\sqrt{2\omega_{\boldsymbol{p}}}} L^{3/2} \left(e^{ipx}A_{\boldsymbol{p}} + e^{-ipx}A_{\boldsymbol{p}}^\dagger \right) \frac{\Delta V}{(2\pi)^3}\,. \tag{A.44}$$

As $L \to \infty$, $\Delta V \to 0$, and the sum becomes the integral

$$\phi(x) = \int_{\mathbb{R}^3} \frac{1}{\sqrt{2\omega_{\boldsymbol{p}}}} \left(e^{ipx}a(\boldsymbol{p}) + e^{-ipx}a(\boldsymbol{p})^\dagger \right) \frac{\mathrm{d}^3\boldsymbol{p}}{(2\pi)^3}\,, \tag{A.45}$$

where $a(\boldsymbol{p})$ is the (natural) distribution corresponding to $L^{3/2}A_{\boldsymbol{p}}$ and $a(\boldsymbol{p})^\dagger$ is the distribution corresponding to $L^{3/2}A_{\boldsymbol{p}}^\dagger$.

The generalization to charged scalar fields, spinor fields, and vector fields is straightforward and follows the same general principles. [7]

A.4 The Scattering Operator

Definition 1. For a system with Hamiltonian H, the scattering operator is given by

$$S = \lim_{\substack{t_0 \to -\infty \\ t_1 \to +\infty}} e^{iH_0 t_1} e^{-iH(t_1 - t_0)} e^{-iH_0 t_0} \,.[7] \tag{A.46}$$

Theorem 1. *For a system with Hamiltonian H and scattering operator S,*

$$S = I + \sum_{n=1}^{\infty} \frac{1}{i^n n!} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathcal{T}[H_I(\tau_1) \dots H_I(\tau_n)] d\tau_1 \dots d\tau_n \,, \tag{A.47}$$

where

$$\mathcal{T}[H_I(\tau_1) \dots H_I(\tau_n)] \coloneqq H_I(\tau_{i_1}) \dots H_I(\tau_{i_n}) \text{ for } \tau_{i_1} > \tau_{i_2} > \dots > \tau_{i_n} \,. \tag{A.48}$$

Proof. Let $U(t) = e^{-itH}$, $U_0(t) = e^{-itH_0}$, $V(t) = U_0(-t)U(t)$, $H_I(t) = U_0(-t)H_I U_0(t)$. Then,

$$\begin{aligned} V'(t) &= iU_0(-t)H_0 U(t) - iU_0(-t)HU(t) \\ &= \frac{1}{i} U_0(-t) H_I U(t) \\ &= \frac{1}{i} H_I(t) V(t) \,. \end{aligned}$$

Now, let $V(t, t_0) = U_0(-t)U(t - t_0)U_0(t_0) = V(t)V(t_0)^{-1}$. Then,

$$\begin{aligned} \frac{d}{dt} V(t, t_0) &= \frac{1}{i} H_I(t) V(t) V(t_0)^{-1} \\ &= \frac{1}{i} H_I(t) V(t, t_0) \,. \end{aligned}$$

The above differential equation is equivalent to the integral equation

$$\begin{aligned} V(t, t_0) &= V(t_0, t_0) + \int_{t_0}^{t} \frac{d}{d\tau_1} V(\tau_1, t_0) d\tau_1 \\ &= I + \frac{1}{i} \int_{t_0}^{t} H_I(\tau_1) V(\tau_1, t_0) d\tau_1 \,. \end{aligned}$$

By iterating the same equation,

$$\begin{aligned} V(t,t_0) &= I + \frac{1}{i}\int_{t_0}^{t} H_I(\tau_1)\left(I + \frac{1}{i}\int_{t_0}^{\tau_1} H_I(\tau_2)V(\tau_2,t_0)d\tau_2\right)d\tau_1 \\ &= I + \frac{1}{i}\int_{t_0}^{t} H_I(\tau_1)d\tau_1 + \frac{1}{i^2}\int_{t_0}^{t}\int_{t_0}^{\tau_1} H_I(\tau_1)H_I(\tau_2)V(\tau_2,t_0)d\tau_2 d\tau_1 \\ &\vdots \\ &= I + \sum_{1}^{\infty}\frac{1}{i^n}\int_{t_0}^{t}\int_{t_0}^{\tau_1}\cdots\int_{t_0}^{\tau_{n-1}} H_I(\tau_1)\ldots H_I(\tau_n)d\tau_n\ldots d\tau_1 \\ &= I + \sum_{1}^{\infty}\frac{1}{i^n n!}\int_{t_0}^{t}\int_{t_0}^{t}\cdots\int_{t_0}^{t}\mathcal{T}[H_I(\tau_1)\ldots H_I(\tau_n)]d\tau_n\ldots d\tau_1\,. \end{aligned}$$

But,

$$S = \lim_{\substack{t_0\to-\infty \\ t_1\to+\infty}} V(t_1,t_0)\,,$$

so we have our result. [7] □

A.5 Feynman Diagrams

The terms in the perturbative expansion of the scattering operator are typically calculated using Feynman diagrams. For example, Figure A.1 is the leading-order Feynman diagram for the electromagnetic interaction between an electron and a muon.

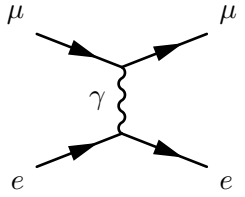


Figure A.1: A Feynman diagram showing an electron interacting with a muon via the photon.

Each line in the diagram represents a particle and each vertex represents an interaction between the particles corresponding to the lines touching the vertex. Time flows to the right so that the incoming state is an electron with momentum p_1 and a muon with momentum p_2, and the outgoing state is an electron with momentum p_1' and a muon with momentum p_2'. Let q denote the momentum of the photon that mediates the interaction. The incoming electron external line contributes a spinor u_1, the incoming muon line contributes a

spinor u_2, the outgoing electron line contributes a spinor $\bar{u_1}$, and the outgoing muon line contributes a spinor $\bar{u_2}$. Each vertex contributes a factor $ig_e\gamma^0\gamma^\mu$ that goes in between the corresponding spinors. The electron-photon vertex additionally contributes its momentum conserving factor $(2\pi)^4\delta^4(p_1-p_1'+q)$, and the muon-photon vertex additionally contributes its momentum conserving factor $(2\pi)^4\delta^4(p_2-p_2'-q)$. The internal photon line contributes a propagator $\frac{-ig_{\mu\nu}}{q^2}$ that goes in between the electron and muon factors, and additionally, contributes an integration factor $\frac{d^4q}{(2\pi)^4}$. Putting this all together, we get

$$\begin{aligned}
&\int \left[u_1 ig_e\gamma^0\gamma^\mu\bar{u_1}\right]\left[\frac{-ig_{\mu\nu}}{q^2}\right]\left[u_2 ig_e\gamma^0\gamma^\nu\bar{u_2}\right](2\pi)^4\delta^4(p_1-p_1'+q)(2\pi)^4\delta^4(p_2-p_2'-q)\frac{d^4q}{(2\pi)^4}\\
&= ig_e^2(2\pi)^4\int\left[u_1\gamma^0\gamma^\mu\bar{u_1}\right]\left[\frac{g_{\mu\nu}}{q^2}\right][u_2\gamma^0\gamma^\nu\bar{u_2}]\delta^4(p_1-p_1'+q)\delta^4(p_2-p_2'-q)d^4q\\
&= (2\pi)^4\delta^4(p_2-p_2'+p_1-p_1')\left[\frac{ig_e^2}{(p_1'-p_1)^2}\right]\left[u_1\gamma^0\gamma^\mu\bar{u_1}\right]\left[u_2\gamma_0\gamma_\mu\bar{u_2}\right]\,. \qquad \text{(A.49)}
\end{aligned}$$

Finally, we remove the overall conservation of momentum factor $(2\pi)^4\delta^4(p_2-p_2'+p_1-p_1')$ and replace it with i to obtain the leading-order probability amplitude:

$$\mathcal{M} = -\frac{g_e^2}{(p_1'-p_1)^2}\left[u_1\gamma^0\gamma^\mu\bar{u_1}\right]\left[u_2\gamma_0\gamma_\mu\bar{u_2}\right]\,. \qquad \text{(A.50)}$$

The internal lines of a Feynman diagram are commonly thought of as *virtual particles*. While the external lines represent particles that are highly localized in four-momentum, the internal lines are quite the opposite. As particles, they are likely to violate energy-momentum conservation. However, the invariant mass of their decay products may follow a distribution that is peaked at their rest mass; a key characteristic of particles. [6][7]

A.6 Left and Right Particles

If u_p is a bispinor in the Dirac basis representing a spin-1/2 particle in a definite four-momentum state p, then it can be shown that

$$\gamma^5 u_p = \begin{pmatrix} \frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{E+m} & 0 \\ 0 & \frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{E-m} \end{pmatrix} u_p\,, \qquad \text{(A.51)}$$

where σ^i is the ith Pauli Matrix for $i = 1, 2, 3$. If the particle is massless, the above equation reduces to

$$\gamma^5 u_p = (\hat{\mathbf{p}} \cdot \mathbf{\Sigma})\, u_p\,, \tag{A.52}$$

where

$$\Sigma = \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix} .[6] \tag{A.53}$$

Note that if the particle is left handed, $\hat{\mathbf{p}} \cdot \boldsymbol{\sigma} u_p = -u_p$, and if the particle is right handed, $\hat{\mathbf{p}} \cdot \boldsymbol{\sigma} u_p = +u_p$. Thus, if we let $L = \frac{1-\gamma^5}{2}$ and $R = \frac{1+\gamma^5}{2}$, we get (for a massless particle)

$$Lu_p = \begin{cases} u_p \text{ if } u_p \text{ is left handed} \\ 0 \text{ if } u_p \text{ is right handed} \end{cases} \tag{A.54}$$

and

$$Ru_p = \begin{cases} u_p \text{ if } u_p \text{ is right handed} \\ 0 \text{ if } u_p \text{ is left handed}\,. \end{cases} \tag{A.55}$$

Similarly, for a spinor $\bar{u_p}$ in the Dirac basis representing a spin-1/2 antiparticle in a definite four-momentum state p,

$$\gamma^5 \bar{u_p} = -\begin{pmatrix} \frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{E+m} & 0 \\ 0 & \frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{E-m} \end{pmatrix} \bar{u_p}\,, \tag{A.56}$$

where only the bottom two components of $\bar{u_p}$ are shown. In the massless case, this reduces to

$$\gamma^5 \bar{u_p} = -\,(\hat{\mathbf{p}} \cdot \mathbf{\Sigma})\, \bar{u_p}\,. \tag{A.57}$$

If we let $\bar{L} = \frac{1+\gamma^5}{2}$ and $\bar{R} = \frac{1-\gamma^5}{2}$, we get (for a massless antiparticle)

$$\bar{L}\bar{u_p} = \begin{cases} \bar{u_p} \text{ if } \bar{u_p} \text{ is left handed} \\ 0 \text{ if } \bar{u_p} \text{ is right handed} \end{cases} \tag{A.58}$$

and

$$\bar{R}\bar{u_p} = \begin{cases} \bar{u_p} \text{ if } \bar{u_p} \text{ is right handed} \\ 0 \text{ if } \bar{u_p} \text{ is left handed}\,. \end{cases} \tag{A.59}$$

These results show a few interesting facts. For a massless particle, it must be left handed in order to participate in weak interactions, and for a massless antiparticle, it must be right handed in order to participate in weak interactions. Thus, according to this model, even if right handed neutrinos existed, they would not interact, so we can assume that they do not exist. Note that for a massive particle u, Lu is *not* a spinor representing a left handed particle (Similarly for R,$\bar{L}$,$\bar{R}$). However, $L + R = I$, so we can decompose u as

$$u = Lu + Ru\,, \tag{A.60}$$

and call $u_L \coloneqq Lu$ the left part of u and $u_R \coloneqq Ru$ the right part of u. Since

$$L(Lu) = Lu \text{ and } L(Ru) = 0\,, \tag{A.61}$$

we can view u_L and u_R as distinct particles and say that u_L participates in weak interactions while u_R does not. Similarly, for an antiparticle $\bar{u}$, call $\bar{u}_{\bar{L}} \coloneqq \bar{L}\bar{u}$ the left part of $\bar{u}$, $\bar{u}_{\bar{R}} \coloneqq \bar{R}\bar{u}$ the right part of $\bar{u}$, and say that $\bar{u}_R$ participates in weak interactions while $\bar{u}_L$ does not. For any leptons with spinors a, b, this allows us to replace our weak vertex factor

$$a^\dagger g_w \gamma^0 \gamma^\mu (1 - \gamma^5) b\,, \tag{A.62}$$

with

$$2a^\dagger g_w \gamma^0 \gamma^\mu L b = 2a^\dagger g_w \gamma^0 \gamma^\mu L^2 b = 2a^\dagger L g_w \gamma^0 \gamma^\mu L b = 2a_L^\dagger g_w \gamma^0 \gamma^\mu b_L\,, \tag{A.63}$$

and replace our electromagnetic vertex factor

$$a^\dagger g_e \gamma^0 \gamma^\mu b \tag{A.64}$$

with

$$\begin{aligned}
&(a_L + a_R)^\dagger g_e \gamma^0 \gamma^\mu (b_L + b_R) \\
&= a^\dagger L g_e \gamma^0 \gamma^\mu L b + \cancel{a^\dagger L g_e \gamma^0 \gamma^\mu R b} \\
&+ \cancel{a^\dagger R g_e \gamma^0 \gamma^\mu L b} + a^\dagger R g_e \gamma^0 \gamma^\mu R b \\
&= a_L^\dagger g_e \gamma^0 \gamma^\mu b_L + a_R^\dagger g_e \gamma^0 \gamma^\mu b_R\,.
\end{aligned}$$

This is nice because we can now look at the weak and electromagnetic interactions as having the same vertex factors modulo a coupling constant, except that the weak interaction only acts on the "left" particles. [6]

Appendix B

Data Selection

Here we show the details of the signal/background selection criteria, for four different cases: signal taken from MC, signal taken from data, background taken from MC, and background taken from data.

Signal - MC Data:

- `L2_tauNoCut && EF_tauNoCut`,
- `(((vxp_type->at(vx)== 1)|| (vxp_type->at(vx)== 3))&& (vxp_nTracks->at(vx)> 3))` for at least one `vx` $\in \{$`0`$,\ldots,$`vxp_n`$\}$,
- L1 RoI must have $E_T > 8\,\text{GeV}$,
- Match L1 RoI to L2 and EF RoIs for the tauNoCut chain,
- EF-Offline $\Delta R < 0.2$ match,
- `(tau_author == 1)|| (tau_author == 3)`,
- `tau_Et >` $15\,\text{GeV}$,
- `fabs(tau_eta)< 2.2`,
- `tau_numTrack > 0`,
- `tau_JetBDTSigMedium`,

- Offline-Truth match with `tau_trueTauAssoc_index`,
- `trueTau_vis_Et >` 10 GeV,
- `fabs(trueTau_vis_eta)<` 2.5.

Signal - Collision Data:

- `L2_tauNoCut && EF_tauNoCut`,
- `(((vxp_type->at(vx)== 1)|| (vxp_type->at(vx)== 3))&& (vxp_nTracks->at(vx)> 3))` for at least one `vx` $\in \{$`0`,...,`vxp_n`$\}$,
- L1 RoI must have $E_T > 8$ GeV,
- Match L1 RoI to L2 and EF RoIs for the tauNoCut chain,
- EF-Offline $\Delta R < 0.2$ match,
- `(tau_author == 1)|| (tau_author == 3)`,
- `tau_Et >` 15 GeV,
- `fabs(tau_eta)<` 2.2,
- `tau_numTrack > 0`,
- `tau_JetBDTSigMedium`.

Background - MC Data:

- `L2_tauNoCut && EF_tauNoCut`,
- L1 RoI must have $E_T > 8$ GeV,
- Match L1 RoI to L2 RoI for the tauNoCut chain.

Background - Collision Data:

- `L2_tauNoCut && EF_tauNoCut`,

- L1 RoI must have $E_T > 8\,\text{GeV}$,
- Match L1 RoI to L2 and EF RoIs for the tauNoCut chain,
- EF tau - Offline jet $\Delta R < 0.2$ match,
- `jet_AntiKt4TopoEM_pt > 15` GeV/c,
- `fabs(jet_AntiKt4TopoEM_eta)< 2.2`,
- There must be another offline jet separated from the offline jet by $|\Delta\phi| > 2.7$ and satisfying the same p_T and η requirements.

`vxp_type` refers to the type of the given primitive vertex. It can either be of type 1 (primary) or type 3 (pile-up). `vxp_nTracks` refers to the number of tracks for the primitive vertex. [64] We use standard selections for these variables [65][66]. `tau_author` determines how the offline tau is seeded. If it is 1, then it is only calorimeter seeded; 2 implies that it is only track seeded; and 3 implies that it is both calorimeter and track seeded. Currently, all offline taus are calorimeter seeded, so the requirement on `tau_author` for the 2012 samples is unnecessary, but harmless [67]. Naturally, `tau_Et` is the transverse energy of the offline tau, `tau_eta` is its pseudorapidity, and `tau_numTrack` is the number of tracks it contains. `tau_JetBDTSigMedium` is a boolean-valued variable that determines whether or not the offline tau passes the medium BDT identification criteria. For the D3PDs with the `p1130` tag, a manual calculation needs to be performed in order to determine the true value of this variable. This is detailed in [67]. A *true tau* is the digital representation of a tau that was generated by the MC simulation, and is thus known to be a tau with absolute certainty. `tau_trueTauAssoc_index` maps an offline tau index to a true tau index by finding the closest match in terms of ΔR. If the resulting index is non-negative, then it is used to obtain the variables for the corresponding true tau. `trueTau_vis_Et` is the visible transverse energy of the true tau and `trueTau_vis_eta` is the visible pseudorapidity of the true tau. By visible, we mean that we only include the decay products that can be detected by the detector; in particular, we ignore neutrinos. Finally, `jet_AntiKt4TopoEM_pt` is the

transverse momentum of the offline jet and `jet_AntiKt4TopoEM_eta` is its pseudorapidity. [68][69]

Appendix C

Source Code

Here we show the source code for a few calculations that our algorithm performs. Section C.1 shows how the variables `HADtoEMenergy` and `EnergyTonCells` are calculated, Section C.2 shows how the variable `scalarPtSumIsoToCore` is calculated, and Section C.3 shows the selection criteria, for any input RoI, of the reference trigger `L2_TAU20_MEDIUM1`. More of the source code will be made available at the author's web site: https://akrmn.org.

C.1 `HADtoEMenergy` and `EnergyTonCells`

```
float const energy_grn = 0.1;
float const energy_max = 1000000000.;
float const nCells_grn = 0.1;

if ((fabs(EMenergy) <= energy_grn)) {
  if (fabs(HADenergy) <= energy_grn) {
    if (((EMenergy >= 0.) && (HADenergy >= 0.)) ||
       ((EMenergy < 0.) && (HADenergy < 0.))) {
      HADtoEMenergy = 1.;
    }
```

```
    else {
      HADtoEMenergy = -1.;
    }
  }
  else if (fabs(HADenergy) < energy_max) {
    if (EMenergy >= 0.) {
      HADtoEMenergy = HADenergy/energy_grn;
    }
    else {
      HADtoEMenergy = -HADenergy/energy_grn;
    }
  }
  else {
    if (((EMenergy >= 0.) && (HADenergy >= 0.)) ||
        ((EMenergy < 0.) && (HADenergy < 0.))) {
      HADtoEMenergy = energy_max/energy_grn;
    }
    else {
      HADtoEMenergy = -energy_max/energy_grn;
    }
  }
}
else if (HADenergy >= energy_max) {
  HADtoEMenergy = energy_max/EMenergy;
}
else if (HADenergy <= -energy_max) {
  HADtoEMenergy = -energy_max/EMenergy;
}
```

```
else {
  HADtoEMenergy = HADenergy/EMenergy;
}

float EMHADenergy = EMenergy + HADenergy;

if (numTotCells == 0) {
  if (fabs(EMHADenergy) <= energy_grn) {
    if (EMHADenergy >= 0.){
      EnergyTonCells = 1.;
    }
    else {
      EnergyTonCells = -1.;
    }
  }
  else if (fabs(EMHADenergy) <= energy_max) {
    EnergyTonCells = EMHADenergy/nCells_grn;
  }
  else {
    if (EMHADenergy >= 0.) {
      EnergyTonCells = energy_max/nCells_grn;
    }
    else {
      EnergyTonCells = -energy_max/nCells_grn;
    }
  }
}
else if (EMHADenergy >= energy_max) {
```

```
  EnergyTonCells = energy_max/((float)numTotCells);
}
else if (EMHADenergy <= -energy_max) {
  EnergyTonCells = -energy_max/((float)numTotCells);
}
else {
  EnergyTonCells = EMHADenergy/((float)numTotCells);
}
```

C.2 scalarPtSumIsoToCore

```
float const energy_grn = 0.1;
float const energy_max = 1000000000.;

if ((fabs(scalarPtSumCore) <= energy_grn)) {
  if (fabs(scalarPtSumIso) <= energy_grn) {
    if (((scalarPtSumCore >= 0.) && (scalarPtSumIso >= 0.))
       || ((scalarPtSumCore < 0.) && (scalarPtSumIso <
       0.))) {
      scalarPtSumIsoToCore = 1.;
    }
    else {
      scalarPtSumIsoToCore = -1.;
    }
  }
  else if (fabs(scalarPtSumIso) < energy_max) {
    if (scalarPtSumCore >= 0.) {
      scalarPtSumIsoToCore = scalarPtSumIso/energy_grn;
    }
```

```
    else {
      scalarPtSumIsoToCore = -scalarPtSumIso/energy_grn;
    }
  }
  else {
    if (((scalarPtSumCore >= 0.) && (scalarPtSumIso >= 0.))
       || ((scalarPtSumCore < 0.) && (scalarPtSumIso <
       0.))) {
      scalarPtSumIsoToCore = energy_max/energy_grn;
    }
    else {
      scalarPtSumIsoToCore = -energy_max/energy_grn;
    }
  }
}
else if (scalarPtSumIso >= energy_max) {
  scalarPtSumIsoToCore = energy_max/scalarPtSumCore;
}
else if (scalarPtSumIso <= -energy_max) {
  scalarPtSumIsoToCore = -energy_max/scalarPtSumCore;
}
else {
  scalarPtSumIsoToCore = scalarPtSumIso/scalarPtSumCore;
}
```

C.3 Reference Trigger: L2_TAU20_MEDIUM1

```
if (etMedium <= 15199.) {
  return(-1);
```

```
}
if(CoreFrac <= 0.7485) {
  return(-2);
}
if ((nCoreTracks < 1) || (nCoreTracks > 4)) {
  return(-3);
}
if (scalarPtSumIso < 0. || scalarPtSumCore <= 0.) {
  return(-4);
}
float scalarPtSumIsoToCore = scalarPtSumIso/scalarPtSumCore;
if (nCoreTracks == 1) {
  if (scalarPtSumIsoToCore >= 0.001) {
    return(-5);
  }
  if ((etOverPtLeadTrk < 0.) || (etOverPtLeadTrk >= 8.84)) {
    return(-6);
  }
  if (trkAvgDist >= 0.105) {
    return(-7);
  }
}
else {
  if (scalarPtSumIsoToCore >= 0.099) {
    return(-8);
  }
  if ((etOverPtLeadTrk < 0.) || (etOverPtLeadTrk >= 2.84)) {
    return(-9);
```

```
    }
    if (trkAvgDist >= 0.073) {
      return(-10);
    }
  }
  return(0);
```

Note: A return value of 0 means that the RoI is accepted; otherwise, it is rejected. [60]

References

[1] **ATLAS** Collaboration, ATLAS detector and physics performance: Technical Design Report, 1. Technical Design Report ATLAS. CERN, Geneva, 1999. CDS:391176.

[2] C. Ohm, D. Milstead, and T. Moa, Searches for exotic stable massive particles with the ATLAS experiment. PhD thesis, Stockholm U., Stockholm, 2011. CDS:1504815.

[3] "Study of the spin of the new boson with up to 25 fb^{-1} of ATLAS data," CERN Document Server (Apr, 2013) , CDS:1542341.

[4] Wikipedia, "Standard Model of Elementary Particles." https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg, 2012.

[5] **ATLAS** Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," Phys. Lett. B **716** (Aug, 2012) 1–29. 39 p.

[6] D. Griffiths, Introduction to Elementary Particles - Second, Revised Edition. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2008.

[7] G. Folland, Quantum Field Theory - a Tourist Guide for Mathematicians. American Mathematical Society, Providence, Rhode Island, United States of America, 2008.

[8] Wikipedia, "Mexican Hat Potential." https://en.wikipedia.org/wiki/File:Mexican_hat_potential_polar.svg, 2009.

[9] R. Moore, "University of Alberta PHYS 485 Lecture 22: Non-standard Models." Available upon request from author, April, 2012.

[10] S. P. Martin, "A Supersymmetry Primer," arXiv (2011) , hep-ph/9709356.

[11] **TOTEM** Collaboration, "Luminosity-independent measurements of total, elastic and inelastic cross-sections at $\sqrt{s} = 7\,\text{TeV}$," EPL (Europhysics Letters) **101** no. 2, (2013) 21004.

[12] C. Lefevre, "LHC: the guide (English version).," CERN Document Server (Feb, 2009) , CDS:1165534.

[13] O. S. Brüning et al., LHC Design Report. CERN, Geneva, 2004.

[14] J. Pinfold et al., "Technical Design Report of the MoEDAL Experiment," CERN Document Server (Jun, 2009) , CDS:1181486.

[15] K. Yurkewicz, "The skinny on the LHC's heavy ions." http://www.symmetrymagazine.org/breaking/2010/11/05/the-skinny-on-the-lhcs-heavy-ions, 2010.

[16] CERN Press Office, "First beam in the LHC - accelerating science." http://press.web.cern.ch/press-releases/2008/09/first-beam-lhc-accelerating-science, 2008.

[17] CERN Press Office, "CERN releases analysis of LHC incident." http://press.web.cern.ch/press-releases/2008/10/cern-releases-analysis-lhc-incident, 2008.

[18] CERN Press Office, "The LHC is back." http://press.web.cern.ch/press-releases/2009/11/lhc-back, 2009.

[19] CERN Press Office, "LHC research programme gets underway." http://press.web.cern.ch/press-releases/2010/03/lhc-research-programme-gets-underway, 2010.

[20] CERN Press Office, "LHC physics data taking gets underway at new record collision energy of 8TeV." http://press.web.cern.ch/press-releases/2012/04/lhc-physics-data-taking-gets-underway-new-record-collision-energy-8tev, 2012.

[21] CERN Press Office, "First three-year LHC running period reaches a conclusion." http://press.web.cern.ch/press-releases/2013/02/first-three-year-lhc-running-period-reaches-conclusion, 2013.

[22] F. Hugging, "The ATLAS pixel detector," Nuclear Science, IEEE Transactions on **53** no. 3, (June, 2006) 1732 – 1736.

[23] J. Pequenao, "Computer generated image of the whole ATLAS detector," CERN Document Server (Mar, 2008) , CDS:1095924.

[24] P. Jenni et al., ATLAS high-level trigger, data-acquisition and controls: Technical Design Report. Technical Design Report ATLAS. CERN, Geneva, 2003. CDS:616089.

[25] **ATLAS** Collaboration, "Performance of the ATLAS Trigger System in 2010," Eur. Phys. J. C **72** (January, 2012) 1849.

[26] **ATLAS** Collaboration, "Performance of the ATLAS tau trigger in 2011," CERN Document Server (Jan, 2013) , CDS:1510157.

[27] **ATLAS** Collaboration, "The DPD format." https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/PhysicsAnalysisWorkBookDPDFormat, 2012.

[28] I. Antcheva et al., "ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization," Computer Physics Communications **180** no. 12, (2009) 2499 – 2512.

[29] P. F. Akesson and E. Moyse, "Event Data Model in ATLAS," CERN Document Server (2005) , CDS:865580.

[30] **ATLAS** Collaboration, "AMI Portal Home." http://ami.in2p3.fr, 2013.

[31] T. Sjöstrand et al., "PYTHIA." http://http://home.thep.lu.se/~torbjorn/Pythia.html, 2013.

[32] **Particle Data Group** Collaboration, "Review of particle physics," Phys. Rev. D **86** (Jul, 2012) 010001.

[33] A. Denner et al., "Standard Model Higgs-Boson Branching Ratios with Uncertainties," arXiv (Aug, 2011) , arXiv:1107.5909 [hep-ph].

[34] **ATLAS** Collaboration, "TAPM Tau Slice Group." https://twiki.cern.ch/twiki/bin/viewauth/Atlas/TapmTau, 2013.

[35] M. Tosi, "Results on the Search for MSSM Neutral and Charged Higgs bosons (CMS)," CERN Document Server (Feb, 2013) , CDS:1525917.

[36] R. Reece, "Searches for charged Higgs bosons, supersymmetry, and exotica with tau leptons with the ATLAS and CMS detectors at the LHC," CERN Document Server (Dec, 2012) , CDS:1499581.

[37] **ATLAS** Collaboration, "Performance of the ATLAS Trigger System in 2010," Eur. Phys. J. C **72** (Oct, 2011) 1849. 63 p.

[38] W. Lampl et al., "Calorimeter Clustering Algorithms: Description and Performance," CERN Document Server (Apr, 2008) , CDS:1099735.

[39] N. Ghodbane, "Performance of Jet Algorithms for the Top Quark Physics at the ATLAS Experiment.," CERN Document Server (Aug, 2009) , CDS:1199764.

[40] G. C. Blazey et al., "Run II Jet Physics: Proceedings of the Run II QCD and Weak Boson Physics Workshop," arXiv (May, 2000) , hep-ex/0005012.

[41] A. Christov et al., "Performance of the tau reconstruction algorithm with release 15.3.1.6 and mc08 data," CERN Document Server (Oct, 2010) , CDS:1302271.

[42] S. Fleischmann and K. Desch, Tau lepton reconstruction with energy flow and the search for R-parity violating supersymmetry at the ATLAS experiment. PhD thesis, Bonn U., 2011. CDS:1504815.

[43] **ATLAS** Collaboration, "The ATLAS Simulation Infrastructure," Eur. Phys. J. C **70** (May, 2010) 823–874. 53 p.

[44] M. Shamim, "ATLAS Hadronic tau trigger," CDS:1452928.

[45] **ATLAS** Collaboration, "Public Tau Trigger Plots for Collision Data." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TauTriggerPublicResults, 2012.

[46] **ATLAS** Collaboration, "Performance of the Reconstruction and Identification of Hadronic Tau Decays with ATLAS," CERN Document Server (Nov, 2011) , CDS:1398195.

[47] C. Cuenca Almenar, "MC study of tracking for L2 tau triggers: Imporving the performance for 2012," CERN Document Server (Feb, 2012) , CDS:1422720.

[48] **ATLAS** Collaboration, "Measurement of τ polarization in $W \to \tau\nu$ decays with the ATLAS detector in pp collisions at $\sqrt{s} = 7\,\text{TeV}$," Eur. Phys. J. C **72** (May, 2012) 2062. 25 p.

[49] **ATLAS** Collaboration, "Search for charged Higgs bosons decaying via $H^{+-} \to \tau\nu$ in $t\bar{t}$ events using $4.6\,\text{fb}^{-1}$ of pp collision data at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector," CERN Document Server (Mar, 2012) , CDS:1429659.

[50] **ATLAS** Collaboration, "Search for the Standard Model Higgs boson in the $H \to \tau^+\tau^-$ decay mode with 4.7 fb^{-1} of ATLAS data at 7 TeV," CERN Document Server (Mar, 2012) , CDS:1429662.

[51] **ATLAS** Collaboration, "$Z \to \tau\tau$ cross section measurement in proton-proton collisions at 7 TeV with the ATLAS experiment," CERN Document Server (Feb, 2012) , CDS:1426991.

[52] L. Tompkins, "FTK: A Hardware Track Finder for the ATLAS Trigger System," CDS:1514162.

[53] P. Speckmayer et al., "The toolkit for multivariate data analysis, TMVA 4," Journal of Physics: Conference Series **219** no. 3, (2010) 032057.

[54] A. Hoecker et al., "TMVA - Toolkit for Multivariate Data Analysis," arXiv (2009) , physics/0703039.

[55] **ATLAS** Collaboration, "Centrally Produced D3PDs for the TauWG." https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TauWGDPDs, 2013.

[56] **ATLAS** Collaboration, "/Trigger/TrigAlgorithms/TrigT2CaloTau." https://svnweb.cern.ch/trac/atlasoff/browser/Trigger/TrigAlgorithms/TrigT2CaloTau, 2013.

[57] **ATLAS** Collaboration, "/Trigger/TrigAlgorithms/TrigT2Tau." https://svnweb.cern.ch/trac/atlasoff/browser/Trigger/TrigAlgorithms/TrigT2Tau, 2013.

[58] **ATLAS** Collaboration, "/Trigger/TrigAlgorithms/TrigT2IDTau." https://svnweb.cern.ch/trac/atlasoff/browser/Trigger/TrigAlgorithms/TrigT2IDTau, 2013.

[59] R. Moore, "University of Alberta PHYS 699 Lecture: Expectations and Uncertainty for Efficiencies." Available upon request from author, June, 2011.

[60] **ATLAS** Collaboration, "/Trigger/TrigHypothesis/TrigTauHypo." https://svnweb.cern.ch/trac/atlasoff/browser/Trigger/TrigHypothesis/TrigTauHypo, 2013.

[61] **ATLAS** Collaboration, "Tau Triggers for Physics Analyses." https://twiki.cern.ch/twiki/bin/viewauth/Atlas/TauTriggerPhysicsTriggerRecommendations2012, 2012.

[62] T. G. Rizzo, "Z' Phenomenology and the LHC," arXiv (2006) , hep-ph/0610104.

[63] **TOTEM** Collaboration, "A luminosity-independent measurement of the proton-proton total cross-section at $\sqrt{s} = 8\,\text{TeV}$," CERN Document Server (Nov, 2012) , CDS:1495764.

[64] **ATLAS** Collaboration, "TopD3PDForTopMass." https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TopD3PDForTopMass, 2011.

[65] **ATLAS** Collaboration, "ZToTwoTausToLepHad7TeVAnalysis2011." https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ZToTwoTausToLepHad7TeVAnalysis2011, 2011.

[66] **ATLAS** Collaboration, "LepJetsSelectionCuts." https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/LepJetsSelectionCuts, 2012.

[67] **ATLAS** Collaboration, "TauRecommendationsSummerConf2012." https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TauRecommendationsSummerConf2012, 2013.

[68] **ATLAS** Collaboration, "How to Make D3PDs for the Tau Group." https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TauD3PDMaker, 2012.

[69] **ATLAS** Collaboration, "Accessing MC truth - Release 15." https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/PhysicsAnalysisWorkBookExMCRel15, 2012.