Associated Production of a Top Quark and a Higgs Boson in *pp* Collisions at 13 TeV Using the ATLAS Detector

Han Na We

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

May 2020

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, ...
Date Signature

1. Gutachter: Prof. Dr. Ian C. Brock
2. Gutachter: Prof. Dr. Jochen Dingfelder

Acknowledgements

First, I would like to thank the Bonn Cologne Graduate School (BCGS) admission committee for giving me an opportunity to pursue my master degree in physics at this wonderful university. I had the fortune to experience an awesome German education system.

I would like to thank Prof. Dr. Ian Brock for giving me an opportunity to write my thesis in his group. I really appreciate you for all your guidance, patience and encouragement during my thesis. I would also like to thank Prof. Dr. Jochen Dingfelder for agreeing to be my second supervisor for this thesis.

Furthermore, I would like to thank all the people in the Brock research group. A big thanks to Tanja, for all your guidance from the beginning till the last day of my thesis submission. I am very grateful for all your efforts for helping me whenever I had questions, and especially for thoroughly correcting this thesis. I would also like to thank Dr. Rui Zhang, Dr. Regina Moles for all your advice and help when I initially joined this group. I also thank Anji for your efforts for spending time to proofread my draft of this thesis. I would like to thank Chris, Christian, Federico and Dr. Oleh for giving suggestions when I had difficulty in solving some problems. Thank you, Maxx, for the good time discussing about interesting topics and helping me to keep a good mood while sharing the same office. A special thanks to all the other members of the Brock group (those I have not mentioned previously) who were in the group during my days there. I enjoyed a lot having nice barbecue parties and board game nights with you all.

Lastly, I would like to sincerely thank my parents who always sent their encouragement and support from the other side of the globe. Without your love and prayers, it would never have been possible for me to continue and finish my studies. I also thank all my friends and colleagues who supported me in Germany, and also those in Korea and other countries even if they were not together with me.

Contents

CHAPTER 1

Introduction

"Equipped with his five senses, man explores the universe around him and calls the adventure Science."

- Edwin Powell Hubble

Revealing the mystery of the birth of the universe and the formation of matter has been an everlong task for humankind to solve. In order to answer an enormous amount of profound questions on this journey, human beings have been developing various theories in the last centuries. Up to now, there are two main physical theories that are known to be well-established: the Standard Model (SM) of particle physics and General Relativity of cosmology. Through countless of analyses and decades of experiments, these theories have been experimentally validated to astounding precision.

Since the discovery of the electron in 1897 [1], which led human beings into the subatomic scale era, more particles have been revealed to be fundamental subsequently. With the discovery of the Higgs boson by the ATLAS and CMS collaborations at the LHC in 2012 [2], one of the last missing pieces of the SM fell into place, marking a great milestone in the history of particle physics. The properties and behavior of the Higgs boson matched many theoretical expectations, such as initially massless particles in the SM acquiring masses through the Brout-Englert-Higgs (BEH) Mechanism. However, there are still many open questions and inconsistencies remained in the SM, such as naturalness problem, matter-antimatter asymmetry, dark matter, dark energy, etc. Therefore, there are various Beyond Standard Model (BSM) theories that are proposed in an attempt to explain the incompatibility between the experimental observations and SM predictions.

One of the major places to test the SM is the regime of heavy elementary particles, such as the top quark, W, Z and Higgs bosons. The production of a single top quark in association with a Higgs boson (tH) allows a direct probe for the coupling of the top quark to the Higgs boson (y_t), as well as the coupling of the W boson to the Higgs boson (g_{HVV}). This process has a great advantage compared to other processes such as $t\bar{t}H$ that, it is uniquely sensitive to the relative sign of the coupling constants in addition to their magnitude. This feature originates from the fact that the tH diagrams, which involve the two distinct coupling constants y_t and g_{HVV}, maximally interfere in the destructive way. The searches for this process can be classified depending on the Higgs decay channel, and the main three channels are: multilepton (ML) channel which involves multiple leptons in the final state, $H \to b\bar{b}$, and $H \to \gamma\gamma$. This analysis focuses on the ML channel, especially on the channel where the Higgs boson decays into a pair of τ leptons, of which one decays hadronically and the other decays leptonically ($H \to \tau_h\tau_l$), namely the lep-had channel. Prior to the tHq analysis, the production of a single top quark in association with a Z boson (tZq) has been investigated for a precursor analysis, as it constitutes one of the most prominent backgrounds to the tHq.

The structure of this thesis is as follows. In chapter 2, a brief overview of the Standard Model, basic theoretical concepts used in collider physics, as well as a detailed description of top quark physics are presented. An introduction of the LHC and the ATLAS detector is given in chapter 3, together with an explanation about how particles are identified and reconstructed within the ATLAS detector. In chapter 4, the data and Monte Carlo simulated samples that are used for both tZq and tHq analyses are described. The tZq analysis is presented in chapter 5. In this chapter, the selection of events which give the tZq ML final state as well as the validation procedure of a framework named `tHqLoop` (for the tHq analysis) are shown. The tHq study is divided into two main parts: 1) origin of the light leptons described in chapter 6; 2) reconstruction of the lep-had channel discussed in chapter 7. In chapter 6, a general description of the event selection procedure of the tHq ML channel as well as a study on the origin of the light leptons are given. The second part of the tHq analysis is focused on the reconstruction of the lep-had final state. At last, in the last chapter a summary of the whole analysis as well as a short outlook of a possible direction for tHq analysis are presented.

CHAPTER 2

Theoretical concepts

In this chapter, a brief overview of the Standard Model of particle physics and theoretical concepts used in hadron collider physics is presented. First of all, an introduction to the constituents of the Standard Model and the relationship between them will be given. Afterwards, basic concepts commonly used in experimental particle physics will be discussed. Finally, the top quark and the Higgs boson will be explained in detail as they are key particles in this analysis.

2.1 The Standard Model

Since late 1970s, as a heart of our understanding of nature, the Standard Model of particle physics (SM) has been developed as a unified theoretical framework that describes the elementary particles[1]. The laws governing the behavior of these fundamental particles and their nature have been theoretically formulated as a set of Quantum Field Theories (QFTs). According to QFTs, a particle is described as an excitation of a quantum field that satisfies its own quantum mechanical field equations [3]. So far, the SM has successfully explained nearly all experimental results and predicted a wide variety of phenomena in high precision, and is therefore known as a well-established model in particle physics.

2.1.1 Ingredients of the SM

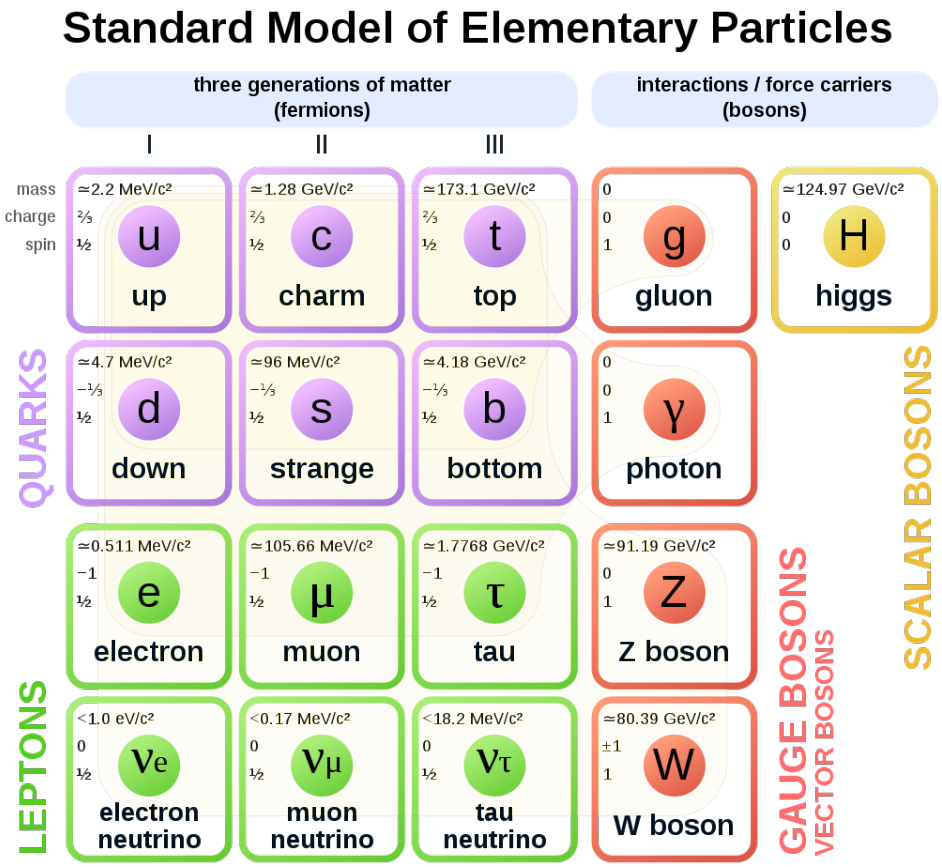
All known elementary particles included in the SM are shown in Fig 2.1 They can be classified into two types: fermions and bosons, based on their spins[2]. Each type of particles is explained in more detail in the following section.

Fermions

The fermions, also known as matter particles, are half-integer spin particles. Their behavior is thus constrained by the Pauli exclusion principle according to the law of spin–statistics [3]. All fermions experience the weak force, and with the exception of neutrinos, which are electrically neutral, all other fermions interact via the electromagnetic interaction which is described by Quantum Eletrodynamics (QED) (more details in Section 2.1.2). There are twelve fermions (and their antiparticles) in the SM in total to date, which can be subdivided further into two different classifications, six quarks and six

[1] Elementary particles or fundamental particles are particles that are thought to have no internal structure, and are the smallest known building blocks of the universe.
[2] Spin is an intrinsic angular momentum carried by the particles.

leptons, depending on whether they interact via the strong interaction, which is theoretically formulated by Quantum Chromodynamics (QCD). Unlike leptons which carry no *color charge*, which is equivalent of electric charge q, quarks carry a color charge and thus interact via the strong interaction. Quarks can be combined into nucleons, for example the proton with two u-quarks and one d-quark, and the neutron with one u-quark and two d-quarks. Atoms consist of these nuclei and electrons around them.



ec

Figure 2.1: An overview of the elementary particles in the SM with mass, charge and spin information. The fermions are classified into quarks (purple) and leptons (green), and bosons can be subdivided into gauge bosons (red) and scalar bosons (yellow). The antiparticles are now shown here [4].

Antiparticles of fermions

Apart from the fermions shown in Fig 2.1, there is a set of *antiparticles*, of which each corresponds to a fermion with the same mass and the opposite charge. The existence of antiparticles was postulated in an attempt to provide a physical interpretation for the negative energy solutions from the Dirac equation [3]. The existence of antiparticles was confirmed by the discovery of antiproton and antineutron in 1995 at the University of California, Berkeley [5].

Generations of fermions

The component of spin along the direction of motion of a particle is known as its helicity. For a spin-half fermion, the component of spin measured along any axis is quantized to be either $\pm 1/2$. These two possible helicity states are termed right-handed and left-handed helicity states. Only fermions that are left-handed (or with a negative chirality[3]) can carry a weak isospin[4], and they can be paired in terms of a weak isospin doublet[5]. The pairs from the quarks and leptons are arranged to form a generation, with corresponding particles exhibiting similar physical behavior. The only differences between generations are the masses and flavors of particles. The first generation has the lightest particles, whereas particles in the third generation have the largest masses. Each generation of quarks consists of one "up-type" quark (u, c, t) and one "down-type" quark (d, s, b), and each lepton generation consists of one charged lepton (e, μ, τ) and an associated neutrino (ν_e, ν_μ, ν_τ). So far, only three generations have been observed, supported by experimental result from LEP Collaboration, that the number of neutrino types is $N_\nu = 2.9841 \pm 0.0083$ [6]. The existence of a fourth generation is also excluded in Higgs boson searches [7].

Bosons

The bosons are particles with integer spin, and their behavior is governed by the Bose-Einstein statistics. There are two types of bosons in the SM: spin-1 gauge bosons and spin-0 scalar bosons. The gauge bosons mediate the three fundamental forces between fermions except the gravitational force, i.e. the electromagnetic force, the weak force and the strong force. The photon (γ) mediates the eletromagnetic force, $W^\pm$ bosons together with Z boson mediate the weak force. The strong interaction is mediated by a particle called a gluon (g), which itself carries color charge unlike the photon. This leads to the gluon undergoing *self-interaction*. Except for gluons and photons, all other gauge bosons have non-zero masses, which arise from the spontaneous symmetry breaking by the Brout-Englert-Higgs mechanism [3]. The only scalar boson that has been observed to date is the neutral Higgs boson (H^0), whereas existence of charged Higgs bosons is proposed in several extended Higgs sectors. The details regarding how the bosons mediate the interactions in the SM are explained in Section 2.1.2.

2.1.2 Theoretical aspects in the SM

Quantum chromodynamics

The strong force is theoretically formulated by QCD, which is a quantum field theory explains the interaction between quarks and gluons. It is associated with a non-abelian gauge group $SU(3)_C$[6]. Here, a gauge group is a group transformation of the field variables that leaves the basic physics of the quantum field unchanged. This condition is called *gauge invariance*, and the group transformation is called *gauge transformation*. Every quark possess one of three different types of color charge, red (r), blue (b) and green (g), while every anti-quark carries an anti-color[7]. As shortly mentioned in Section 2.1.1 there are two features that QCD exhibits: color confinement and asymptotic freedom.

[3] The chirality of a particle is determined by whether the particle transforms in a right- or left-handed representation of the Poincaré group (a non-abelian Lie group). For massless particles, chirality is the same as helicity, which is the sign of the projection of the spin vector onto the momentum vector: “left” is negative, “right” is positive.

[4] Weak isospin is a quantum number relating to the weak interaction, usually denoted as T.

[5] Fermions with negative chirality have $T = \frac{1}{2}$ and can be grouped into doublets with $T_3 = \pm\frac{1}{2}$ which behave the same way under the weak interaction. Here, T_3 is the third component of weak isospin T. $T_3 = 0$ for right-handed fields.

[6] A non-abelian group is a group that has elements which do not necessarily commute. The C stands for color charge.

[7] Anti-color: anti-red ($\bar{r}$), anti-blue ($\bar{b}$) and anti-green ($\bar{g}$)

Color confinement is referred to as a phenomenon that particles with color charges are always confined into color-neutral (or color singlet) bounded states. This phenomenon is believed to originate from the *self confinement* of gluons, which arise from the fact that gluons themselves carry color charges unlike electrically neutral photons. The gluons have a combination of two color charges (one of r, g or b and one of $\bar{r}$, $\bar{g}$ and $\bar{b}$) in a superposition of states. Although there is no analytic proof of the concept of color confinement, it can be qualitatively understood by the interactions between two free quarks when they are pulled apart. Whereas the electric field decreases rapidly as two electrically charged particles are separated, rather a tube-like field is formed when a pair of color charges are moving apart, and the potential energy per distance between the pair stays constant. If the input of energy to the gluon field is high enough, which is above $\sim 1\ \mathrm{GeV\,fm^{-1}}$ (the potential energy of the color field), then an additional quark-antiquark pair will be created out of the vacuum. [3]

The QCD coupling strength, α_S, which is often called as QCD coupling constant, is actually *not* a constant. It is also called "running" coupling constant, as its value depends on the energy scale of the interaction being considered. The QCD coupling strength becomes larger at low-energy scales (large distance), which explains the confinement phenomenon, and becomes smaller at low-energy scales (short distance), which is called asymptotic freedom. In this way, the perturbation theory[8] can be used at the high-energy regime, whereas non-perturbation theory applies to the low-energy regime, where the strength is strong enough for hardonization to take place.

Eletroweak unification

The electromagnetic interaction and the weak interaction appear very different at everyday low energies, however, they would merge into a single electroweak interaction above the unification energy, at the order of 246 GeV, by the Glashow-Salam-Weinber (GSW) model. The weak and electromagnetic interactions are unified based on the gauge symmetry group $\mathrm{SU(2)_L \times U(1)_Y}$[9]. The Lagrangian of the weak interaction is invariant under the $\mathrm{SU(2)_L}$ local phase transformation, whereas the local gauge invariance of the Lagrangian of the electromagnetic interaction is associated with the abelian symmetry group $\mathrm{U(1)_Y}$. The introduction of new parameters, weak hypercharge Y and the third component of weak isospin T_3 which are the generators of $\mathrm{SU(2)_L}$ and $\mathrm{U(1)_Y}$ respectively, is crucial for the electroweak unification since they give rise to four electroweak gauge fields, the weak isospin fields W^+, W^-, W^0, and the weak hypercharge field B^0, which correspond to four massless bosons. The B^0 field mixes with the W^0 field with the weak mixing angle θ_W and this produces the observed Z gauge boson and the photon of quantum electrodynamics. These bosons (except the massless photon) become massive after the spontaneous symmetry breaking and the associated Higgs mechanism (see the next section).

Higgs mechanism

As mentioned in Section 2.1, particles are described by excitations of a quantum field that satisfies the appropriate quantum mechanical field equations. The dynamics of a quantum field theory can be expressed in terms of the Lagrangian density. The Lagrangian density of the SM, denoted as $\mathcal{L}_{\mathrm{SM}}$, is

[8] Perturbation theory is a set of approximation schemes directly related to mathematical perturbation for describing a complicated quantum system in terms of a simpler one. The idea is to start with a simple system for which a mathematical solution is known, and add an additional "perturbing" Hamiltonian representing a weak disturbance to the system. If the disturbance is not too large, the various physical quantities associated with the perturbed system can be expressed as "corrections" to those of the simple system.

[9] The index L stands for "left" based on the fact that the weak charged-current interaction is observed to couple only to left-handed chiral particles; Y stands for the weak hypercharge, $\mathrm{Y} = q - T_3$, where q is electric charge and T_3 is the third component of weak isospin.

required to satisfy gauge invariance defined by the local symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The $\mathcal{L}_{SM}$ can be simply written as:

$$\mathcal{L}_{SM} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}D\psi + \mathcal{L}_{H,\,bosons} + \mathcal{L}_{H,\,fermions}\,, \tag{2.1}$$

where the first term contains the energy-momentum tensor $F_{\mu\nu}$ of the electroweak fields and the gluon fields; second term contains the fermion field ψ and a covariant derivative D; and last two terms $\mathcal{L}_{H,\,bosons}$ and $\mathcal{L}_{H,\,fermions}$ are the Lagrangian of Higgs field associated with bosons and fermions. A direct addition of the mass term to SM Lagrangian breaks the local gauge invariance, and leads to non-renormalization of the SM Lagrangian. To solve this problem, a complex scalar doublet field ϕ called the Higgs field, along with the Higgs potential $V(\phi)$ is added to the Lagrangian as shown below:

$$\mathcal{L}_{H,\,bosons} = D_\mu\phi^\dagger D^\mu\phi - V(\phi)\,, \tag{2.2}$$

where $V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$, which is in a mexican-hat shape with $\mu^2 < 0$ and $\lambda > 0$. Although the potential itself is symmetric under the $SU(2)_L \times U(1)_Y$ transformation, its minimum is not at zero but at a non-zero value with $v = \sqrt{\frac{\mu^2}{\lambda}}$ (vacuum expectation value, *VEV*). Therefore, this leads to the spontaneous breaking of the original symmetry, since the perturbations around the minimum are not symmetric any more. The perturbative expansion includes both the radial and longitudinal directions in the potential. Each independent motion in which the fluctuations can take place without a potential gradient provides a degree of freedom, and as a complex doublet field, the Higgs field has four independent degrees of freedom. The fluctuation in the radial motion around the minimum will give rise to a massive Higgs boson, while three massless scalar bosons (also called Goldstone bosons) are created in tangential direction. These three degrees of freedom for the Goldstone bosons are "eaten" (absorbed) by $W^\pm$ and Z bosons. In terms of fermions, they acquire mass through so-called *Yukawa coupling* with the Higgs field. The photon remains massless since the symmetry breaking does not distort the electroweak symmetry that is responsible for the electromagnetic interaction to take place, i.e. the vacuum state stays still symmetric under $U(1)_{em}$[10] transformation.

2.2 Feynman diagrams

A Feyman diagram serves as a graphic tool for the description of particle interactions, but also a mathematical tool for the probability amplitudes evaluation. Example Feynman diagrams at Leading Order (LO) for the process $e^+e^- \to e^+e^-$ (bhabha scattering) are given in Fig 2.2. A Feynman diagram consists of points called vertices and lines with two different types: external and internal lines. The vertices indicate particle interactions that happen at the same point in space and at the same time. There are two different types of external lines: incoming lines extend from the past to a vertex which represent an initial state, and outgoing lines extend from a vertex to the future which represent the final state. The lines of antiparticles, which travel backward in time, are drawn in the opposite direction to particles. The lines connecting two vertices are internal lines, also called as propagators, and they are virtual particles (*off-shell* particles) which do not obey the energy-momentum conservation law.

Mathematically, as defined in the Feynman rule, all the lines and vertices contribute parts to the amplitude of a Feynman diagram, and the calculation of the probability amplitude can be done by multiplication of the contributions of all components within the diagram. For a given process, there could

[10] The $U(1)_{em}$ is an abelian gauge group which represents the electromagnetic interaction.

be more than one Feynman diagrams which share the same initial and final states. The total amplitude for a physical process is calculated by summing over all possible individual amplitude $\mathcal{M}_i$ of individual diagram i. A physical observable, such as cross-section, is evaluated by the taking the square of the sum. The amplitude of a process could be calculated more accurately if higher order[11] diagrams (e.g. diagrams contain internal loops) are included.

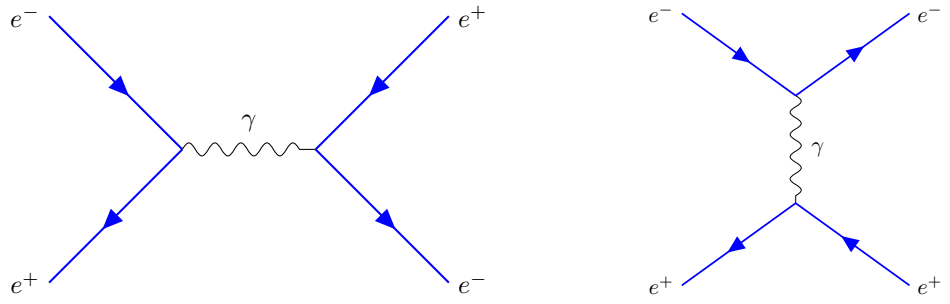


Figure 2.2: Example Feynman diagrams at LO for the process $e^+e^- \to e^+e^-$. Diagram 2.2(a) shows an annihilation process, and Diagram 2.2(b) shows a scattering process.

2.3 Collider physics

2.3.1 Basic concepts

Natural units

Natural units are physical units of measurement that are believed to be universal physical constants. The term of "natural" refers to that their definition originate purely from nature properties and not from human construct. They are generally used to simplify some specific equations in the laws of physics or to normalize certain physical quantities.

In this system of units, the speed of the light (denoted as c), the reduced Planck constant (denoted as $\hbar$), as well as the Bolztman constant (denoted as k_B) are defined to be equal to 1, as follows:

$$c = \hbar = k_\mathrm{B} = 1\,. \tag{2.3}$$

This allows many physical quantities with units of mass [kg], length [m] and time [s] to be able to expressed as the units of energy and momentum, electronvolt [eV][12], as well as their inverse units:

$$\begin{aligned} [\text{mass}] = [\text{energy}] = [\text{momentum}] = \text{eV}\,, \\ [\text{length}] = [\text{time}] = \text{eV}^{-1}\,. \end{aligned} \tag{2.4}$$

[11] According to the Feynman rules, each vertex in a Feynman diagram contributes a factor of the coupling constant. Thus, the order of each coupling constant is the number of vertices of that interaction.

[12] 1 eV = $1.602176634 \times 10^{-19}$ J

Cross-section and decay width

A cross-section, denoted as σ, represents the probability of an interaction between two particles during a collision. It has a unit of area [m^2], and can be imagined as an effective area where the reaction actually takes place, orthogonal to their relative motion. The unit of a cross-section in particle physics is generally expressed as a unit of barn (b), where 1 b = $10^{-28} m^2$. In particle physics, the commonly used cross-section units are picobarn (pb) and femtobarn (fb). (1 pb = 10^{-12} b, 1 fb = 10^{-15} b)

The cross-section can be calculated by integration of the matrix element over all possible phase spaces in a given process, with additional multiplication by factors arise from 4-vector momentum conservation, normalization and etc. Thus, it can be expressed as:

$$\sigma \propto \int \mid \mathcal{M} \mid^2 d\rho \,, \tag{2.5}$$

where $\mathcal{M}$ refers to the matrix element, and ρ the possible phase space.

There are two different measurements of cross-section depending on whether the phase space is calculated as a function of a specific observable or not. If a cross-section is measured only with respect to a certain variable such as an scattering angle or a momentum, then it is called as a *differential* (or exclusive) cross-section, and if it is not, it is called as an *inclusive* (or total) cross-section.

Decay rate

There is another quantity called a decay rate (or a decay width), denoted by Γ. It is used to reflect the probability of a particle process in a unit of time, but limited to a decay process. In a decay mode of a sufficiently large ensemble which consists of identical particles, the expected number of surviving particles after a certain time exhibits a function of an exponential. The time span after which the ensemble is reduced to $1/e$ of its initial size is called $lifetime$ (τ) of a such particle, and it is inversely proportional to the decay rate, $\tau = 1/\Gamma$. The unit of the lifetime is eV^{-1}, and consequently, the decay rate has a unit of eV.

In general, in a particle collision, not only one decay mode but multiple decay modes could take place, and the total decay rates can be simply written as the sum of decay rates of all individual decay (or partial decay width) modes, denoted as Γ_i:

$$\Gamma_{\text{tot}} = \sum_{i=1}^{n} \Gamma_i \,. \tag{2.6}$$

With a partial decay width known, the branching ratio could be calculated, which defines the fraction of a partial decay width with respect to the total decay width:

$$\mathcal{B}_i = \frac{\Gamma_i}{\Gamma_{\text{tot}}} \,. \tag{2.7}$$

Luminosity and event rate

In high-energy collision experiments, one of the most important quantities is the number of events, which is denoted as N:

$$N = \sigma \int \mathcal{L}(t) dt = \sigma \mathcal{L} \,, \tag{2.8}$$

where σ is the cross-section, $\mathcal{L}(t)$ is the *instantaneous luminosity*, which characterizes the performance and capability of an accelerator, and $\mathcal{L}$ is *integrated luminosity*, the integral of the instantaneous

luminosity with respect to time in a unit of b^{-1}.

As the distribution of the intensity of two head-on colliding[13] beams are assumed to be Gaussian-shaped, the instantaneous luminosity can be expressed as a function of a set of beam parameters, the crossing frequency (f), the number of particles in the beam ($n_{1/2}$) and the beam size in horizontal (σ_x) and vertical direction (σ_y):

$$\mathcal{L}(t) = \frac{f n_1 n_2}{4\pi\sigma_x\sigma_y} \,. \tag{2.9}$$

Technically, it is very tricky to precisely measure the instantaneous luminosity as running conditions vary with time. Thus, the integrated luminosity is more commonly used in data analysis, and the maximization of this luminosity is regarded as one of the main goals in most of collider experiments for larger data collection.

The event rate is defined as the number of events within in a certain time:

$$\frac{dN}{dt} = \sigma \cdot \mathcal{L}(t) \,, \tag{2.10}$$

and has a unit of $\left[\frac{\text{number of events}}{\text{m}^2 \cdot \text{s}}\right]$.

2.3.2 Parton distribution function

The parton model is proposed by Richard Feynman in the purpose of interpretation of cascades of radiation (a parton shower) produced from QCD processes and the interactions in particle collisions. According to this model, in high-energy hadron collision, hadrons such as protons or neutrons, can be considered as composite particles of so-called *partons*, and not the whole hadrons but a bunch of these partons interact with each other. The partons are referred to as valence and sea quarks[14] as well as gluons. Each parton carries away some certain amount of the longitudinal momentum of a hardon, and the fraction of the momentum carried by individual partons (p_i) can be described by a variable x called *Bjorken x*:

$$x_i = \frac{p_i}{p_{\text{hadron}}} \,, \tag{2.11}$$

As a consequence, the recorded collision energy can be lower than the total amount of the colliding beam energies. The probability density for finding a parton with a momentum fraction x_i at a given energy scale Q^2 ($= -q^2$, where q is a four momentum of modulus) can be described by the parton distribution function (PDF). [8] With the PDF of each parton known, the cross-section for a process $p_1 p_2 \to X$ (p_1, p_2 denotes two partons; X denotes a final state) at a given center-of-mass energy $\sqrt{s}$ can be calculated by:

$$\sigma_{p1p2\to X}(s) = \sum_{i,j=\text{partons}} \int dx_i dx_j \cdot f_1(x_i,\mu^2) f_2(x_j,\mu^2) \sigma_{ij\to X}(\hat{s},\mu^2) \,. \tag{2.12}$$

Here, f_1 and f_2 are the PDFs of the proton p_1 and p_2, respectively, at the energy scale of μ (also known as a *fractorization scale*). $\sigma_{ij\to X}$ is the cross-section for the process of two partons colliding into a final state of X with $\hat{s} = x_i x_j s$.

[13] A collider is a type of particle accelerator which accelerates particles to very high kinetic energy and letting them impact on other particles. There are two possible setups: 1) fixed target setup, where one beam of particles (the projectiles) is accelerated with a particle accelerator, and the other one is put into the path of the beam as a stationary target; and 2) head-on collider, where two beams of particles are accelerated and the beams are directed against each other. [3].

[14] The sea quarks are also known as virtual quarks, and they are always produced in quark-antiquark pairs by a gluon.

2.3.3 Kinematics

Center-of-mass energy

The center-of-mass energy of a particle system is the energy measured in the center-of-mass reference frame, and is defined as follows:

$$\sqrt{s} = \sqrt{\left(\sum_{i=1}^{2} E_i\right)^2 + \left(\sum_{i=1}^{2} \vec{p}_i\right)^2}, \tag{2.13}$$

where E_i and $\vec{p}_i$ are the energy and momentum of the initial-state particle i $(= 1, 2)$. Assuming two colliding particles as massless, the center-of-mass can be simplified as $\sqrt{s} = 2E_{\text{beam}}$ in the center-of-mass frame, $\sqrt{s} = 2m_{\text{a}}E_{\text{b}}$ in the laboratory frame.

Since the energy of the motion of the center of mass itself stays fixed at the center of mass and the internal properties of the system are unchangeable, it is available to probe new particles or to explore the internal structure of particles with all available energy. [9]

Transverse momentum and Missing transverse momentum

In collision experiments such as ATLAS, the partons are accelerated along the beam pipe, which is set to be the z-axis, therefore, in the ideal case, the particles in the final states have no longitudinal but only transverse components in their momentum[15]. Consequently, the total sum of the transverse momentum of the final objects must be zero to fulfill the momentum conservation law. The transverse momentum is described as:

$$p_{\text{T}} = \sqrt{p_{\text{x}}^2 + p_{\text{y}}^2}, \tag{2.14}$$

and here, p_{x} and p_{y} denote the momentum in x and y direction, respectively.

However, not all the final objects are detectable in a (reasonable-sized) general purpose detectors, such as neutrinos which rarely interact with other fundamental particles. The *missing* transverse momentum resulted in by these undetectable objects can be calculated by deduction of the total sum of momentum of all visible objects from zero:

$$\overrightarrow{E_{\text{T}}^{\text{miss}}} = -\sum_{i} \overrightarrow{p_{\text{T}}^{i}}. \tag{2.15}$$

Rapidity and pseudorapidity

In the frame of center-of-mass of a proton-proton collision, one of two colliding partons can be boosted along the beam direction due to the difference in the longitudinal momentum of the partons. Therefore, it is especially useful to define quantities which are invariant under longitudinal Lorentz boost. The rapidity (y) is one of such quantities, and is expressed as:

$$y = \frac{1}{2} \ln\left(\frac{E + p_z}{E - p_z}\right), \tag{2.16}$$

where E denotes the energy of the particle, and p_z the particle momentum along the beam axis.

[15] The transverse (or xy) plane is perpendicular to the beam direction (z-axis)

Another quantity called *pseudorapidity*, however, is more commonly used, which is defined as:

$$\eta = \frac{1}{2}\ln\left(\frac{|\vec{p}| + p_{\mathrm{L}}}{|\vec{p}| - p_{\mathrm{L}}}\right) = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{2.17}$$

where θ is the polar angle between the particle and the beam axis. Pseudorapidity asymptotically approaches the rapidity in the high energy limit, and it tends towards positive infinity at $\theta = 0^\circ$ (along the beam axis) and approaches zero at $\theta = 90^\circ$. Combined with azimuthal angle ϕ (the angle of a particle around the beam axis), a boost-independent angular distance ΔR is defined as follows:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}\,, \tag{2.18}$$

where $\Delta\eta$ and $\Delta\phi$ is referred to as the differences of pseudorapidity and azimuthal angle between two particles.

Transverse mass

The transverse mass, denoted as m_{T}, is invariant under Lorentz boost along the beam direction. It is defined as:

$$m_{\mathrm{T}}^2 = E_{\mathrm{T}}^2 - \vec{p}_{\mathrm{T}}^{\,2}\,, \tag{2.19}$$

where $E_{\mathrm{T}} = \sqrt{m^2 + (\vec{p_{\mathrm{T}}})^2}$ is the transverse energy that is defined using its true invariant mass m, and $\vec{p}_{\mathrm{T}}$ is the transverse momentum.

Pile-up

At a hadron collider, such as the LHC, the beam is bunched into bunches of which each consists of $\sim 10^{19}$ protons, and is accelerated in the ring bunch by bunch. Thus it is very likely that multiple protons interact with each other and produce not only interesting particles but also some other particles that are not interesting. All these collisions of no interest are collectively referred to as *pile-up* (PU). There are two different PUs: in-time PU and out-of-time PU. The in-time PU originates from multiple hadron interactions in the same bunches, whereas out-of-time PU is produced by earlier or later bunches. One of our ultimate goals is to maximize the number of interesting collisions, and one possible way is to maximize the number of particles per bunch. However, this means more pile-up, thus new methods of dealing with high pile-up condition are in need.

2.4 Top quark physics

In 1973, Kobayashi and Maskawa predicted the existence of the third generation of fermions in the SM, the bottom and the top quark, in order to explain CP violation in kaon decays. After the discovery of the bottom quark, the top quark, which is the weak isospin partner of the bottom quark, was discovered by the CDF and D0 colloborations at Tevatron in 1995. [10, 11] The basic properties and the main production processes of the top quark will be given in this section.

2.4.1 Top quark properties

As a member of the third generation, the top quark forms a weak isospin doublet with the bottom quark, with weak isospin value $T_3 = \frac{1}{2}$. It has spin $\frac{1}{2}$, charge $q = +\frac{2}{3}e$. The top quark has very unique and

interesting features compared to the other fermions in the SM. It is the heaviest particle that has been observed so far, with a mass of 173.34 ± 0.76 GeV. The mean lifetime of the top quark is too short to form bound states, with $\tau_t \approx 5 \times 10^{-25}$ s [12], thus it is also called "bare quark". It is believed to strongly couple to the Higgs boson due to its large mass. Since the mass of the top quark together with the mass of the W boson are linked to the mass of the Higgs boson, any deviation from this connection could indicate new physics. Thus, studying the properties of the top quark can provide a direct probe for the Higgs sector and the Beyond Standard Model (BSM) theories.

2.4.2 Top quark production

At hadron colliders, top quarks are produced mainly in two different processes: top quark pair production via the strong interaction and single top quark production by the electroweak interaction.

Top quark pair production

Top quark pairs can be produced either via gluon-gluon fusion ($gg \rightarrow t\bar{t}$) or quark-antiquark annihilation ($q\bar{q} \rightarrow t\bar{t}$). The LO Feynman diagrams of these processes are shown in Fig 5.6. The contribution of each process to the total $t\bar{t}$ production cross section depends on the center-of-mass energy of the collider and the PDF of the interacting particles. At the LHC, gg fusion is the dominant process, whereas at the Tevatron[16], $q\bar{q}$ annihilation is the dominant. At the center of mass energy $\sqrt{s} = 13$ GeV, the gg fusion accounts for ~ 90%. [12] The total cross-section for $t\bar{t}$ production has been measured experimentally by the ATLAS collaboration: [13]

$$\sigma_{t\bar{t}} = 826.4 \pm 3.6\ (\text{stat}) \pm 11.5\ (\text{syst}) \pm 15.7\ (\text{lumi}) \pm 1.9\ (\text{beam})\ \text{pb}\,. \tag{2.20}$$

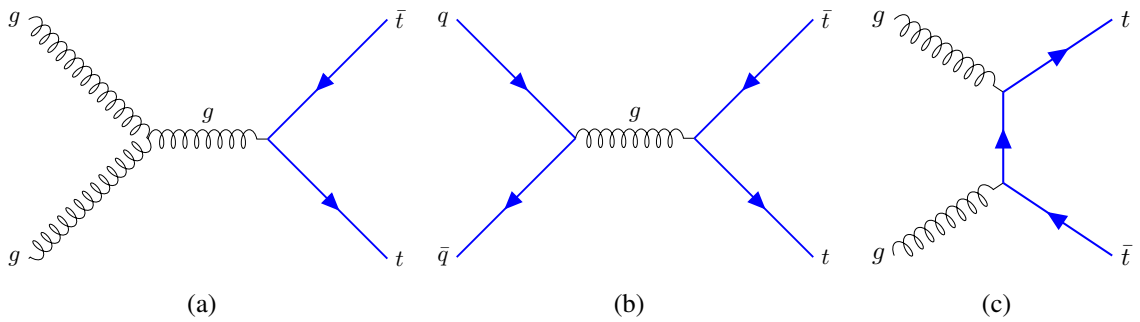


Figure 2.3: Example LO Feynman diagrams of top quark pair production by gluon-gluon fusion 2.3(a), quark-antiquark annihilation 2.3(b) and quark-antiquark annihilation 2.3(c).

Single top production

A single top quark is produced via the electroweak interaction. There are three main processes of single top quark production, and their LO Feynman diagrams are shown in Fig 2.4. The difference between these processes is whether the exchanged particle is on-shell or not.

[16] The Tevatron is a circular particle accelerator at the Fermi National Accelerator Laboratory (also known as Fermilab) in the United States.

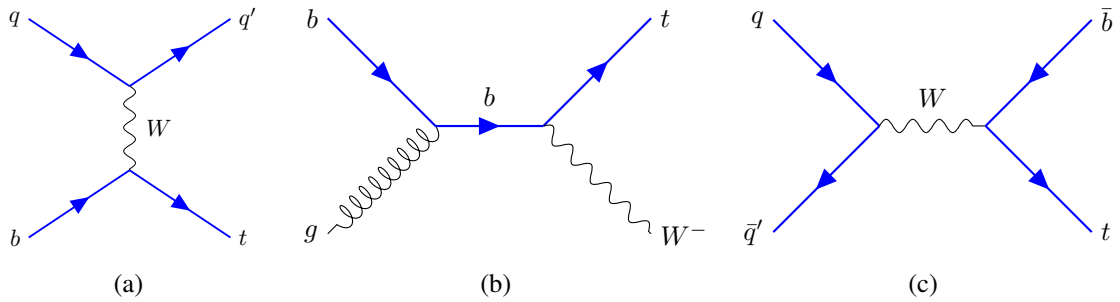


Figure 2.4: Example LO Feynman diagrams of single top quark production. Diagram 2.4(a) shows t-channel, diagram 2.4(b) corresponds to tW channel, and diagram 2.4(c) depicts s-channel.

Single top quark is dominantly produced via the interaction of a b quark and a virtual W boson in t-channel, as shown in Fig 2.4(a). In this process, a spectator quark q is produced along with the top quark, and transforms into a different flavor (q') via interacting with the W boson. This process accounts for $\sim 70\%$ of the whole single top quark production cross section. This process contributes differently to the total cross section σ_{tq}, depending on whether single top quark or anti-quark is produced (see Section 2.4.4). The cross sections of both processes have been theoretically calculated up to NNLO precision. The fiducial and total cross section have been experimentally measured at center-of-mass energy of 8 TeV at ATLAS. [14] The most recent results measured at 13 TeV [15] are:

$$\sigma_{tq} = 156 \pm 5\ (\text{stat}) \pm 27\ (\text{syst}) \pm 3\ (\text{lumi})\ \text{pb}\,, \tag{2.21}$$

$$\sigma_{\bar{t}q} = 91 \pm 4\ (\text{stat}) \pm 18\ (\text{syst}) \pm 2\ (\text{lumi})\ \text{pb}\,. \tag{2.22}$$

The second largest cross-section for electroweak top quark production is the associated tW production. As shown in the Feynman diagram in Fig 2.4(b), the top quark is produced with a real W boson in the final state in this process. The process when both the top quark and the W boson decay leptonically has been first discovered by the ATLAS collaboration. [16] The measured total cross section for this process including both top and anti-top quark production is: [17]

$$\sigma_{tW} = 94 \pm 10\ (\text{stat})^{+28}_{-22}\ (\text{syst}) \pm 2\ (\text{lumi})\ \text{pb}\,. \tag{2.23}$$

The least dominant process of single top quark production occurs via the interaction of a top quark with a virtual W boson in s-channel, as shown in Fig 2.4(c). The discovery of this process has been reported by the ATLAS collaboration at 8 TeV in [18], and the theoretical predictions for the cross sections of top and anti-top quark are: [19]

$$\sigma^{t}_{\text{s-ch}}\ (m_t = 173\text{GeV},\ \sqrt{s} = 14\text{TeV}) = 7.93 \pm 0.14\ (\text{scale})\ ^{+0.31}_{-0.28}\ (\text{PDF})\ \text{pb}\,, \tag{2.24}$$

$$\sigma^{\bar{t}}_{\text{s-ch}}\ (m_{\bar{t}} = 173\text{GeV},\ \sqrt{s} = 14\text{TeV}) = 3.99 \pm 0.05\ (\text{scale})\ ^{+0.14}_{-0.21}\ (\text{PDF})\ \text{pb}\,. \tag{2.25}$$

2.4.3 Rare top quark processes

With increasing luminosity and collision energy at the LHC, it is possible to probe rare top quark processes with larger data. In general, the rare top quark processes involve a top quark pair production associated with an eletroweak gauge boson ($V = Z, W$) or a Higgs boson, or with even lower cross-section,

involve a single top quark production associated with a Z, W or Higgs boson. In the following, several rare processes are explained.

- $t\bar{t}V$ ($V = Z, W$): The predicted NLO production cross-section of this process at $\sqrt{s} = 13$ TeV is lower than 1 pb, and the observed result at the ATLAS collaboration is [20] :

$$\sigma_{t\bar{t}Z} = 0.9 \pm 0.3 \text{ pb}, \tag{2.26}$$

$$\sigma_{t\bar{t}W} = 1.5 \pm 0.8 \text{ pb}. \tag{2.27}$$

- $t\bar{t}H$: This process involves a Higgs boson, which provides a direct probe for the Yukawa-coupling constant. The observed production cross-section of this process is recently published by the ALTAS [21] and CMS [22] collaborations, and the result from the ATLAS is:

$$\sigma_{t\bar{t}H} = 670 \pm 90 \text{ (stat)} \, ^{+110}_{-100} \text{ (syst) fb}. \tag{2.28}$$

- tZq: In this process, a single top quark is produced along with a Z boson, and it has a very similar final state depending on the decay channel as that of the tHq process which is the focus of this thesis. Therefore, it is one of the main backgrounds to our signal, but at the same time, a study of tZq production can be regarded as a precursor to a tHq analysis. Its production cross section and more details are given in Section 2.4.4.

- tHq: The process where a single top quark is produced in association with a Higgs boson is the signal process in this thesis. It is of particular interest as not only it provides a direct probe for the magnitude of the Yukawa coupling constant, but also its sign, which $t\bar{t}H$ has not sensitivity to. A search for this process in pp collisions at $\sqrt{s} = 13$ TeV is recently published by the CMS collaboration. [23] More details are given in Section 2.4.5.

2.4.4 Single top quark production associated with a Z boson

As mentioned in Section 2.4.4, a study of the single top quark production in association with a Z boson can be seen as a precursor to a tHq analysis. The process occurs via the electroweak interaction through the t-channel diagram. The Z boson radiates either from any quarks that are associated with a virtual W boson or from the W boson. Thus, it is available to probe for the coupling between a top quark and a Z boson as well as a W boson and a Z boson (WWZ). The LO Feynman diagrams of tZq productions are illustrated in Fig 2.5 in the 4-flavor scheme[17]. The $tllq$ process which does not necessarily contain a Z boson has been taken into account, since it leads to the same final state and it could potentially add interference effects to the tZq production.

The cross section of a process which involves a top quark is different from the one which involves an anti-top quark. The former occurs through:

$$u + b \rightarrow d + t + Z, \tag{2.29}$$

or

$$\bar{d} + b \rightarrow \bar{u} + t + Z, \tag{2.30}$$

[17] In the 4-flavor scheme (4FS), all quarks except the bottom and top quark are assumed to be massless. In the 5-flavor scheme (5FS), only the top quark is regarded as massive.

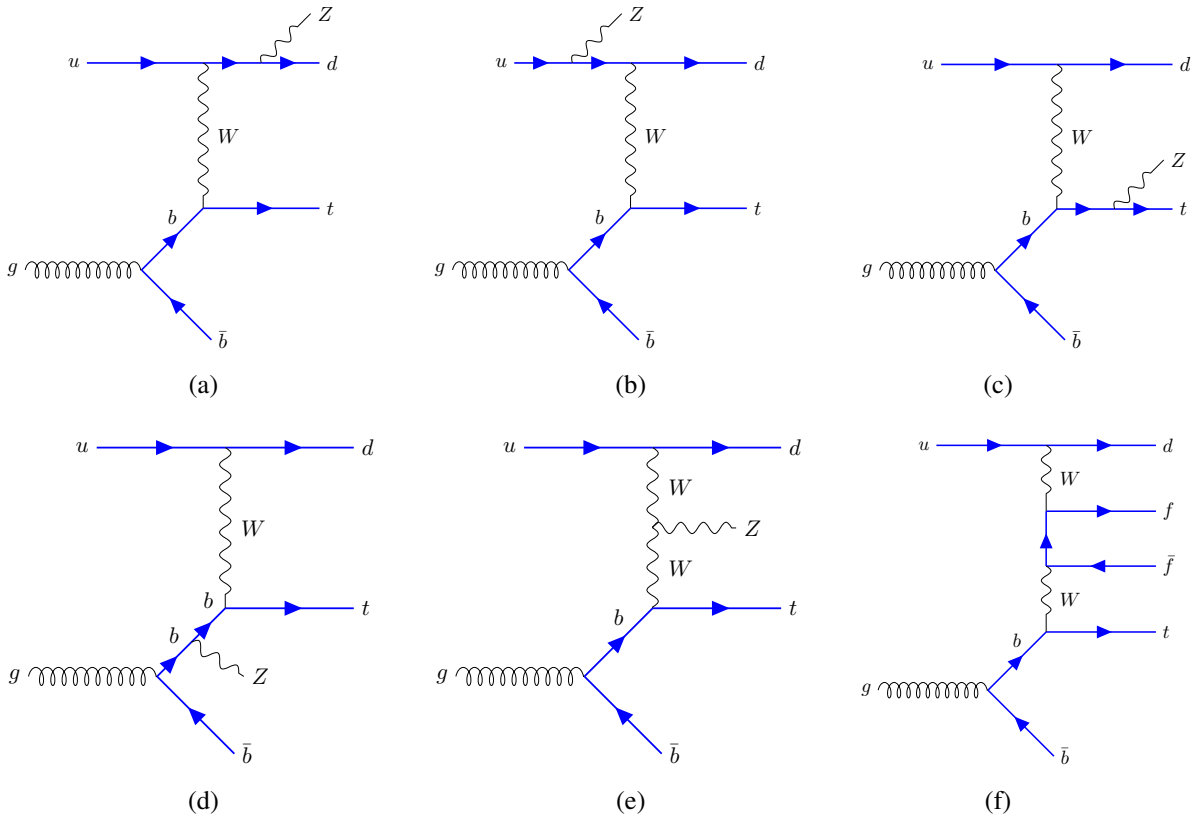


Figure 2.5: Possible LO Feynman diagrams of the tZq production in the 4-flavor scheme. From diagram 2.5(a) to diagram 2.5(e) show the processes which involve a Z boson. Diagram 2.5(f) represents the $tllq$ process which does not necessarily involve a Z boson.

whereas the process which involves an anti-top quark is produced through:

$$d + \bar{b} \rightarrow u + \bar{t} + Z\,, \tag{2.31}$$

or

$$\bar{u} + \bar{b} \rightarrow \bar{d} + \bar{t} + Z\,, \tag{2.32}$$

Here, the anti-quarks in above equations are sea-quarks[18]. Whether a top quark or an anti-top quark is produced depends on whether the spectator quark is an u- (or a $\bar{d}$-) quark or a d- (or an $\bar{u}$-) quark. Since the proton contains one u-quark and two d-quarks, the cross section ratio of the tZq and the $\bar{t}Zq$ production should be roughly 2. Assuming that the Z boson and the top quark decay leptonically, the theoretical prediction of the cross-section in the 5-flavor scheme of total tZq (including both top quark and anti-top quark) production is: [24]

$$\sigma^{\text{theo.}}_{tZq} = 102 \text{ fb } {}^{+5.2\%}_{-1.3\%} \text{ (scale)} \pm 1.0\% \text{ (PDF)}\,. \tag{2.33}$$

The tZq ($tllq$) process was firstly observed by the CMS collaboration, in the channel where the Z boson

[18] Hadrons contain, along with the valence quarks that contribute to their quantum numbers, virtual quark–antiquark pairs known as sea quarks. Sea quarks form when a gluon of the hadron's color field splits.

decays leptonically. The measured cross-section is: [25]

$$\sigma_{tllq}^{\text{CMS}} = 111 \pm 13\ (\text{stat})\ ^{+11}_{-9}\ (\text{syst})\ \text{fb}\,, \tag{2.34}$$

and the cross section measured by the ATLAS collaboration is: [24]

$$\sigma_{tllq}^{\text{ATLAS}} = 97 \pm 13\ (\text{stat}) \pm 7\ (\text{syst})\ \text{fb}\,. \tag{2.35}$$

2.4.5 Single top quark production associated with a Higgs boson

Single top production in association with a Higgs boson is the main process of interest for this thesis. In this thesis, the Higgs boson refers to the neutral Higgs boson that has been discovered in 2012 by the ATLAS and CMS collaborations at the LHC at CERN with a mass of $m_{\text{H}} = 125.18 \pm 0.16\ \text{GeV}/c^2$ [2]. Since the discovery of the Higgs boson by the CMS and ATLAS collaborations at the LHC, many researches have been carried out to measure the Higgs couplings to bosons (g_{HVV}) and fermions (y_{f}) in high precision. In addition to measuring the absolute value of the couplings, a search for potential relative phases of the couplings as well as their Lorentz structure[19] is of significance to reveal any deviations from the SM expectations.

Recently, the Yukawa coupling of a Higgs boson to a top quark, as well as the Higgs boson decaying into bottom quark pairs ($H \to b\bar{b}$) have been observed in the production of a Higgs boson with a top quark pair ($t\bar{t}H$) by [21, 22, 27]. However, the production rate of $t\bar{t}H$ is sensitive only to the magnitude of the top quark-Higgs boson coupling (y_t) and not to its sign. On the other hand, the processes involving loop interactions, such as the production associated with a Z boson or a Higgs boson via gg fusion or Higgs decays into photon pairs do have sensitivity to the sign of y_t. The production of a Higgs boson in association with a single top quark and Higgs decays into two tau leptons is of main interest in this thesis. Based on the measurements to date, a positive sign for the coupling is currently more favored. [28, 29]

In the process where a top quark is produced along with a Higgs boson, the Higgs boson can be radiated either from the top quark or the W boson. Depending on which the baseline process is, there can be two distinct processes: tHq and tWH. The two most important LO t-channel Feynman diagrams for tHq and one of the five LO diagrams for tWH are shown in Fig 2.6. The interference between these diagrams which involve distinct coupling constants (tHq and tWH) makes the production cross section particularly sensitive to the magnitude in addition to the relative sign and the phase of the couplings. The maximally destructive interference leads to a small cross section of approximately 71, 16, and 2.9 fb for the t-channel, tWH and s-channel respectively at $\sqrt{s} = 13$ TeV. In our analysis, only the tHq process has been considered due to the small cross section of tWH. The overall measurements of the production cross section of several processes by the ATLAS collaboration are illustrated in Fig 2.7.

[19] A Lorentz structure on a manifold V is defined as a reduction of the bundle of linear frames to a subbundle of frames invariant under the homogeneous Lorentz group. [26]

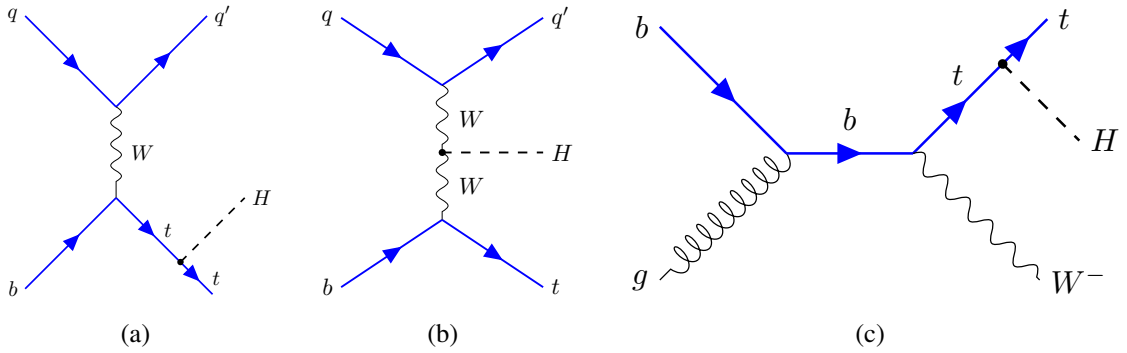


Figure 2.6: Example Feynman diagrams for the tHq process. Diagram 2.6(a) and 2.6(b) shows the t-channel of tHq production, where the Higgs boson couples either to the top quark or to the W boson, respectively. Diagram 2.6(c) represents the tWH production, where the Higgs boson couples to the top quark.

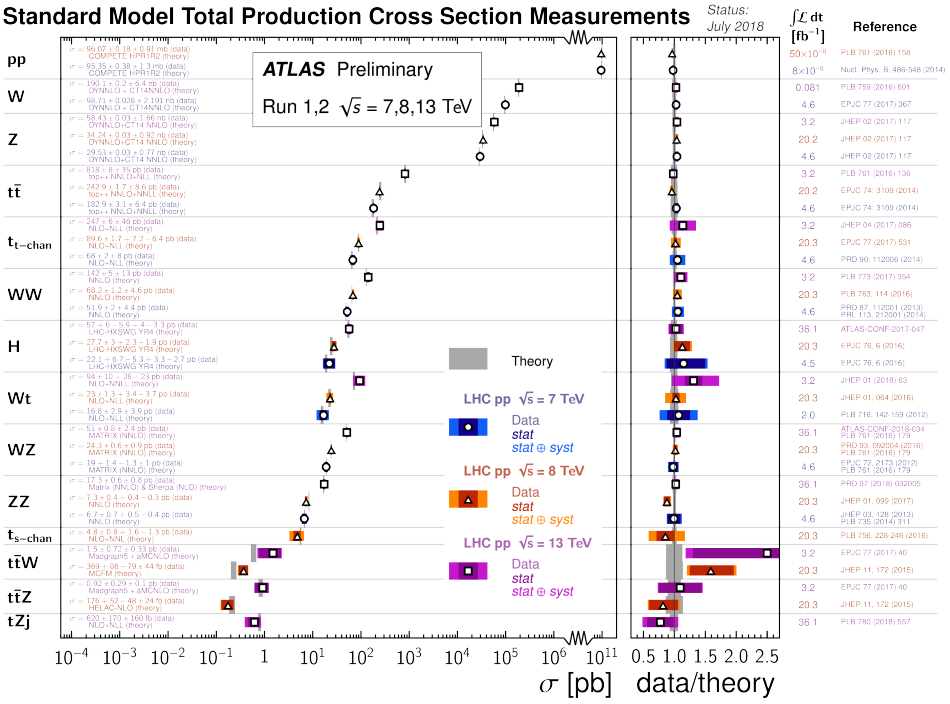


Figure 2.7: An overview of total production cross section measurements of several process in the SM. The right plot shows the cross section measurements in picobarn (pb). The left plot illustrates the ratio of experimental observations to the theoretical expectation. [30]

CHAPTER 3

The LHC and the ATLAS experiment

The data used in this thesis is generated at the Large Hadron Collider (LHC) and collected at the ATLAS detector located at CERN . In this chapter, a general overview of the LHC and the ATLAS detector will be given, and at last, the details of physical object reconstruction in ATLAS will be explained.

3.1 The Large Hadron Collider

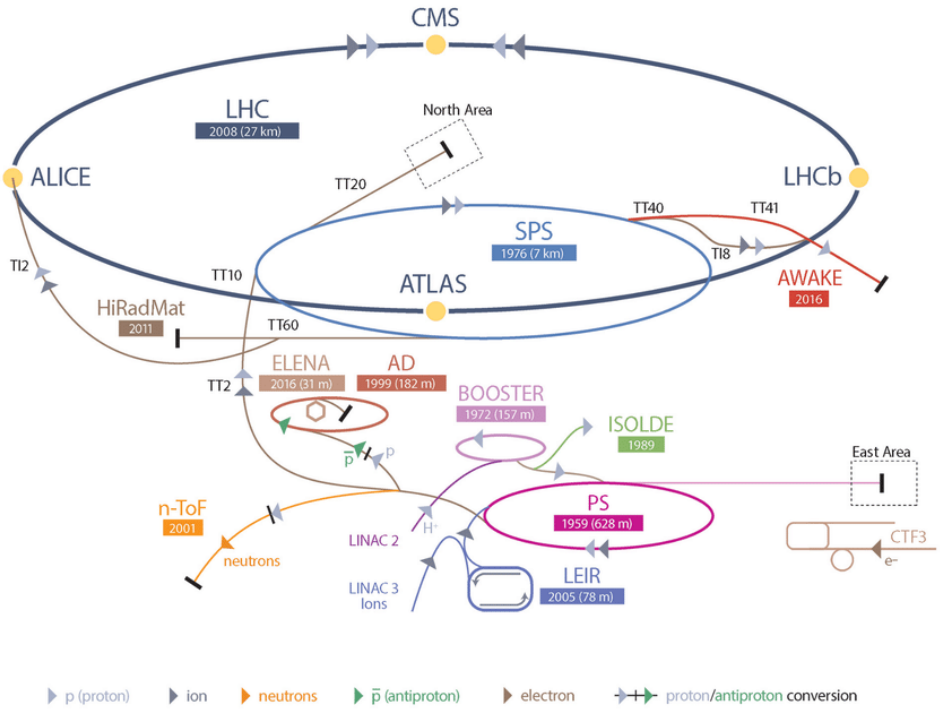


Figure 3.1: A schematic overview of the CERN accelerator complex. The four interaction points marked with yellow color around the LHC ring represent the places where four major experiments are located. [31]

The CERN is a European research organization that operates the largest particle physics laboratory in the world. Since 1954 when it was established, numerous experiments have been constructed at

CERN by international collaborations. [32] The Large Hadron Collider (LHC) is a circular proton-proton accelerator with a circumference of 27 km, and it is located in underground at a depth of 100m, at the border of Switzerland and France, Geneva. Its design is descended from the former the Large Electron Positron (LEP) collider. An overview of the LHC accelerator complex is illustrated in Fig 3.1. More details regarding the design of LHC can be found in [33].

The first operation of the LHC was carried in 2008. The LHC is not under operation everyday for the whole year, but only within certain data-taking periods with shutdowns in between. The data was collected in two separated operational runs so far, Run1 and Run2, from 2009 to 2013 and from 2015 to 2018, respectively. The center-of-mass energies up to 7 and 8 TeV were reached during Run1 with an instantaneous luminosity of $3.0 \times 10^{34} \mathrm{cm}^{-2}\mathrm{s}^{-1}$, and a center-of-mass energy of 13 TeV was reached during Run2 with a total integrated delivered[1] luminosity of 156 fb^{-1} and of 139 fb^{-1} available for physical analyses. Run3 is planned be started in May 2021 with a center-of-mass of 14 TeV. [35] Fig 3.2 illustrates the integrated luminosity achieved during the time between 2015 and 2018.

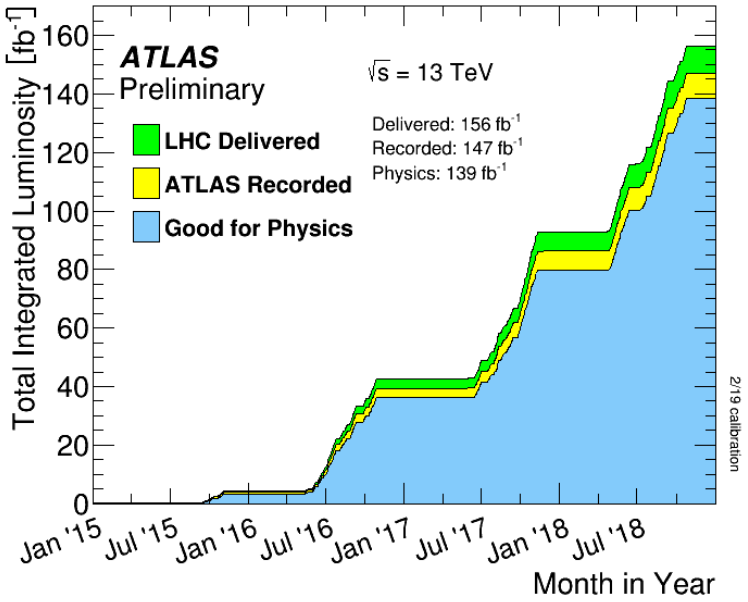


Figure 3.2: Integrated luminosity for different data-taking periods during 2015-2018. The total luminosity delivered by the LHC and recorded by ATLAS are marked in green and yellow, respectively. The luminosity that is considered good for physics is marked in blue. [36]

In order to achieve such high energies, a chain of several accelerators has been built. The protons are obtained by the ionization of hydrogen atoms in an electric field. The proton sources are then injected to a pre-accelerator called linear accelerator (LINAC2) up to the energy of 50 MeV, and accelerated further by a series of circular rings: the Proton Synchroton Booster, the Proton Synchroton and the Super Proton Synchroton, up to the energy of 50 MeV, 1.4 GeV, 25 GeV, and 450 GeV respectivly. The protons are first grouped into bunches of 1×10^{11} protons in the Proton Synchroton Booster, and are injected to the LHC ring once they have reached the energy of 450 GeV. In the LHC, a total of 2808 proton bunches

[1] Delivered luminosity refers to the luminosity delivered to the detectors by the LHC. In the ideal case, the amount of luminosity recorded should be the same as the amount delivered. However, the detectors are unable to take data in some cases, either because their data acquisition chain is busy or because one or more of the detector subsystems is temporarily unavailable. The recorded luminosity includes only the luminosity actually logged by the dectectors. [34]

are accelerated in two separate beam pipes, where one of the proton bunches circulate in the clockwise direction, and the other one in the counter-clockwise direction. In order to ensure the particle beams to travel in a circular path, about 1300 dipole magnets with a magnetic filed of 8.4 T are employed. Additionally, approximately 400 quadrupole magnets are used to focus the beam. Since the proton beams are not perfectly collimated, about 5000 magnets with higher orders, such as sextupoles and octapoles, are used for the further correction of the beams focusing. Liquid helium is particularly used to cool down the temperature of superconducting material of the magnets to 1.9 K for the least resistance losses. [35]

The LHC is designed to collide at four different interaction points, where a specialized detector is built at each colliding point to detect created particles. The location of these four detectors are shown in Fig 3.1. There are four experiments under conduction: A Toroidal LHC ApparatuS (ATLAS) [37], Compact Muon Soleniod (CMS) [38], A Large Ion Collider Experiment (ALICE) [39] and Large Hadron Collider beauty (LHCb) [40]. The first two are for general purposes, which include the SM-based searches as well as BSM theories. The LHCb experiment is focused on b-physics, and searches for charge-parity (CP) violation in the decays of b-hadrons in the SM. ALICE is a heavy-ion detector which is specified to detect quark-gluon plasma[2].

3.2 The ATLAS detector

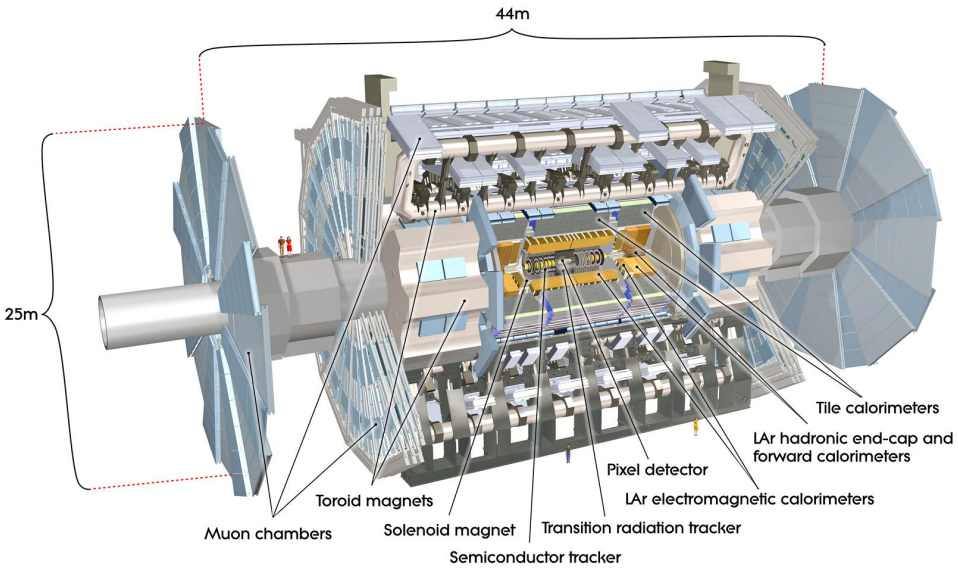


Figure 3.3: A schematic overview of the ATLAS detector. [37]

The ATLAS detector is a general purpose detector, which is designed to detect and identify the particles from the collisions. It is a 46 m long and 25 m high detector with a total weight of 7000 t. The ATLAS detector consists of several layers of sub-detectors, where each of them has been designed to perform a specific function for the detection of particles. An overview of these sub-components can be seen in Fig 3.3. There are mainly four components: the Inner Detector (ID), the calorimeter system, which

[2] The quark-gluon plasma is a state of matter thought to have formed just after the big bang.

is comprised of Electromagnetic Calorimeter (ECAL) and Hadron Calorimeter (HCAL), the Muon Spectrometer (MS) and the magnet system. [37]

The coordinate system of the ATLAS detector can be defined with right-handed cartesian coordinates: the x-axis points towards the center of the LHC ring; the y-axis is defined to be perpendicular to the x-axis; and the z-axis is the direction of the LHC beam pipe. More details regarding the kinematic observables are described in details in Section. 2.3.3.

3.2.1 Inner detector

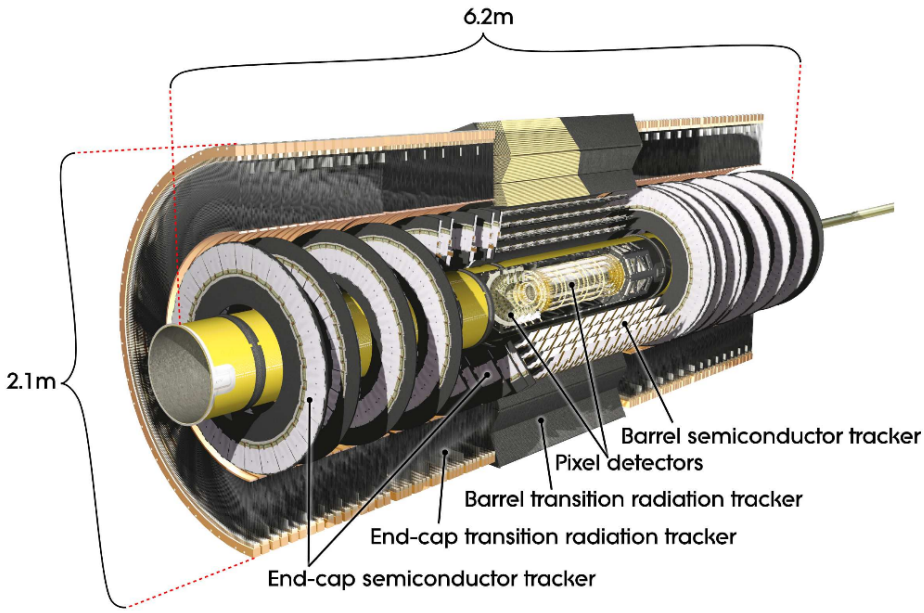


Figure 3.4: A schematic overview of the inner detector of the ATLAS detector. [37]

The role of the Innder Detector (ID) is to reconstruct the primary and secondary vertices[3] and to measure the momentum, direction, and charge of electrically charged particles passing through the detector. It consists of three different cylindrical systems of sensors all placed in a magnetic field along the beam axis: Pixel Detector, Semiconductor Tracker (SCT) and Transition Radiation Tracker (TRT). A schematic view of this system is shown in Fig 3.4. The tracking scope of the ID is limited to $|\eta| = 2.5$. At each side of the ID, a so-called "end-cap" detector is built in order to detect the particles with higher η. [37] The central solenoid magnet system is surrounding the ID to bend trajectories of charged particles with a magnetic field of 2 T.

Pixel detector

The main purpose of the pixel detector is to precisely reconstruct the primary and secondary vertices of the collisions. It is a silicon detector which has four layers of silicon pixel sensors with high granularity, and it is placed in the most central part of the ID with distance of 33 mm from the beam line. [33] The pixel detector is built with 80 million pixels with the pixel size of $50 \times 400\ \mu m^2$, and it has 1744 modules

[3] The primary vertex is where the hard scattering, i,e, the initial interaction takes place; and the secondary vertex refers to the place where the decays of initially interacted particles takes place.

with approximately 46000 readout channels per module. There is an additionally layer which is inserted to the innermost of the pixel detector, called Insertable B-layer (IBL). It has been inserted for Run 2 in order to improve the reconstruction of impact parameters by providing an additional measurement point closer to the interaction point. [41]

Semiconductor tracker

The semiconductor tracker (SCT) is located outside of the ID with a distance of 300 mm from the beam pipe. A total area of 60 m^2 silicon system consists of four layers of cylindrical silicon sensor modules in the central (or barrel) region and nine annular (ring-shaped) disks in each end-cap. There are 4088 two-sided individual sensor modules in the SCT, of which one half is located in the barrel region and the rest in each end-cap. Over 6 million readout strips are implanted in the SCT, with a length of 80 μm for each. The readout strips can measure the position of charged particles with a resolution of 17 μm for each layer in the transverse plane. The SCT covers the radial region with hermetic azimuthal angle out to $|\eta| = 2.5$. [42]

Transition radiation tracker

The outermost component of the ID is the TRT, which covers the largest part of the ID with a total volume of 12 m^3. It has 50000 straw-tubes in the barrel region and five times more of straws in the end-caps. Each straw is filled with a gas mixture depending on their location, and is functioning as an individual drift chamber. The straws in the barrel region are 4 mm wide and 144 cm long, and contain a gold-plated tungsten wire with a diameter of 0.03 mm in the center. The straws in the end-caps, on the other hand, have a length of 39 cm. [43]

A highly negative voltage is applied to the straw wall, leading to an ionization of the gas mixture within each straw when particles pass through it. An avalanche can be induced when the primary electrons are accelerated towards the anode, and thus it leaves a detectable current signal. In addition to the track reconstruction, the TRT is also capable of the particle identification by creating Transition Radiation (TR)[4]. For example, an electron and a hadron such as a pion can be distinguished by the amount of the TR they deposit to the detector. [44]

3.2.2 Calorimeters

The calorimeters are built for identification of highly energetic particles by measuring their energy deposition. There are two different calorimeters in the ATLAS: the electromagnetic calorimeter (ECAL), which is responsible for the detection of electrons and photons, and the hadron calorimeter (HCAL), which measures the particle showers induced by hardons.

The electromagnetic calorimeter

The electromagnetic calorimeter is used to measure the energy deposition of particles that are produced via the electromagnetic interaction. These particles, which are typically electrons and photons, deposit their energies through a so-called *particle shower* in the ECAL. The electrons lose their energy by emitting a photon when they are passing through the material of the ECAL, while photons deposit their energies by eletron-positron pair production. The alternating successions of these processes lead to

[4] Transition radiation (TR) refers to the radiation of an electrically charged particle emits when it passing through the boundary between two media with different dielectric constant. [37]

particle showers and continue until the particles are stopped by losing all of their energies in the matter. The radiation length, denoted as X_0, describes the mean distance that the particles travel until their energies shrink to $1/e$ of their initial energies. Since the atomic number Z, is inverse proportional to the X_0, the material with a higher Z is usually preferred for the minimization of the size of the ECAL. [37, 44]

A pre-sampling detector[5] is located in the inner side of the ECAL ($|\eta| < 1.8$) to avoid leakage of information regarding the energy loss of particles before entering the main calorimeter. The ECAL is a sampling calorimeter, which consists of alternating layers of an active material, which records the energy deposit, and a passive absorbing material, where the particles are absorbed. In ATLAS, liquid argon (LAr) and lead (Pb) are used as the active material and the passive absorber. The alternating layers are designed in an accordion-like shape in order to extend the coverage range of the detector up to $|\eta| < 3.2$. The barrel region and two end-cap regions cover $|\eta| < 1.5$ and $1.4 < |\eta| < 3.2$, respectively. The ECAL is subdivided into three different regions depending on their granularities: the first layer which has the highest granularity of $\eta - \phi$ resolution of 0.003×0.1 to differentiate photons and neutral pions; the second layer with lower granularity due to its priority in the energy measurements; the third layer with coarser cells for the separation between electromagnetic and hadronic showers. [37, 44]

The hadron calorimeter

The hadron calorimeter is used to measure the energy deposited by particles that interact mainly via the strong interaction and/or electromagnetic interaction, typically hadrons like neutrons and pions. Similar to electrons and photons, the hadrons also induce particle cascades called hadronic showers by interacting with the nucleus of the calorimeter material and leading to ionization and/or secondary hadrons. In general, the hadronic showers have electromagnetic shower components as well, such as neutral pions generate electromagnetic component of the shower, i.e. $\pi^0 \rightarrow \gamma\gamma$. In addition, the hadronic showers have longer transverse shower sizes than the electromagnetic showers, thus the HCAL is built within a larger volume.

The HCAL is comprised of mainly three sub-components with different choices of active and passive materials: the tile calorimeter, the hadronic end-cap calorimeter (HEC) and forward calorimeter (FCAL). In the tile calorimeter, plastic scintillators are used as active material and iron as passive material. The alternating layers of these materials are designed to have a staggered shape to prevent any leakage of the energy deposition. The tile calorimeter can be subdivided into two sections with different coverage of $|\eta|$: the barrel tile region with a range of $|\eta| < 1.0$, and the end-cap tile regions with a range of $1.0 < |\eta| < 1.7$. The HEC covers the range $1.7 < |\eta| < 3.2$, and uses copper as absorbing material and the LAr as active material. The FCAL is located in the forward region which covers the range of $3.1 < |\eta| < 4.9$. It is a highly-densed calorimeter that is composed of three layers with different absorption medium. The first layer is made to allow the eletromagnetic showers, and it uses LAr as its active material and copper as the absorption medium. The last two layers use the same active material LAr, but Tungsten as the passive material, which is very suitable for the measurement of energy deposit from the particles produced with a high rapidity. [37, 44]

3.2.3 Muon spectrometer

Muons are able to pass through the detector without losing a large fraction of their energies in the calorimeter system. Since it is very tricky to measure the accurate transverse momentum of muons only

[5] A pre-sampling detector, or a pre-sampler is composed of thin layers (with an active depth of 11 mm) of liquid argon (Ar) equipped with readout electrodes , but no lead (Pb) absorber.

with the information from the calorimeters, an additional component which is specified to track muons is needed. The MS is made for this purpose and is located at the outermost part of the ATLAS detector to cover a wide range of p_{T} and η.

The MS is surrounded by a system of toroidal magnets, which produces a very strong magnetic field to deflect the muons effectively. It consists of four different types of muon chambers: muon drift tubes (MDTs), cathode-strip chambers (CSCs) (which is a multi-wire proportional chamber), resistive plate chambers (RPCs) and thin gap chambers (TGCs). The MDT and CSC are located in the barrel and in the end-cap regions, respectively. They mainly contribute to the precise measurements of the momentum of the muons and their tracks reconstruction. The tracks are reconstructed by using the information of the locations where the muons are initially bended. The RPCs and TGCs are built in the barrel and in the end-cap regions respectively, and are used as trigger systems with fast time response. [37, 44]

3.2.4 Magnetic system

In the ATLAS detector, a system of superconducting magnets is used to measure the momentum and determine the charge of particles with high precision by bending their trajectories. It is comprised of two different magnet systems: a central solenoid magnet surrounding the ID and a toroidal magnet system which is built as a component of the MS. The central solenoid provides a homogeneous magnetic field of 2T along the beam axis, bending the charged particles in the transverse plane. The energy loss of the particles when passing through the solenoid is corrected by the pre-sampler as mentioned in Section 3.2.2. The toroid magnet system in the MS consists of three toroids, one located in the barrel region and the other two in the end-caps. The toroid system produces a total magnetic field of 4 T perpendicular to the solenoid field, with 0.5 T from the barrel section and 1 T from the two end-cap sections. [37, 44]

3.2.5 Trigger system

As the center-of-energy and the luminosity in the LHC are significantly high to date, storing the data of each event is impossible and unnecessary. The trigger system is introduced to reject events of no interest, and extract only events that we are interested in for physics analysis. The trigger system is divided into two levels: Level-1 (L1) trigger and High-Level trigger (HLT). The L1 trigger is a hardware-based system, while HLT trigger is softward-based. The L1 trigger defines Regions-of-Interests (ROIs) based on the information from calorimeters and muon spectrometer. It is responsible for decreasing the event rate from 40 MHz to 100 kHz, and the decision rate for each event is 2.5 μs. The events which pass the L1 trigger are then filtered in the HLT by a factor of 100, at an average time of 200 ms for each event. The events which pass the HLT are stored and used for different usages. In case of the data analysis, the full data of each event is stored, while only partial event information is recorded for trigger level analysis or detector calibration. [37, 44]

3.3 Physics object reconstruction in the ATLAS

When particles passing through the ATLAS detector, they leave distinctive signatures depending on their particle types within each sub-component of the detector. Fig 3.5 shows some examples of the signatures that various particles leave in the each part of the ATLAS detector. This section describes more details about how the particle objects are reconstructed from the signals in each sub-detector.

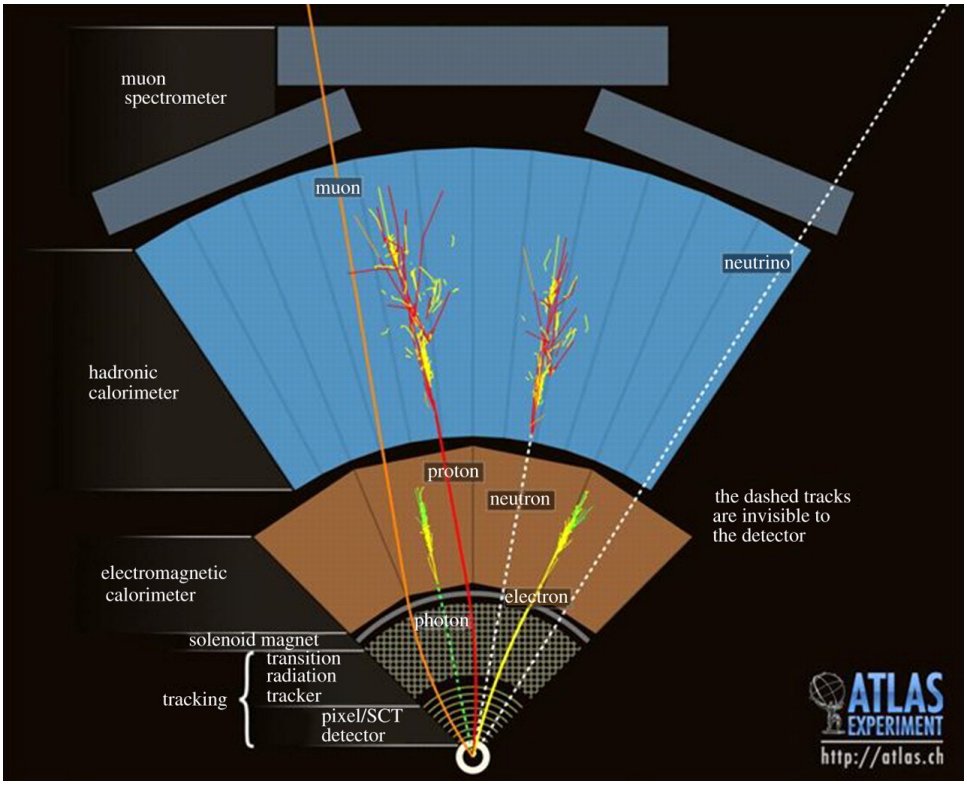


Figure 3.5: A schematic overview of typical signatures that various particles leave in sub-components of the ATLAS detector. [45]

3.3.1 Low-level analysis

In order to fully reconstruct high-energy physics objects, so-called *low-level* objects need be identified in advance. These are associated with the trajectories, vertices where hard interactions or particle decays occur, as well as energy deposition that is registered in the detector.

Tracks

When electrically charged particles pass through the detector, they leave a trajectory of consecutive three dimension space points, called *hits*, in the pixel and SCT detector. The reconstruction of these particle tracks starts from finding the track candidates with at least three hits, using the Kalman filter algorithm. The track candidates must satisfy a set of quality criteria, such as transverse momentum threshold, the number of hits, χ^2 of the track fits, etc. Each of the track candidates is then given a score by being evaluated from these criteria, and is ranked according to its score. The tracks fulfilling the requirements are then extended to the TRT, which extends the track reconstruction procedure up to a radius of 1082 mm. More details can be found in [46].

Vertices and pile-up

The places where the hard scatterings occur are called primary vertices. They are reconstructed using a primary vertexing algorithm which takes track parameters such as d_0 (transverse impact parameter)

and z_0 (longitudinal impact parameter) as input. Several vertex candidates are created after all iterations that passed the quality criteria, and the one with the highest $\sum p_T^2$ is selected among all valid candidates. Apart from the primary vertices, so-called *secondary* vertices are also reconstructed, which refer to the positions where a particle participating in the primary scattering decays into its daughter particles. Other pp inelastic collisions can also take place and give rise to additional vertices. These vertices are called pile-up (explained in Section 2.3.3), and can be quantified either by the number of reconstructed vertices N_{PV}, or the average number of additional interactions per bunch crossing $<\mu>$. Here, μ is defined as:

$$\mu = L_{\text{bunch}} \times \sigma_{\text{inel}} / f_r, \tag{3.1}$$

where L_{bunch} represents the instantaneous luminosity per bunch, σ_{inel} the inelastic cross-section and f_r the revolution frequency of the LHC. The recorded luminosity by the ATLAS detector as a function of $<\mu>$ during Run 2 is shown in Fig 3.6.

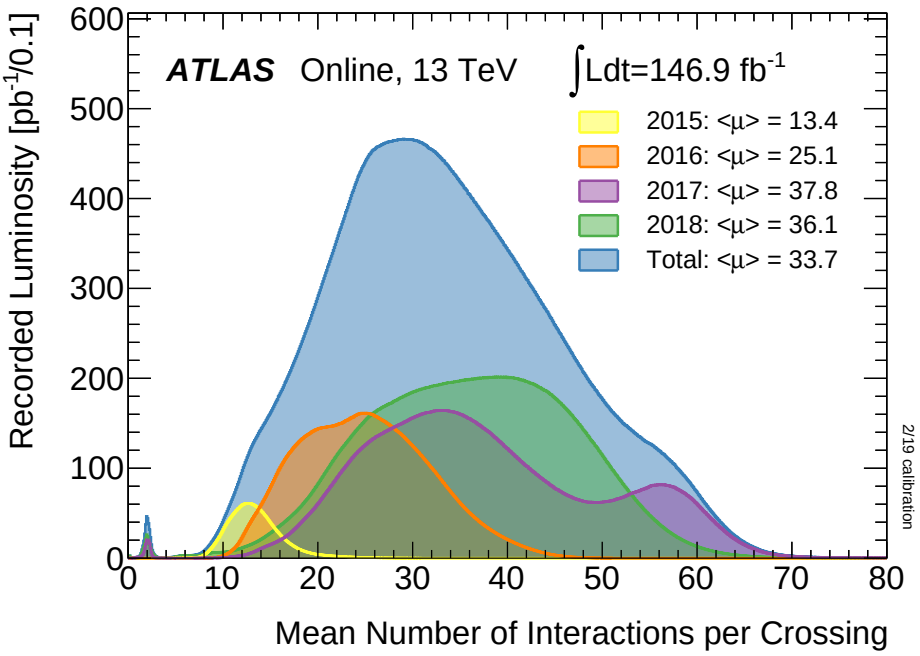


Figure 3.6: The recorded pileup profile by the ATLAS detector during Run 2 (2015-2018). The total mean number of interactions per bunch crossing is $<\mu> \approx 33.7$. [47]

Topological clusters

Topological clusters (topo-clusters) are a combination of individual calorimeter cells which are used to reconstruct the energy of hadrons and jets (also electrons). There are two different clustering algorithms: the topological algorithm and the sliding-window algorithm. The topological algorithm starts from finding a seed cell, whose energy deposition is at least 4 times larger than the expected average noise level. The neighboring cells which have a signal-to-noise (S/N) ratio of at least 2 are additionally included to the seed to form a proto-cluster. This procedure is repeated until no any new cells are found to be added. The sliding-window algorithm is based on summing up all energy depositions of each cell within a box-shaped frame with a fixed height and length. The position of this window is altered so

that the transverse energy inside the frame is at its local maximum. This algorithm is known to be very efficient for energy calibration as well as precise reconstruction of electromagnetic showers and jets from tau-lepton decay. [48]

3.3.2 Electrons

Electrons are expected to leave very clear tracks and clusters in the ID and in the ECAL respectively, due to their electrical charges and smallest mass of a charged particle in the SM. The electron reconstruction requires combined information of the tracks in the ID and the energy deposition in the ECAL. The sliding-window algorithm is used to identify potential electron clusters in a range of $|\eta| < 2.47$ (electrons that are produced in Athena release 21[6] is identified via the supercluster algorithm [49]). These clusters are then matched with ID tracks that are originated from primary vertices, based on the angular distance between the barycenter[7] of the cluster and extrapolated track position into the calorimeter. [50] In order to reduce the probability of fake electron identification (or misidentification), a multivariate technique is exploited. In this algorithm, a likelihood profile is built from the information of tracks and clusters, shower shape, radiation-energy measurement in the TRT, cluster-track matching and etc. Depending on the chosen threshold for the likelihood value, the identified electrons are classified into the main three working points: "Loose", "Medium" and "Tight". In addition to reconstruction and identification, isolation of electrons is also needed for the discrimination between prompt and non-prompt leptons.[8] Two variables, $E_{\mathrm{T}}^{\mathrm{coneXY}}$ which checks for energy deposit within a cone of $\Delta R = 0.\mathrm{XY}$, and $p_{\mathrm{T}}^{\mathrm{varconeXY}}$ which is the sum of p_{T} of all tracks within a cone of $\Delta R = \min(0.\mathrm{XY}, 10\ \mathrm{GeV}/p_{\mathrm{T}})$ are used to ensure whether an electron is isolated.

3.3.3 Muons

Similar to electrons, muons leave charged tracks in the ID, but leave no energy in the calorimeter. Muons lose most of their energy in the MS due to a larger mass than that of electrons as described in Section 3.2.3. The reconstruction of muons is done independently in both sub-detectors, and so-called *combined muons* are formed depending on how the ID and MS information is combined. For the reconstruction of combined muons, a global fitting is applied to find the track candidates in these two sub-detectors. For the identification of muons, a set of variables is used to check the quality of combined tracks, such as momentum measurement in the ID and MS, the normalized χ^2 of the global fit, etc. Four different identification operating points can be defined for muons: "Loose", "Medium", "Tight" and "High-p_{T}". The isolation procedure and criteria apply similar to the ones described for electrons.

3.3.4 Tau leptons

The reconstruction of the tau leptons starts from choosing the tau candidates, which are seeded by jets formed using the anti-k_t algorithm with a distance parameter of 0.4 (see the next section). Three-dimensional clusters of calorimeter cells called topo-clusters (see the next section), calibrated using a

[6] Athena is a software functional for event generation, simulation, digitisation reconstruction and DAOD derivations, based on the common Gaudi framework that is used by ATLAS, LHCb and FCC. The release series 21 is being used for Run 2 simulation and data taking in 2017 and 2018.

[7] The barycenter is defined as the point between two objects where they balance each other. It is the center of mass of a distribution of mass in space where the weighted relative position of the distributed mass sums to zero.

[8] Prompt leptons are referred to as the leptons decaying from heavy bosons, such as W or Z bosons. Non-prompt leptons are the leptons that originate from photon conversions and heavy flavor hadron decays such as b-hadrons, or sometimes even from light hadron decays like c-hadrons.

local hadronic calibration (LC) serve as inputs to the jet algorithm. Jets seeding hadronic tau candidates are additionally required to have $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2$. A tau vertex is chosen as the track candidate vertex with the largest p_{T}-weighted fraction from tracks associated ($\Delta R < 0.2$) with the jet. The tracks are required to pass certain conditions on the number of hits in the tracker, as well as on the shortest distance from the track to the tau vertex in the transverse and longitudinal planes. Afterwards, the tracks that passed these requirements are then associated to core ($0 < \Delta R < 0.2$) and isolation ($0.2 < \Delta R < 0.4$) regions around the tau candidate. The direction (η or ϕ) of the tau candidate is calculated using the vectorial sum of the topo-clusters within $\Delta R < 0.2$ of the seed jet barycenter, exploiting the tau vertex as the origin. A tau-specific energy calibration is then performed on the tau candidate in order to correct the energy deposition measured in the detector to the average value of the energy carried by the measured decay products at the generator level. There are two calibrations: the baseline calibration, and the boosted regression tree (BRT) based calibration.

The tau identification algorithm is dedicated to reject backgrounds from quark- and gluon-initiated jets. The identification process formerly used the Boosted Decision Tree (BDT) based method and now uses a Recurrent Neural Networks (RNN) classifier that provides largely improved jet rejection. The performance of the RNN tau identification algorithm is evaluated on statistically independent test samples of $Z/\gamma^* \to \tau\tau$ for signal and di-jet events (which are selected from data) for background. Four working points with increasing background rejection "Very loose", "Loose", "Medium" and "Tight" are defined to be used by physics analyses are provided, and correspond to different tau identification efficiency values. The input variables to the RNN are adjusted such that the mean value of their distribution for signal samples is constant as a function of pile-up. This makes sure that the efficiency for each working point does not depend strongly on the pile-up conditions. [51]

3.3.5 Jets

Jets are identified as a bunch of collimated particles, which are predominantly $\pi^{\pm}$, $K^{\pm}$ and photons from π^0 decays. They are produced from quarks and gluons via hadronization and fragmentation. There are two different types of jet collections: 1) topo-cluster jets and 2) Particle Flow (PFlow) jets. Depending on the type of the jets, different methods are used for the reconstruction.

Topo-cluster jets

The topo-cluster algorithm and an algorithm [52] called anti-k_t algorithm are used for the reconstruction of topo-cluster jets. Since the jets are expected to deposit large energies in the calorimeters, reconstruction of clusters which present such objects are done prior to the jet reconstruction algorithm. Once the topo-clusters are formed by the topo-cluster algorithm, they are reconstructed as jet objects by the anti-k_t algorithm. The anti-k_t algorithm is the most commonly used jet reconstruction algorithm in ATLAS, which takes four-vectors of individual particles as input and combines them until they meet certain threshold requirements. In this process, a distance between two objects i and j, denoted by d_{ij}, is defined:

$$d_{ij} = \min\left(\frac{1}{p_{\mathrm{T},i}^2}, \frac{1}{p_{\mathrm{T},j}^2}\right)\frac{\Delta R_y^2}{R^2}\,. \tag{3.2}$$

Here, p_{T} is the transverse momentum of each object, and ΔR_y is a quantity that can be calculated from $\Delta R_y^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ where R represents a radius parameter. Once the distance between two objects i and j is calculated, the distance between the object and the beam, defined as $d_{iB} = p_{\mathrm{T},i}^{-2}$, is

calculated. The object pair is merged into one object (jet) if d_{iB} is larger than d_{ij}. This process is repeated until it fulfills $d_{iB} < d_{ij}$.

For the identification of jets, several discriminants, such as the jet-vertex-fraction (JVF), corrected jet-vertex-fraction (corrJVF), R_{p_T} as well as a combined discriminant called jet-vertex-tagger (JVT) are exploited. [53] They are also used for the separation between the jets which originate from the hard scattering and the ones from pile-up events. A discriminant called forward jet-vertex-tagger (fJVT) is used in the forward region ($2.5 < |\eta| < 4.5$) of the detector, while JVT is used in the central region. [54]

Particle flow jets

Unlike the reconstruction technique used for the topo-cluster jets which uses solely the information obtained in the calorimeters, the particle flow jet reconstruction technique combines the information from both the calorimeters and tracking detectors. This technique is usually preferred for optimal event reconstruction, as the tracking system provides a better p_T resolution[9] at low p_T for charged particles. Fig 3.7 shows the p_T resolution as a function of p_T for charged particles in the ECAL and HCAL, as well as in the tracking system.

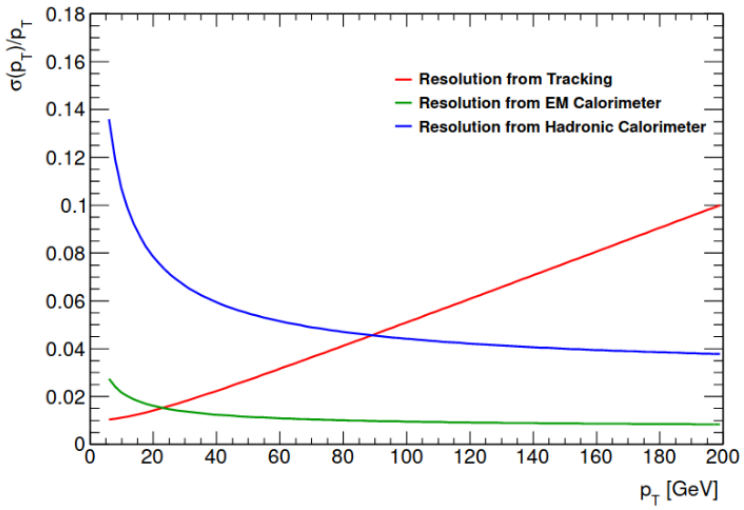


Figure 3.7: The transverse momentum resolution σ_{p_T} as as function of p_T for charged particles. [37]

So-called PFlow-objects are created by combining the tracks and clusters through following four steps: 1) track-cluster matching; 2) charged shower subtraction; 3) cluster annihilation; and 4) neutral particle calibration. In the first step, the reconstructed track is extrapolated to the calorimeteres, providing the impact parameter coordinates of the track to different layers of the calorimeters. Then a topo-cluster nearest to the extrapolated track is chosen using these impact parameters. The information from the calorimeters is used if no topo-cluster is matched to the track. The second step is essential for avoiding double counting a particle, which leaves tracks and the deposition of energy in the calorimeters. In this

[9] The p_T resolution (σ_{p_T}) of a charged particle in the ATLAS tracking system is expressed as: $\frac{\sigma_{p_T}}{p_T} = 0.05\%\ p_T \oplus 1\%$

step, the deposited energy of this particle is removed from the calorimeter cells from the closest[10] topo-cluster (cell-by-cell), until the total amount of the subtracted energy is in consistency with the fraction from the distribution of the calorimeter response (E/p)[11]. [55] In the last steps, the calorimeter clusters with associated tracks are removed during cluster annihilation, and calibration of neutral clusters (the clusters made from the neutral particles) is performed. This is needed for an accurate jet reconstruction, as the neutral deposits remained in the calorimeter after charged shower subtraction are not fully calibrated (the deposited energy in the calorimeter is calibrated at the electromagnetic scale), and need to be re-scaled appropriately.

3.3.6 *b*-jets

The jets that stem from a hadron containing a *b*-quark are called *b*-jets. A *b*-hadron can be produced by the hadronization of a *b*-quark and it has distinctive features that can be used to distinguish the *b*-jets from other jets. The *b*-hadron has a relatively long lifetime and thus travels a distance in order of a few millimeters before it decays, making the identification of the secondary vertex and further decay vertices such as tertiary vertices much easier. It also has a relatively large mass, thus its decay products are likely to be produced with a larger transverse momenta and a wider opening angle. The *b*-jet identification, which is referred to as *b-tagging*, is performed by a multivariate algorithm. [56] In the algorithm, the training samples with *b*-jets as signals and a combination of approximately 90% light-flavoured jets and 10% *c*-jets as backgrounds are used. There are several working points[12] provided, which are defined by a single cut associated with *b*-jet efficiency fulfillment. The information regarding the *b*-tagging efficiency as well as the rejection rate of *c*-jets and light-flavoured jets for different working points can be found in [56].

3.3.7 $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction

As briefly mentioned in Section 2.3.3, the missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) is defined as the negative vector sum of the transverse momenta of all visible particles pertaining to a collision. The $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is done by including all calibrated hard objects which are reconstructed with calorimeter signals in the following order: electrons, photons, hadronically decaying τ leptons, jets, muons (the object passing specific requirements are added to the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation if the signal does not overlap with the objects in the previous step). Apart from these, soft objects originating from soft scattering also need to be considered. These can be reconstructed by taking detector signals such as ID tracks or calorimeter signals that are not involved in any hard object selections stated above. [57]

The $E_{\mathrm{T}}^{\mathrm{miss}}$ components are obtained as below:

$$E_{\mathrm{x(y)}}^{\mathrm{miss}} = E_{\mathrm{x(y)}}^{\mathrm{miss},\,e} + E_{\mathrm{x(y)}}^{\mathrm{miss},\,\gamma} + E_{\mathrm{x(y)}}^{\mathrm{miss},\,\tau} + E_{\mathrm{x(y)}}^{\mathrm{miss,\,jets}} + E_{\mathrm{x(y)}}^{\mathrm{miss},\,\mu} + E_{\mathrm{x(y)}}^{\mathrm{miss,\,soft}} \,, \tag{3.4}$$

[10] The distance to the topo-cluster from deposited energy is given by:

$$\Delta R' = \sqrt{\left(\frac{\Delta\phi}{\sigma_\phi}\right)^2 + \left(\frac{\Delta\eta}{\sigma_\eta}\right)^2}\,, \tag{3.3}$$

where σ_ϕ and σ_η are the width of the cluster with respect to the associated angle in the ATLAS coordinate system.

[11] The calorimeter response (E/p) is defined as the ratio of the energy of a cluster in the calorimeter to the momentum of an associated track.

[12] Working Point (WP) describes the efficiency of tagging. For example, if the tagging is performed at a WP of 70% on jets from a decay of a particle, then it means that 70% of all the jets originating from the decay are tagged.

where $E_{\mathrm{x(y)}}^{\mathrm{miss},\,i}$ refers to the $x(y)$ momentum components of object i with opposite sign. The magnitude as well as the azimuthal angle ϕ^{miss} of $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated as:

$$\begin{aligned} E_{\mathrm{T}}^{\mathrm{miss}} &= \sqrt{(E_{\mathrm{x}}^{\mathrm{miss}})^2 + (E_{\mathrm{y}}^{\mathrm{miss}})^2}\,, \\ \phi^{\mathrm{miss}} &= \tan^{-1}\left(E_{\mathrm{x}}^{\mathrm{miss}}/(E_{\mathrm{y}}^{\mathrm{miss}})\right). \end{aligned} \tag{3.5}$$

CHAPTER 4

Data and Monte Carlo simulated samples

In order to extract signal events that we are interested in from pp collision data recorded by the LHC, a set of reliably simulated samples which reflects our current understanding of particle interactions is needed. Monte Carlo (MC) simulation is used for this purpose and will be described in details in this chapter. By comparing the real data with MC simulated samples, one can have a general idea about the validity of the SM and BSM hypotheses. At first, details about the data and MC samples used for two different analyses, tZq and tHq analyses, are presented. Then a brief overview of how the data samples are collected and how the MC simulation is carried out will be given.

4.1 Datasets

The starting point for both tZq and tHq analyses is the single-top ntuples which are produced from `TOPQ1` derivations. The data and MC ntuple samples are produced using a framework called `AnalysisTop`, which is used by the ATLAS Top Working Group for sample generation. For both tZq and tHq studies version 29 (v29) samples produced with `AnalysisTop-21.2.75` are used. However, in the tZq study, the samples that are used in precedent tZq analyses [58, 59] are in version 28 (v28) which are produced with `AnalysisTop-21.2.58`. These up to v29 ntuple samples contain no truth-level information (see Section 7.2.6 for the definition of truth-level) of the Higgs system.

4.1.1 Data samples

The data samples used for the analyses were recorded at the LHC and collected by the ATLAS detector at the center-of-energy of $\sqrt{s} = 13$ TeV in Run 2, from 2015 to 2018. Only the data samples which were recorded during stable LHC conditions and guaranteed full functionality of the ATLAS detector were selected for this analysis. The data fulfilling these conditions are stored as a set of "Good Run Lists (GRL)" and are evaluated to be in sufficiently good quality for physics analysis. The total amount of these data corresponds to a total integrated luminosity of 139 fb^{-1}. The data collected in individual years are listed in Table 4.1.

4.1.2 Monte Carlo simulated samples

MC simulation plays a key role in high energy physics and is used to generate various physics processes based on stochastic methods. The MC simulation is used to estimate the feasibility of specific measurements, by predicting the event numbers of a signal process as well as contributions from background

Year	Int. Lumi. [fb^{-1}]
2015	3.2
2016	33.0
2017	44.3
2018	58.5
Total	139.0

Table 4.1: A summary of integrated luminosities of the data in each year during Run 2 operation (2015-2018). [58]

processes. It also allows us to investigate the event topology, signatures in the final states, and detector effects. The generation procedure of MC samples can be divided into three steps. First, the hard scattering process[1] is simulated based on the calculation of the matrix element of this interaction and integration of its phase spaces. The parton distribution functions (see Section 2.3.2) are determined using experimental data, and are derived by performing global fits up to NNLO in QCD. Together with the hard interaction, the underlying events are also generated using a general-purpose event generator such as SHERPA [61]. In order to match the PU conditions that are observed in real data samples, both in-time and out-of-time PU are modeled in the MC samples by reweighting. Then the hadronization process (where partons form bound states) of these events as well as parton showering processes are simulated by software such as SHERPA or HERWIG, and POWHEG as a specialized one. Finally, the simulated events are passed to a simulation of the ATLAS detector which provides detector effects by a software called GEANT4 [62]. This step is done either via a full simulation (FS), or via a fast simulation (AFII) which makes use of parameterized parton showers in the calorimeters [63].

The MC samples are modelled based on the common Physics Modelling Group (PMG) recommendations. The MC samples generated for Run 2 are divided into three different categories: labeled as MC16a, MC16d and MC16e, corresponding to the PU conditions during the data taking period in 2015 to 2016, 2017 and 2018 respectively. As a consequence, the MC samples need to be scaled to match the corresponding integrated luminosities in the individual data taking periods, as shown in Table 4.1. All MC samples used for the signal process as well as background processes are listed in Appendix A and an overview of them is shortly given in the following sections.

MC signal samples

Two separate MC signal samples, tZq samples for the tZq analysis and tHq samples for the tHq analysis, are discussed here. The tZq samples are generated at NLO in QCD with MadGraph5_aMC@NLO v2.6.0 [64] and NNPDF30_NLO_AS_0118_NF_4 [58] as the parton distribution function (PDF) set. The renormalization scale μ_R and factorization scale μ_F are set to $\mu_R = \mu_F = (m_t + m_Z)/4 = 66$ GeV [65]. Inelastic collisions were simulated using PYTHIA 8.186 [66] and the parton shower as well as the hadronization were simulated with PYTHIA 8.230 [67]. In order to take the interference effect between the processes which involve a Z boson and those which do not (see Section 2.4.4), the tZq sample is technically including the tl^+l^-q process. The MC simulation is performed using the four-flavor scheme (4FS), in which all quarks are considered to be massless except for the b- and top quark. Afterwards,

[1] The *hard* scattering refers to a 2-to-2 parton scattering with non-vanishing transverse momentum (p_T). The resultant event contains particles that originate from the two outgoing partons (including initial- and final-state radiation) and particles that come from the breakup of the proton and antiproton. The *underlying event* is everything except the two outgoing hard-scattered jets (for more complicated processes like tZq there can be more outgoing hard-scattered jets). [60]

all resulted distributions of kinematics and any physical properties need to be rescaled to the five-flavor scheme (5FS) at NLO with the theoretical prediction of $\sigma = 102^{+5}_{-2}$ fb for the tZq cross-section. [58] The top quark mass in the event generators was set to $m_t = 172.5$ GeV. The tHq samples are generated in two different channels: the $H \to b\bar{b}$ and multilepton (ML) channel which includes $H \to WW/ZZ/\tau\tau$. They are generated at NLO with `MadGraph5_aMC@NLO` v2.6.0 and their parton showers are simulated with `PYTHIA` 8.230. However, due to low statistics of the current samples as described in Section 4.2.2, the production of new tHq ML samples with increased number of generated events has been proposed.

MC background samples

The backgrounds to the signal (here, we consider the tZq process as signal) is estimated by using simulated samples that contain at least two leptons and at least two jets [58]. These samples include the production of $t\bar{t}$, $t\bar{t}H$, $t\bar{t}V$ ($V = Z, W$), tW, tWZ, diboson (WW, WZ or ZZ), and Z+jets. For the tHq analysis, the tZq production samples are included as one background source. The generators used for the most important background processes described in Section 5.2 are summarized in Table 4.2.

Process	MC generator	Parton showering
$t\bar{t}$	`PowhegBox` v2	`Pythia` 8.230
$t\bar{t}H$	`PowhegBox` v2	`Pythia` 8.230
$t\bar{t}V$	`MadGraph5_aMC@NLO` v2.3.3	`Pythia` 8.210
tW	`PowhegBox` v2	`Pythia` 8.230
tWZ	`MadGraph5_aMC@NLO` v2.3.3	`Pythia` 8.212
Diboson	`Sherpa` v2.2.2	`Sherpa`
Z + jets	`Sherpa` v2.2.1	`Sherpa`
tZq (for tHq analysis)	`MadGraph5_aMC@NLO` v2.6.0	`PYTHIA` 8.230

Table 4.2: An overview of all background samples and the MC generators used to simulate them. Note that, in the tHq analysis, the tZq process samples are included as a background source. More details about the samples can be found in [58].

4.2 Monte Carlo reweighting

4.2.1 Event-by-event and luminosity reweighting

Simulation of events taking place in the real world is not always perfect, and may not correctly reproduce the data-taking conditions. Furthermore, the selection efficiency may alter for different physical objects. Therefore, the simulated samples need to be reweighted to be used for accurately estimating the real data samples. There are two different reweighting procedures: event-by-event reweighting and luminosity reweighting. In event-by-event reweighting, every single MC simulated event needs to be multiplied with an event-by-event factor w_{event} (w stands for *weight*) which can be written as below:

$$w_{\text{event}} = w_{\text{MC}} \times w_{\text{pile-up}} \times w_{\text{lepton}} \times w_{\text{JVT}} \times w_{\text{trigger}} \times w_{\text{b-tagging}} \,. \tag{4.1}$$

The total weight for each event is the product of following individual weights:

- w_{MC} : this factor stands for MC event weight. The MC generators generate events which carry a certain weight w_{MC} which differs from 1 due to the way how the MC integration is implemented

in the generator. Therefore, w_{MC} needs to be applied for each event such that the sum of these weights from all generated events ($\sum w_{\text{MC}}$) gives the correctly estimated number of events for the simulated sample.

- $w_{\text{pile-up}}$: this corrects for the simulated pile-up conditions to match the pile-up conditions in real data taking runs.
- w_{lepton} : this term represents the efficiency of reconstruction, identification and isolation of leptons. It is applied by multiplying the individual weights corresponding to each of the selected leptons for each event. In the tHq analysis, this weight is further multiplied by w_{tau} which takes into account those efficiencies of hadronically decaying τ leptons (no τ lepton is selected in the tZq signal regions).
- w_{JVT} : this reduces backgrounds based on primary vertex when applying a cut on the Jet Vertex Tagger (JVT) in high pile-up environment.
- w_{trigger} : this term represent the mismatch related to trigger selection efficiencies in data and MC.
- $w_{\text{b-tagging}}$: this correction factor accounts for b-tagging efficiencies in data and MC.

Apart from the event-by-event reweighting by multiplying a factor of w_{event} to each MC event, luminosity reweighting also needs to be applied in order for the simulation to reach the correctly estimated integrated luminosity as well as a cross-section similar to the one in real data. Furthermore, as MC simulation demands for a high computing power, the total number of events that can be generated is limited. Therefore, the MC simulation is carried out up to a certain number of events[2] according to production cross-section of the process, and a reweighting procedure to compensate this difference between MC and data is crucially done. The luminosity reweighting is done by multiplying a factor of w_{lumi}, called luminosity weight, to each w_{event}:

$$w_{\text{total}} = w_{\text{event}} \times w_{\text{lumi}} \,, \tag{4.2}$$

where w_{lumi} is defined as:

$$w_{\text{lumi}} = \frac{\sigma \mathcal{L}_w}{N_w} \times w_c \,. \tag{4.3}$$

Here, σ is the cross-section of the process in unit of [pb], N_w is the number of weighted events in the original MC samples, and $\mathcal{L}_w$ is the integrated (nominal) luminosity that the events need to be normalized to, of which the default is set to 1000 pb^{-1} in our framework, such that the integrated luminosity of the dataset to be in [fb^{-1}]. The last term w_c stands for campaign weight which is determined as the fraction from individual luminosity to the total luminosity:

$$w_c = \frac{\mathcal{L}_c}{\sum_i \mathcal{L}_i} \,, \tag{4.4}$$

where $\mathcal{L}_c$ is the luminosity associated with the campaign that the weight is generated for. Note that all these mentioned weights are only applied to MC simulated events.

4.2.2 Monte Carlo weight problem

As shortly mentioned in Section 4.1.2, the current tHq ML samples are suffering from severe low statistics. This issue is thought to originate from a high percentage of negative weights. Here, the weight refers to

[2] The MC samples are actually produced at an arbitrary luminosity and is scaled to the luminosity of the real data (in some cases there can be more MC simulated events than we expect from the data).

nominal weight. In the framework `tHqLoop`, there are two different types of weights, nominal weight (w_{nominal}) and raw nominal weight ($w_{\text{nominal raw}}$). The w_{nominal} is defined as below:

$$\begin{aligned} w_{\text{nominal}} &= \Pi_i w_i \times w_{\text{lumi}} \\ &= \Pi_i w_i \times \left[\frac{\sigma \mathcal{L}_w}{\sum_i w_{\text{MC}_i}} \times w_c \right] \end{aligned} \tag{4.5}$$

Here, the w_{lumi} is differently defined from the previous section, where the luminosity is normalized using the sum of MC generator event weights $\sum_i w_{\text{MC}_i}$ [68, 69]. The $w_{\text{nominal raw}}$ is defined the same as w_{nominal} but without w_c. As you can see from Fig 4.1, the number of negative weights in the nominal weight distribution accounts for $\sim 35\%$. In order to search for the origin of this large fraction of negative weights, each weight distribution of w_{MC}, $w_{\text{pile-up}}$, w_{lepton}, w_{JVT} and $w_{\text{b-tagging}}$ has been investigated as shown in Fig 4.2 (w_{trigger} is not shown). As you can see, it is the w_{MC} which contributes to the negative value of weights mostly. Due to this large population of the negative weights, the *tHq* sample which has low statistics shows significant fluctuations in some kinematic distributions. This indicates that the statistics is not sufficient enough for Multi-Variative Analysis (MVA), background suppression, fake studies, etc. Therefore, to resolve this problem, a new production with increased statistics is proposed and is in progress.

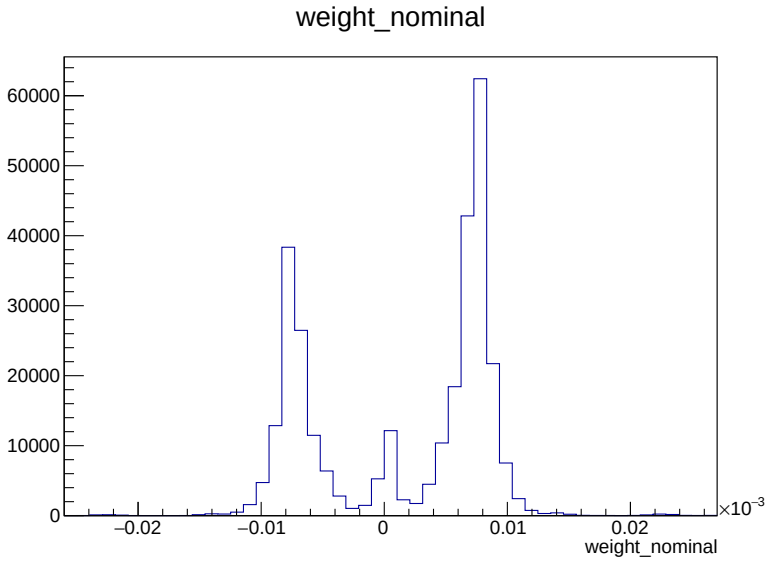


Figure 4.1: The distribution of nominal weight (w_{nominal}) or event weight (w_{total}) of the v29 *tZq* sample (all campaigns).

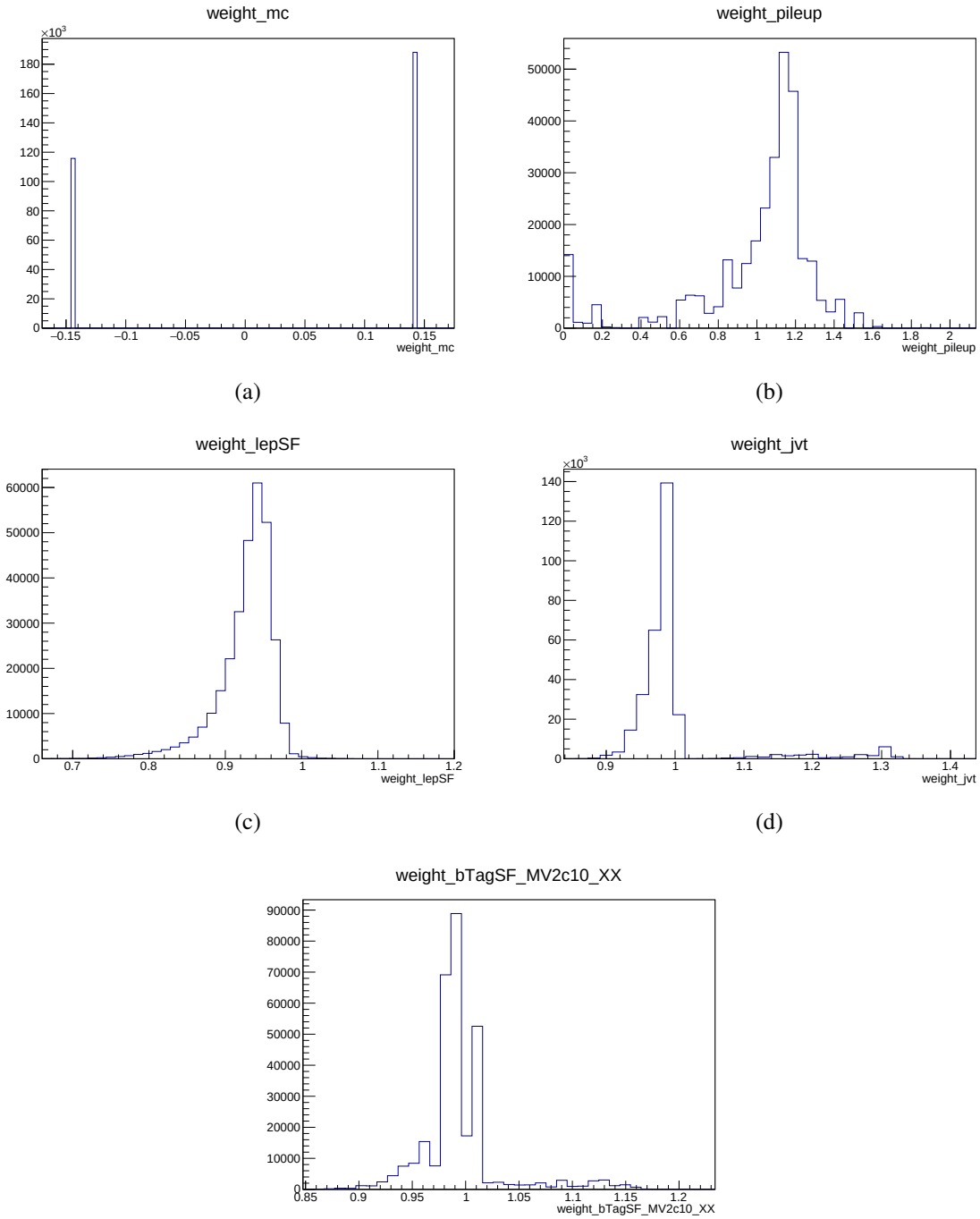


Figure 4.2: The distribution of w_{MC}, $w_{\text{pile-up}}$, w_{lepton}, w_{JVT} and $w_{\text{b-tagging}}$ the v29 *tZq* sample (all campaigns).

CHAPTER 5

tZq event selection and tHqLoop framework validation

In this chapter, an overview of the tZq analysis as a precursor to the tHq analysis is presented. As mentioned in Section 2.4.4, the tZq process can result in a similar final state as the tHq process, constituting a background source to the tHq process. Therefore, studying the event topology of the tZq final state can provide us a better understanding of the tHq events as well as an efficient method for the tZq background suppression. This chapter mainly describes how the validation of a framework named `tHqLoop` (for the tHq study) has been carried out. First, the strategy used for selecting signal regions will be given in details. Then the results obtained with `tHqLoop` will be compared to those from precedent tZq studies [58, 59] for the cross-check.

5.1 tZq trilepton final state

In the tZq process, a single top quark associated with a Z boson as well as a forward[1] (spectator) quark are produced. Since the direct detection of the Z boson and top quark is not possible in parton level due to their relatively short lifetimes, they can be observed only through the reconstruction of their decay products in the final state. The top quark decays almost exclusively ($\sim$ 99%) into a W boson and a b-quark, and the W boson can decay either leptonically resulting in a charged lepton and its corresponding neutrino ($W \to l\nu_l$), or hadronically ($W \to q\bar{q}'$). The Z boson can decay into a charged opposite sign, same flavor (OSSF) lepton[2] pair, or a pair of neturinos, or hadronically. An overview of corresponding branching ratios for the previously mentioned individual decay channels as well as the probability for all possible combined final states are shown in Fig 5.1.

In this analysis, the trilepton final state of the tZq process is investigated as it has similar signal signatures as the ML final state of the tHq process (see Section 6.1 for details about the tHq ML final state). In the trilepton tZq final state, the Z boson and W boson both decay leptonically and result in exactly three leptons where at least one OSSF pair must be included. Although the branching ratio of this channel is very small ($\sim$ 2%), this channel is expected to have a better signal-to-background ratio and thus a higher chance to be observed. An example of a Feynman diagram for the event topology of this

[1] A forward jet generally refers to a jet which is produced in the forward direction of the detector. However, in this thesis, the jet which is defined as the forward jet tends to be more forward than the others, and it is not necessarily in the forward region.

[2] In previous tZq searches [58, 59] the leptons refer to only charged light leptons, i.e. electrons, muons and those which originate from leptonically decaying τ leptons. However, hadronically decaying τ leptons are also taken into account later on in this thesis (tHq analysis).

channel is shown in Fig 5.2. The *tZq* trilepton final state is expected to be composed of exactly three leptons, one neutrino, one *b*-jet, and one forward jet.

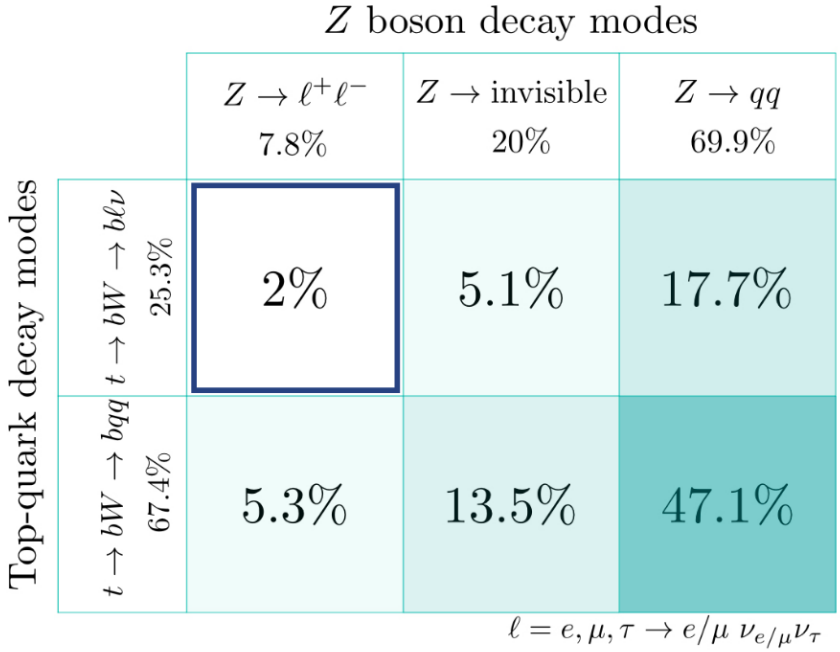


Figure 5.1: An overview of the branching ratios for the *Z* boson and top quark decay modes as well as the probability for all possible combined final states. The combined channel marked with dark blue box, where both the *Z* boson and top quark decay leptonically, gives the *tZq* trilepton final state. [70]

5.2 Source of background

In this section, the main sources of background to the *tZq* trilepton channel are discussed. The *tZq* process shares most of its background sources with the *tHq* signal process. The backgrounds can be divided into two different categories: the processes which involve prompt or real leptons, and the ones involve at least one non-prompt or fake lepton. Here, prompt leptons are referred to the leptons originating from the main collisions as direct products of the particular decay of a heavy particle such as a *W*, *Z* boson or a τ lepton. On the other hand, non-prompt leptons originate from the decays of *b*- or *c*-hadrons or photon conversions. Fake leptons are mostly jets that are misidentified as electrons and muons. The background processes such as diboson, $t\bar{t}V$ and $t\bar{t}H$ are the main backgrounds involving prompt leptons, while $t\bar{t}$ and *Z*+jets are the main background sources including non-prompt or fake leptons. In the following sections, the main background sources are explained in details.

5.2.1 Diboson

Diboson process is one of the largest backgrounds for the *tZq* final state. This process involves two bosons, *WZ* or *ZZ*, where both bosons decay leptonically and enter the signal selection criteria. The LO Feyman diagrams for both processes are illustrated in Fig 5.3. As can be seen from Fig 5.3(a), a *WZ* process results in a final state which consists of three real leptons with leptonically decaying *W* and

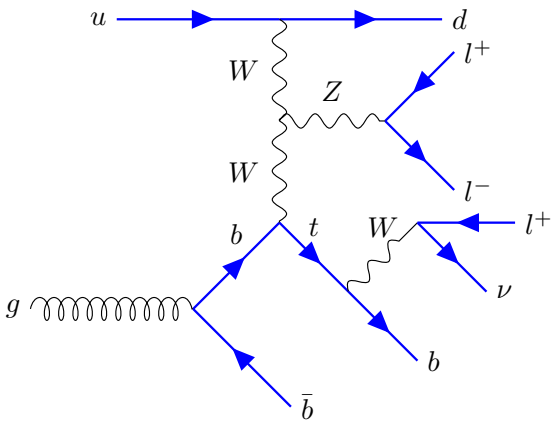


Figure 5.2: One example of a LO Feynman diagram of the tZq processs, with leptonically decaying Z boson and top quark.

Z bosons, mimicking the final state of the tZq trilepton and tHq multilepton channels. One example Feynman diagram of the ZZ process is shown in Fig 5.3(b). The ZZ process takes place when both Z bosons decay leptonically and one of the leptons is either not reconstructed or fails to pass the object selection criteria.

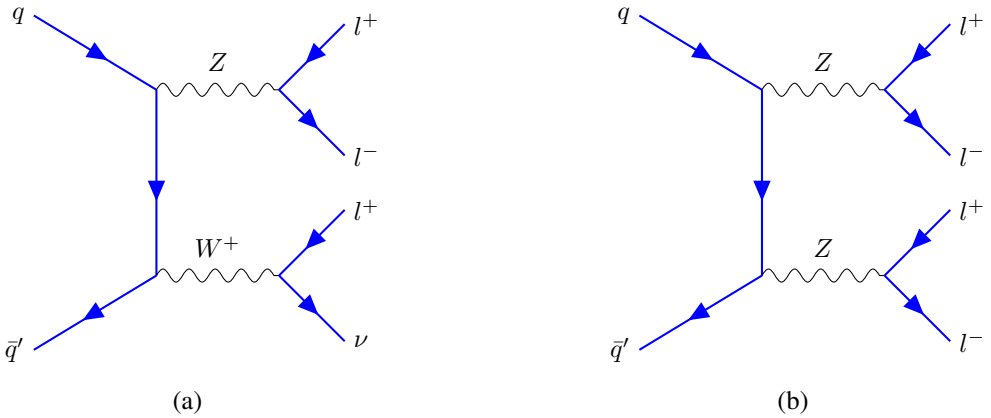


Figure 5.3: Two example Feynman diagrams for the diboson process. Diagram 5.3(a) shows the WZ process, and diagram 5.3(b) corresponds to the ZZ process.

5.2.2 $t\bar{t}V$ and $t\bar{t}H$

$t\bar{t}V$ and $t\bar{t}H$ processes are those in which a top quark pair is produced in association with a heavy boson such as a Z or a Higgs boson. One example of this background event topology is displayed in Fig 5.4. A $t\bar{t}Z$ event can have a similar signature since it involves a Z boson and a pair of top quarks, resulting in three leptons with additional two jets including one b-jet. The unidentified b-jet at the reconstruction level can lead to the event passing the signal selection criteria.

The $t\bar{t}H$ process has a similar event topology to the $t\bar{t}V$ process with only $V = Z$ replaced by $V = H$.

Given the fact that the production cross-section of a pair of top quarks and a Higgs boson as well as the branching ratio for leptonical decay channels of the Higgs boson are small, a lower contribution than from the $t\bar{t}V$ process can be expected.

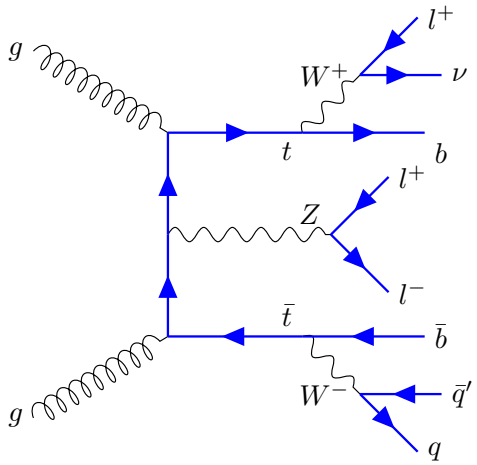


Figure 5.4: One example Feynman diagram for the $t\bar{t}V$ ($V = Z$) processs. The $t\bar{t}H$ process has a similar event topology where only the Z boson is replaced by a Higgs boson.

5.2.3 tWZ

The production of a single top quark in association with a W and a Z boson has not been experimentally observed so far. However, it is expected to be produced with a cross-section of approximately 12.2 pb [71]. One example of a NLO Feynman diagram of the tWZ process where both the W and Z bosons originate from a top quark is shown Fig 5.5. The leptonically decaying Z and W bosons can originate either from hard-scattering or a top quark decay and result in three leptons in the final state but with a high jet multiplicity.

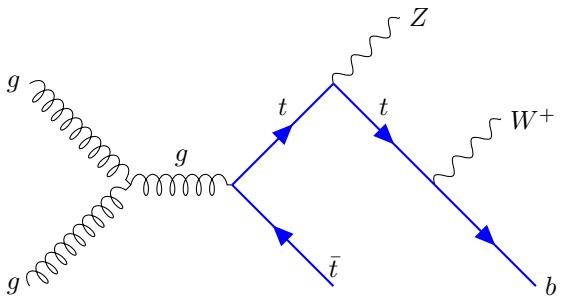


Figure 5.5: One example of a NLO Feynman diagram of the tWZ processs.

5.2.4 $t\bar{t}$

The production of $t\bar{t}$ events is one of the main background sources to the signal event. As can be seen in Fig 2.7, the production cross-section of a pair of top quarks is listed as one the largest production

cross-sections following those of the processes involving a W or a Z boson. One possible LO Feynman diagram of the $t\bar{t}$ process is displayed in Fig 5.6. The pair of top quarks decay leptonically, leading to the similar signature to the signal event when one of the leptons originate from a b-hadron decay.

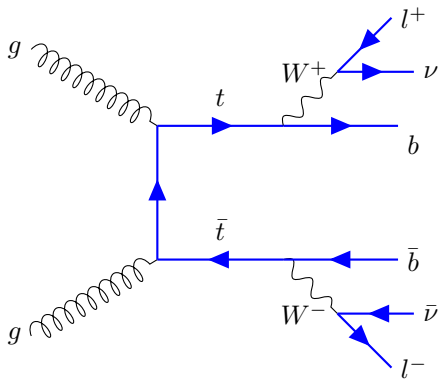


Figure 5.6: One example of a LO Feynman diagram of the $t\bar{t}$ processs, where both W bosons decay leptonically.

5.2.5 Z+jets

Z+jets is considered to be one of the major non-prompt backgrounds. A corresponding event topology is illustrated in Fig 5.7. In this process, the Z boson decays leptonically, while the gluon emitted from the quark decays into a pair of b-quarks. One lepton originates from either of these b-hadrons can mimic the tHq final state.

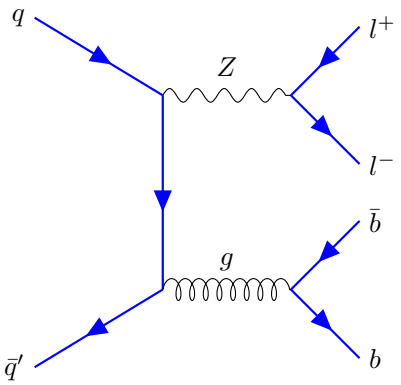


Figure 5.7: One example of a LO Feynman diagram of the Z+jets processs, where the Z boson decays leptonically and a gluon splits into a $b\bar{b}$ pair.

5.3 Event selection

In this section, an overview of the strategy for selecting the signal event, i.e. the tZq trilepton event, as well as the event yields of all processes including the signal and background processes are given. For the development as well as the validation of the framework that will be used for the tHq multilepton channel

analysis, a framework called `tHqLoop` is exploited to apply selection cuts. For the framework validation, a comparison of the obtained event yields with those from a recently published *tZq* research [58] is performed.

5.3.1 Signal regions

The signal regions (SRs) are defined as the regions enriched with the *tZq* events. For the cross-check, we follow the same selection criteria that are stated in precursor *tZq* studies [58, 59]. For the signal events, two orthogonal SRs are defined: 2j1b SR and 3j1b SR. The 2j1b SR refers to the regions where three leptons, one *b*-jet as well as one untagged (forward) jet are selected. The 3j1b SR is also defined in addition, since one additional untagged jet could be resulted in the final state by gluon radiation. This jet is referred to as *radiation* jet. In the 3j1b SR, the same selection criteria is used as in the 2j1b SR except the requirement of two untagged jets, of which the one that gives the highest value of invariant mass $m_{\mathrm{bj_f}}$ between the *b*-jet is selected to be the forward jet. Unlike the 3j1b SR, the forward jet is uniquely defined in the 2j1b SR.

The common selection criteria used for the 3j1b and 2j1b SRs are shown below:

- Exactly three leptons are required (e or μ). The three leptons are sorted by their p_{T} regardless of their flavor into l_1, l_2, and l_3 (l_1 has the highest p_{T}). The three leptons are required to fulfill the transverse momentum conditions $p_{\mathrm{T}}(l_1) > 28$ GeV, $p_{\mathrm{T}}(l_2) > 20$ GeV and $p_{\mathrm{T}}(l_3) > 20$ GeV. All leptons must be reconstructed in the central detector region and thus must satisfy $|\eta| < 2.5$.
- Jets are required to have $p_{\mathrm{T}} > 35$ GeV and $|\eta| < 4.5$, as the untagged jet (or one of the two untagged jets in the 3j1b SR) is likely to be in the forward direction.
- The *b*-jet is required to satisfy $p_{\mathrm{T}} > 35$ GeV and $|\eta| < 2.5$. The 70% working point is used for selecting the *b*-jets, and for *b*-tagging the `MV2c10` algorithm [72] is used.

The followings are specific selection criteria used to define the 2j1b and 3j1b SRs:

- In order to suppress the backgrounds which do not contain a Z boson, at least one OSSF lepton pair associated to the Z boson is required. Additionally, the invariant mass of the OSSF leptons need to fulfill $|m_{ll} - m_{\mathrm{Z}}| < 10$ GeV. In the case where there are more than one OSSF lepton pairs, the pair whose invariant mass is closest to the Z boson mass is selected.
- Exactly one *b*-jet is required in both 2j1b and 3j1b SRs. In 2(3)j1b SR, two (three) jets satisfying the previously mentioned jet condition are required, of which exactly one jet must be *b*-tagged.

The selection cuts applied in the SRs are summarized in Fig 5.8. As can be seen in this figure, six control regions (CRs) are defined in parallel with the SRs. The CRs are the regions designed to be enriched with background events. Similar cuts are applied to CRs as those applied to the SRs, such that they are perpendicular to the SRs, as can be seen in Fig 5.9. Both SRs and CRs are used for the extraction of the signal event yields. The six CRs are the regions designed to address background contributions: CR-2j0b and CR-3j0b for diboson events. For more details about the CRs, see [58].

5.3.2 Signal region yields

The event yields of all simulated processes including the signal and background processes after applying the full selection criteria in the two SRs can be found in Table 5.1. In order to ensure the framework `tHqLoop` is applying the selection cuts properly, the obtained event yields are compared to those in

Common selections			
Exactly 3 leptons (e or μ) with $\lvert\eta\rvert < 2.5$ $p_{\mathrm{T}}(\ell_1) > 28\,\mathrm{GeV}$, $p_{\mathrm{T}}(\ell_2) > 20\,\mathrm{GeV}$, $p_{\mathrm{T}}(\ell_3) > 20\,\mathrm{GeV}$ $p_{\mathrm{T}}(\mathrm{jet}) > 35\,\mathrm{GeV}$			
SR 2j1b	CR diboson 2j0b	CR $t\bar{t}$ 2j1b	CR $t\bar{t}Z$ 3j2b
≥ 1 OSSF pair $\lvert m_{\ell\ell} - m_Z\rvert < 10\,\mathrm{GeV}$ 2 jets, $\lvert\eta\rvert < 4.5$ 1 b-jet, $\lvert\eta\rvert < 2.5$	≥ 1 OSSF pair $\lvert m_{\ell\ell} - m_Z\rvert < 10\,\mathrm{GeV}$ 2 jets, $\lvert\eta\rvert < 4.5$ 0 b-jets	≥ 1 OSDF pair No OSSF pair 2 jets, $\lvert\eta\rvert < 4.5$ 1 b-jet, $\lvert\eta\rvert < 2.5$	≥ 1 OSSF pair $\lvert m_{\ell\ell} - m_Z\rvert < 10\,\mathrm{GeV}$ 3 jets, $\lvert\eta\rvert < 4.5$ 2 b-jets, $\lvert\eta\rvert < 2.5$
SR 3j1b	CR diboson 3j0b	CR $t\bar{t}$ 3j1b	CR $t\bar{t}Z$ 4j2b
≥ 1 OSSF pair $\lvert m_{\ell\ell} - m_Z\rvert < 10\,\mathrm{GeV}$ 3 jets, $\lvert\eta\rvert < 4.5$ 1 b-jet, $\lvert\eta\rvert < 2.5$	≥ 1 OSSF pair $\lvert m_{\ell\ell} - m_Z\rvert < 10\,\mathrm{GeV}$ 3 jets, $\lvert\eta\rvert < 4.5$ 0 b-jets	≥ 1 OSDF pair No OSSF pair 3 jets, $\lvert\eta\rvert < 4.5$ 1 b-jet, $\lvert\eta\rvert < 2.5$	≥ 1 OSSF pair $\lvert m_{\ell\ell} - m_Z\rvert < 10\,\mathrm{GeV}$ 4 jets, $\lvert\eta\rvert < 4.5$ 2 b-jets, $\lvert\eta\rvert < 2.5$

Figure 5.8: An overview of the requirements applied when selecting the tZq events in the signal and control regions. Here, an OSSF pair stands for an opposite-sign same-flavor lepton pair, and an OSDF pair stands for an opposite-sign different-flavor lepton pair. This summarized table is taken from [58].

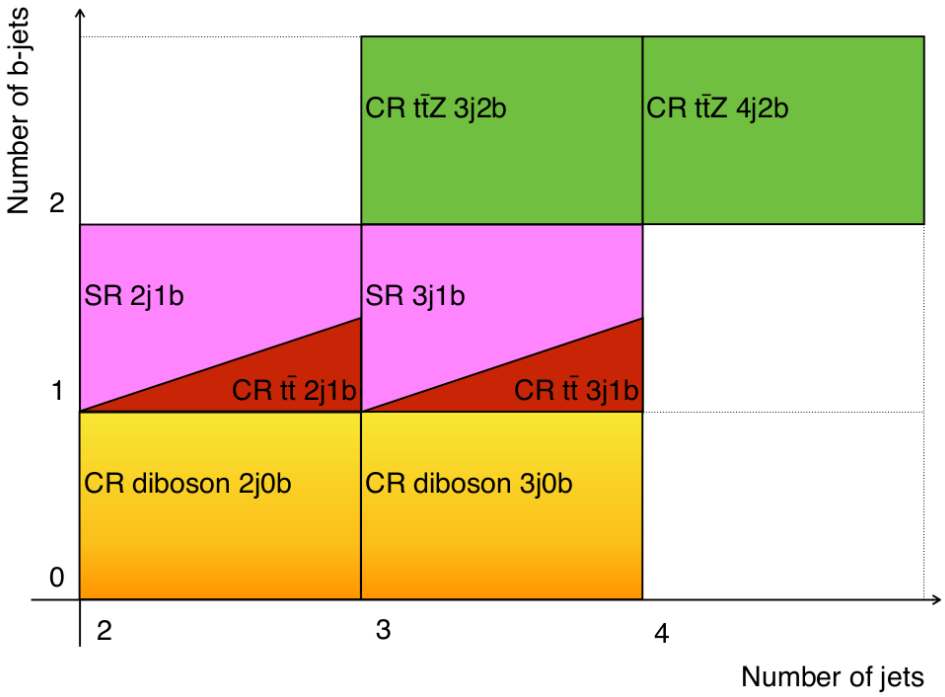


Figure 5.9: A summary of all defined SRs and CRs for the tZq analysis. The SRs and $t\bar{t}$ CRs are different regarding the requirement for lepton pairs: the SRs are required to have one OSSF lepton pair satisfying $|m_{ll} - m_Z| < 10$ GeV, while $t\bar{t}$ CRs are required to have one OSDF lepton pair. This figure is taken from [59].

Fig 5.10 and 5.11, which show the number of predicted and observed event yields for the 2j1b SR and 3j1b SR respectively, taken from [59]. In Fig 5.10 and 5.11, some background samples are treated together as a single background group following the conventions from [58, 59]. The $t\bar{t}$ and tW backgrounds, as well as $t\bar{t}Z$ and tWZ, are combined as they share similar production mechanisms and kinematic properties. The $t\bar{t}W$ and $t\bar{t}H$ contributions are also grouped together, as both are referred to as very small backgrounds. These background groupings are done on events with $m_{\mathrm{T}}(W) > 20$ GeV, and the justification of these groupings can be found in [59]. The diboson contribution is split according to the origin of the associated jets using generator-level information: diboson + heavy flavor (VV + HF) denotes the event where one of the jets contains a b- or a c- hadron; and diboson + light flavor (VV + LF) denotes the case for one of the jets contains lighter quarks.

The obtained event yields with tHqLoop are in a good consistency with those from the previous studies. Here, the event yields of the tZq samples are rescaled to the 5FS[3]. We presume that the discrepancies that exist in the event yields between our and previous studies might come from the differences between two versions of the used ntuples, v29 and v28. Apart from that, the two SRs show similar signal-to-background ratios.

Process	Number of events
$t\bar{t}$V	64.1 ± 0.6
$t\bar{t}$H	1.9 ± 0.0
tWZ	16.8 ± 0.6
Diboson	122.5 ± 1.3
Z + jets	11.1 ± 1.8
$t\bar{t}$	24.9 ± 1.1
tW	1.1 ± 0.4
tZq	80.9 ± 0.6
Total expected	322.8 ± 2.7
Data	336

(a) 2j1b SR

Process	Number of events
$t\bar{t}$V	95.8 ± 0.7
$t\bar{t}$H	2.3 ± 0.1
tWZ	19.7 ± 0.7
Diboson	72.2 ± 0.8
Z + jets	4.4 ± 0.5
$t\bar{t}$	11.7 ± 0.8
tW	0.4 ± 0.3
tZq	47.3 ± 0.5
Total expected	253.7 ± 1.7
Data	263

(b) 3j1b SR

Table 5.1: Numbers of expected and observed (reweighted) event yields in the 2j1b SR (Table 5.1(a)) and 3j1b SR (Table 5.1(b)) broken down by process, obtained by tHqLoop. The uncertainties include only statistical uncertainties. Here, the event yields of the tZq samples are scaled to the 5FS.

Complementary to the event yield tables, so-called *cutflow* plots which represent the efficiency of each applied cut are illustrated in Fig 5.12 for the 2j1b SR and Fig 5.13 for the 3j1b SR. The entry at each bin in these plots corresponds to the number of events of the signal and background processes passing each respective cut described in Fig 5.8. The number of events shown in the cutflow plots are reweighted to the luminosity. As you can see, the requirements on the minimum number of jets, leptons and b-jets, the maximum number of τ leptons as well as the mass constraint on the two leptons that originate from the Z boson cut away quite a large amount of events for most of the backgrounds compared to the signal in both SRs. Especially the required minimum number of (b-) jets as well as leptons is the most efficient cut for the suppression of two major backgrounds Z+jets and $t\bar{t}$.

[3] Scale factor is calculated as a ratio of the SM expectation of the tZq cross-section in the 5FS to that in the 4FS: SF $= \frac{101.7\text{ fb}}{88.5\text{ fb}} \sim 1.15$. This scaling factor has been applied to the event yields of the tZq process [24].

Process	Number of events
$t\bar{t}Z$	62.7 ± 1.4
$t\bar{t}W$	4.3 ± 0.3
$t\bar{t}H$	2.1 ± 0.1
tWZ	17.9 ± 1.0
VV + HF	101.3 ± 1.0
VV + LF	22.7 ± 0.6
Z + jets	10.2 ± 0.0
$t\bar{t}$	23.7 ± 0.1
tW	1.1 ± 0.0
tZq	81.1 ± 0.7
Total expected	327.0 ± 2.3
Data	359

Process	Number of raw events
$t\bar{t}Z$	45 837
$t\bar{t}W$	1513
$t\bar{t}H$	4487
tWZ	2992
VV + HF	39 010
VV + LF	5137
Z + jets	173 703
$t\bar{t}$	172 074
tW	1969
tZq	307 562
Total expected	754 284

Figure 5.10: Numbers of predicted and observed event yields in the 2j1b SR broken down by process, taken from [59]. The left table shows the number of events after luminosity weighting, while the right table shows the raw number of simulated events available. The uncertainties include only statistical uncertainties.

Process	Number of events
$t\bar{t}Z$	96.6 ± 1.9
$t\bar{t}W$	2.2 ± 0.2
$t\bar{t}H$	2.4 ± 0.1
tWZ	20.2 ± 1.1
VV + HF	59.4 ± 0.6
VV + LF	12.3 ± 0.4
Z + jets	4.8 ± 0.0
$t\bar{t}$	11.1 ± 0.0
tW	0.4 ± 0.0
tZq	46.1 ± 0.5
Total expected	255.3 ± 2.4
Data	259

Process	Number of raw events
$t\bar{t}Z$	77 643
$t\bar{t}W$	765
$t\bar{t}H$	4442
tWZ	3601
VV + HF	24 821
VV + LF	3164
Z + jets	77 167
$t\bar{t}$	73 697
tW	736
tZq	208 092
Total expected	474 128

Figure 5.11: Numbers of predicted and observed event yields in the 3j1b SR broken down by process, taken from [59]. The left table shows the number of events after luminosity weighting, while the right table shows the raw number of simulated events available. The uncertainties include only statistical uncertainties.

5.3.3 Signal region plots

For further cross-check, we compare the distributions of lepton-, jet- and reconstructed event-related variables in the two SRs. The distributions of lepton-related quantities are shown in Fig 5.14 to 5.15, and the distributions for other quantities are presented in Appendix B.

[4] In the plots, the place where 'None' is located is supposed to stand for the selected SR name, but is not shown.

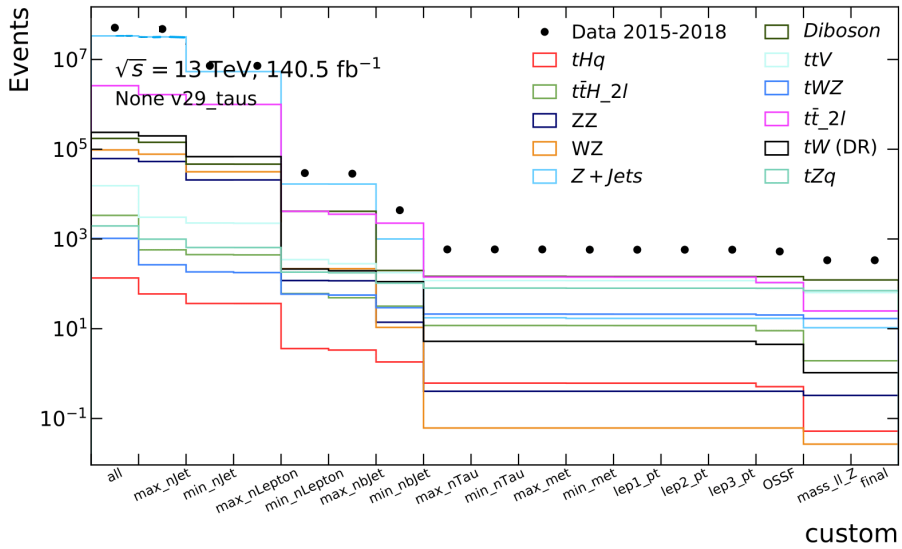


Figure 5.12: The cutflow plot for the measured event yields in log-scale after each SR selection requirement on the signal and background processes in the 2j1b SR.

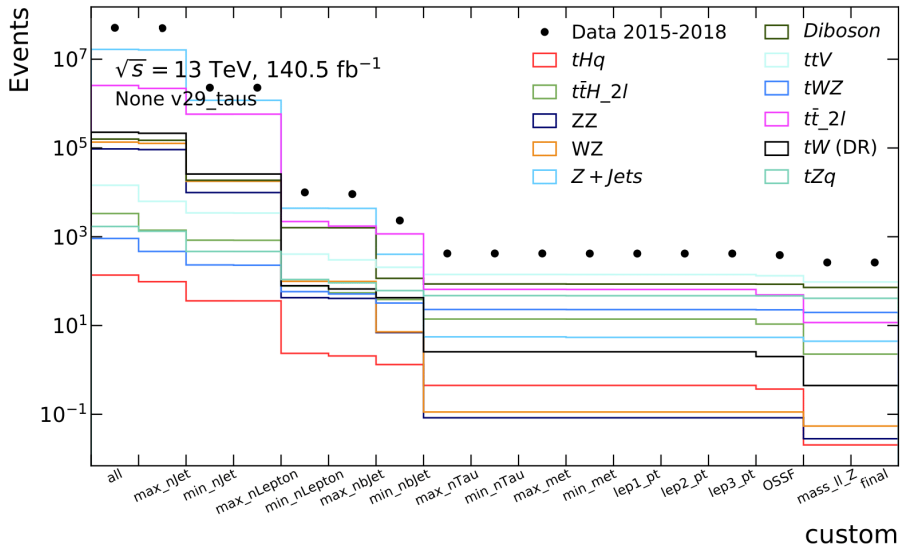


Figure 5.13: The cutflow plot for the measured event yields in log-scale after each SR selection requirement on the signal and background processes in the 3j1b SR.

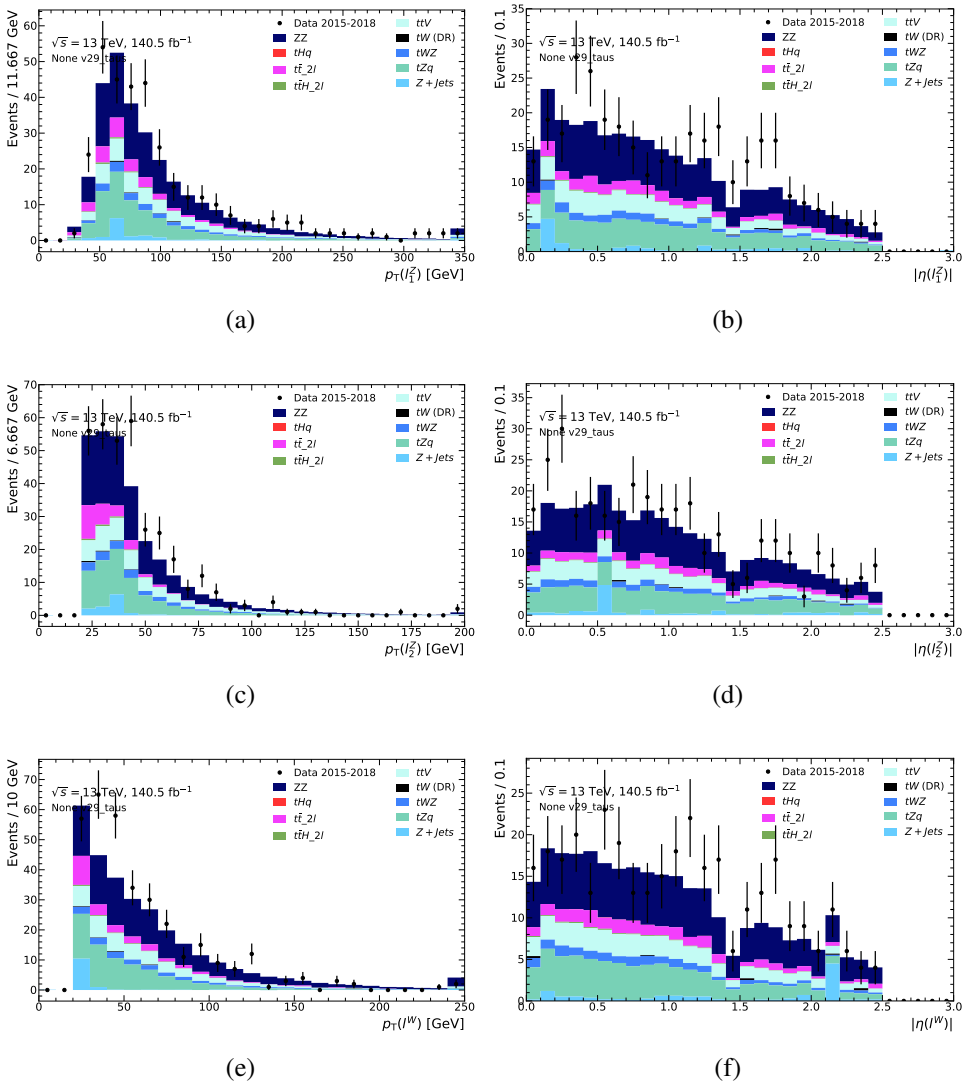


Figure 5.14: Comparison of data and MC predictions for reconstructed lepton-related quantities for events in the 2j1b SR. The uncertainty band includes only statistical uncertainties for the signal and backgrounds[4].

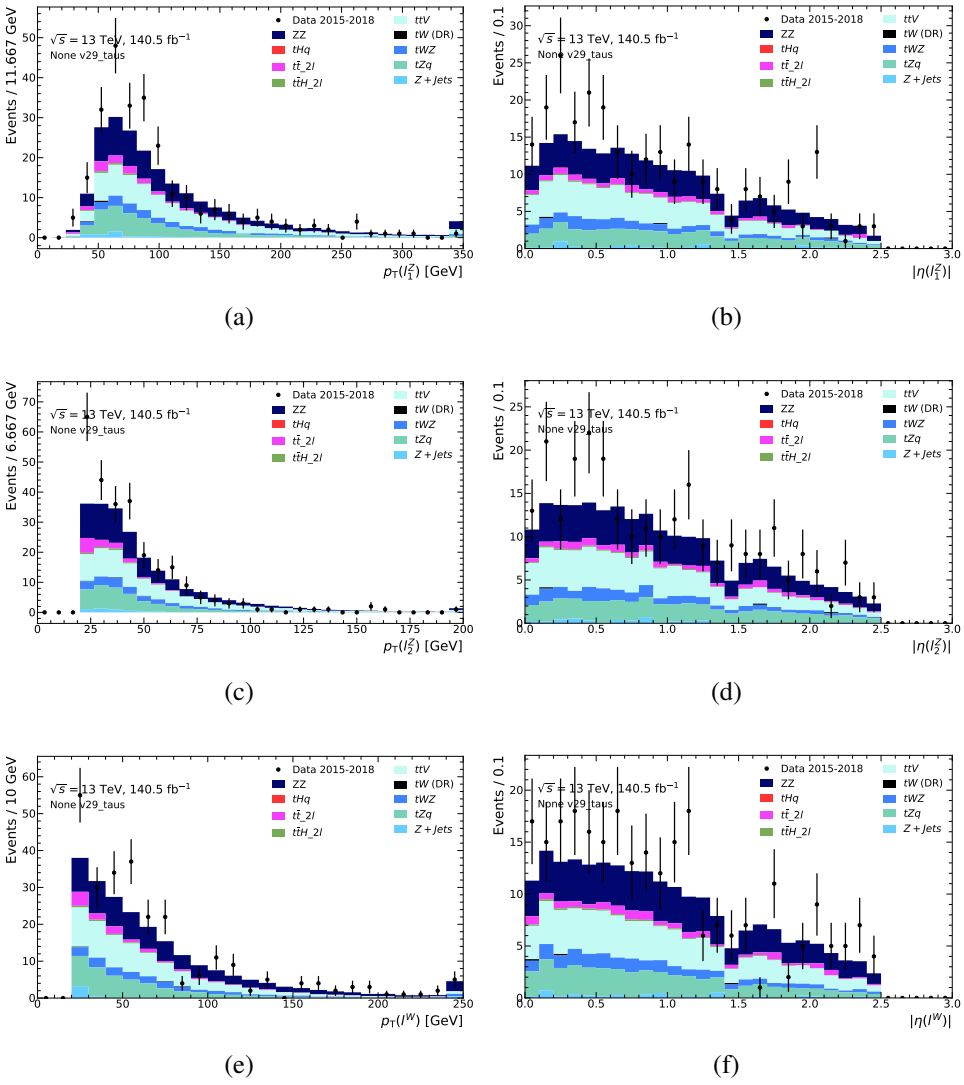


Figure 5.15: Comparison of data and MC predictions for reconstructed lepton-related quantities for events in the 3j1b SR. The uncertainty band includes only statistical uncertainties for the signal and backgrounds.

5.4 Final state reconstruction

As described in Section 5.3.1, the final state of the tZq ML channel consists of exactly three charged leptons, of which two originate from the Z boson and one from the top quark, one b-jet, one forward jet and one neutrino. Reconstructing the mother particles, i.e. the Z boson and top quark is important for separating the signal from backgrounds which do not involve those particles.

First, the Z boson is reconstructed with an opposite-sign, same-flavor (OSSF) lepton pair. If all three leptons in the final state have the same flavor, then the pair which has the invariant mass closest to the Z boson mass is chosen. Then the leptonically decaying W boson is reconstructed from the remaining lepton and the neutrino (missing transverse momentum). However, the longitudinal component of the neutrino (E_z^{miss}) is unknown, thus it is calculated by applying an invariant mass constraint on the W boson as follows:

$$p_W^2 = (p_l + p_\nu)^2 = M_W^2 \,, \tag{5.1}$$

where p_l as well as p_ν represent the four-momentum vectors of the light lepton and the neutrino that decay from the W boson respectively, and the mass of the W boson is set to $M_W = 80.4$ GeV. The quadratic equation 5.1 can be expressed in terms of E_z^{miss}:

$$E_z^{\text{miss}} = \frac{\alpha \cdot P_z^l \pm \sqrt{(E^l)^2 \cdot (\alpha^2 - P_{\text{T}}^l \cdot E_{\text{T}}^{\text{miss}})}}{(P_{\text{T}}^l)^2} \,, \tag{5.2}$$

where α is given by:

$$\alpha = \frac{M_W^2}{2} + P_{\text{T}}^l \cdot E_{\text{T}}^{\text{miss}} \,. \tag{5.3}$$

Depending on the sign under the square root in equation 5.2, there can be different types of solutions: 1) if it is positive ($\alpha^2 \geq P_{\text{T}}^l \cdot E_{\text{T}}^{\text{miss}}$), then there are two real solutions; 2) if negative ($\alpha^2 < P_{\text{T}}^l \cdot E_{\text{T}}^{\text{miss}}$), then the solutions are not real; 3) if it is equal to 0, then one real exact solution exist. For the first case, the smallest value in the magnitude is taken based on the expectation that the W boson is likely to be produced with small rapidity. The problematic case is the second one, and simply ignoring the square root in equation 5.1 will violate the initial assumption on the invariant mass of the W boson. Therefore, one possible remedy could be fixing $m_{\text{T}}(W) = M_W$, vanishing the square root in equation 5.2. Then a quadratic dependence of E_x^{miss} and E_y^{miss} appears, and this can be resolved by introducing a new scale factor β, which is defined as:

$$\beta = \frac{M_W^2}{2 \cdot P_{\text{T}}^l} \cdot E_{\text{T}}^{\text{miss}} - P_{\text{T}}^l \cdot E_{\text{T}}^{\text{miss}} \,, \tag{5.4}$$

where β is used to scale E_x^{miss}, E_y^{miss} and $E_{\text{T}}^{\text{miss}}$. Then α is recalculated as shown in equation 5.3. E_z^{miss} can be calculated by considering only the offset part of equation 5.2. [59]

CHAPTER 6

tHq study part 1: origin of two light leptons

In this chapter, an overview and the first part of the tHq analysis, which is our main interest in this thesis, is presented. The tHq analysis is mainly focused on the ML decay channel of the tHq process which includes the Higgs boson decay channels $H \to \tau\tau/WW/ZZ$. Of particular interest, we search for the $tHq \to t\tau_l\tau_h q$ (lep-had) channel, where one of the τ lepton decays leptonically and the other decays hadronically. After a general description of the event selection procedure for the lep-had channel, a study on addressing the ambiguities in the origin of light leptons has been carried out, since each of the two light leptons in the final state can either originate from the Higgs boson or the top quark.

6.1 $tHq \to t\tau_l\tau_h q$ channel

As mentioned in Section 2.4.5, the Higgs boson can be radiated either from the top quark or from the exchanged W boson in two dominant LO diagrams (see Fig 2.6). Depending on which channel the Higgs boson decays through, several possible tHq final states can be listed. The target Higgs decay modes that have been used for direct searches of the tHq production in the 8 TeV [73] and 13 TeV dataset [23] by the CMS collaboration are: $H \to \tau\tau$, $H \to WW$ and $H \to ZZ$ modes with ML signatures in the final states; $H \to b\bar{b}$ with a signature of a single lepton and a $b\bar{b}$ pair; and $H \to \gamma\gamma$ with a high signal significance for a reinterpretation of the measurement. Of particular interest for us among those tHq final states are those with the ML final state. A recent study on the tHq ML final state using the full 2016 LHC dataset of at 13 TeV by the CMS collaboration has been reported in [74].

Due to diverse decay channels of the Higgs boson, several distinct decay channels can lead to the signatures of the ML final state in each of $H \to WW$, $H \to \tau\tau$ and $H \to ZZ$ channels. The decay channel that we are interested in is $H \to \tau\tau$, where the τ leptons decay either hadronically or leptonically, resulting in three distinctive channels: $H \to \tau_h\tau_h$ (had-had channel), $H \to \tau_l\tau_h$ (lep-had channel) and $H \to \tau_l\tau_l$ (lep-lep channel). This analysis targets an unique topology of the $tHq \to t\tau_l\tau_h q$ channel. Total five possible signal regions in the ML final state are defined and illustrated in Fig 6.1, depending on the number of light leptons[1] and hadronically decaying τ leptons. The lep-had channel is marked as $2l+1\tau_{had}$. One example of a Feynman diagram of the event topology for this channel is illustrated in Fig 6.2.

[1] Here, light leptons refer to electrons and muons including those which originate from τ lepton decays.

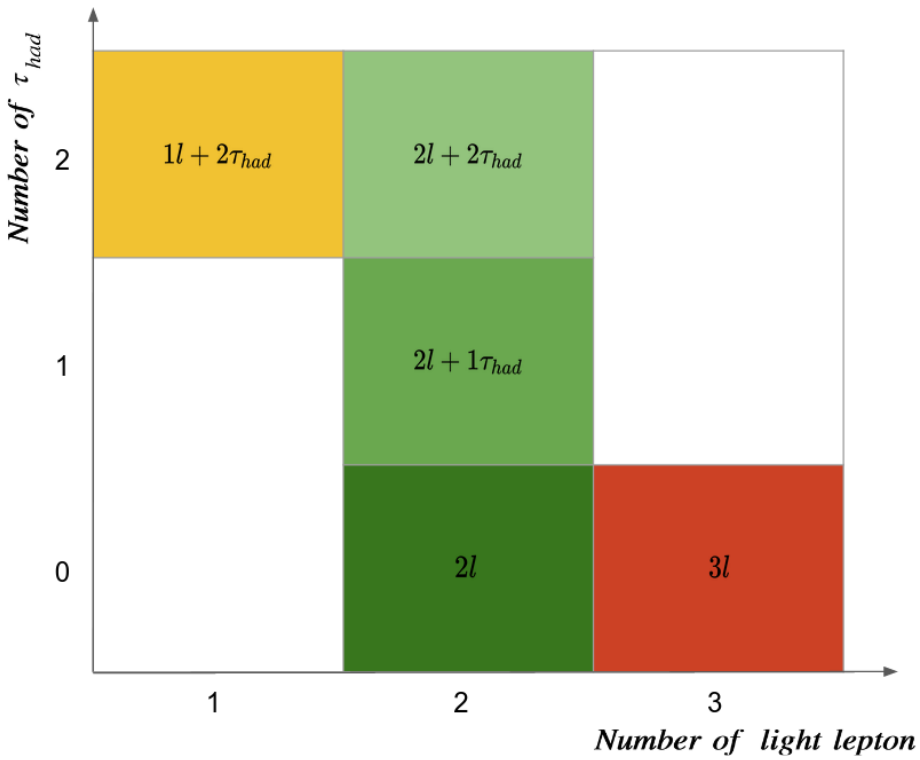


Figure 6.1: All possible signal regions which result in the tHq multilepton final states.

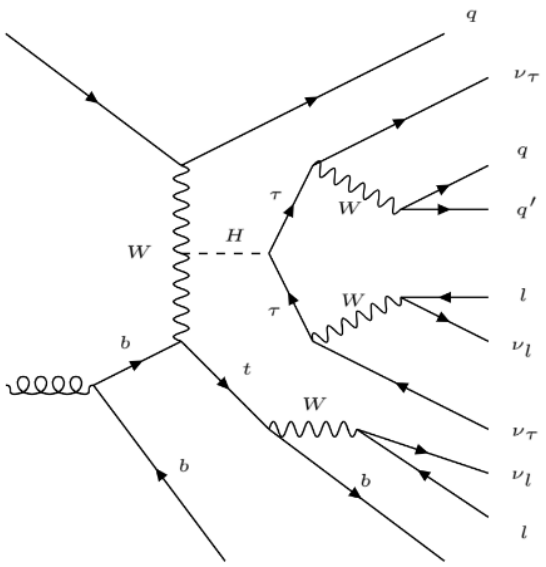


Figure 6.2: One example Feynman diagram for the $tHq \rightarrow t\tau_l\tau_h q$ (lep-had) channel.

6.2 Event selection

In this section, the strategy for selecting the $tHq \to t\tau_l\tau_h q$ channel, as well as the event yields of all processes including the signal and backgrounds are given. The used data and MC samples can be found in Appendix A. The main background sources to the signal process are those described in Section 5.2 in the tZq study (the tZq process is a background source here).

6.2.1 Signal region

The final state of the lep-had channel consists of exactly three leptons including two light leptons and one hadronically decaying τ lepton, 2-6 jets including radiation, forward-jets and exactly one b-jet , four neutrinos of which one from top quark and three from the Higgs boson. In the followings are selection criteria used for defining the signal region:

- Exactly three leptons are required, of which two are light leptons (e or μ) and one is the hadronically decaying τ lepton. The three leptons are required to fulfill the conditions $p_{\mathrm{T}}(l_1) > 28$ GeV, $p_{\mathrm{T}}(l_2) > 20$ GeV and $p_{\mathrm{T}}(l_3) > 20$ GeV (l_1 has the highest p_{T}). All leptons must be reconstructed in the central detector region and thus must satisfy $|\eta| < 2.5$. In addition, as the overlap removals for the muons and τ leptons are not working between the cracks in the detector, vetos have been applied: for the muons, a veto of $0.0 < |\eta| < 0.01$ is applied; and the τ leptons are vetoed in region $1.37 < |\eta| < 1.57$. Additionally, the leptons are required to pass the single lepton trigger and tight isolation working point. For the hadronic τ leptons, passing RNN medium is required, and the overlap removal recommended by ASG is used to make sure that no energy deposition is double counted.

- We constrain the jet multiplicity to be from 2 to 6 in order to increase the statistics of the tHq signal events. The jets include inclusively the forward- and radiation jets, and are required to satisfy $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 4.5$. Exactly one b-jet is required out of those jets, fulfilling the conditions $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$. The 70% working point is used for selecting the b-jets, and for the b-tagging the `MV2c10_PC` algorithm is used.

- The total missing transverse energy is required to be at least larger than 5 GeV, as the total missing transverse energy in this boosted system is less likely to be canceled out.

tHq lep-had selections
2 ll (e or μ), 1 τ_h
p_{T} (l_1) > 28 GeV, p_{T} (l_2) > 20 GeV, p_{T} (l_3) > 20 GeV, $\lvert\eta\rvert < 2.5$
2-6 jets, $p_{\mathrm{T}} > 25$ GeV, $\lvert\eta\rvert < 4.5$
1 b-jet, $p_{\mathrm{T}} > 25$ GeV, $\lvert\eta\rvert < 2.5$
$E_{\mathrm{T}}^{\mathrm{miss}} > 5$ GeV

Table 6.1: An overview of the requirements applied when selecting the signal regions of the lep-had channel.

Process	Number of events
tZq	38.4 ± 0.5
Diboson	155.0 ± 2.6
Triboson	1.1 ± 0.1
$t\bar{t}$	3 788.5 ± 14.6
$t\bar{t}$V	168.2 ± 1.2
$t\bar{t}$H	47.6 ± 0.2
W+jets	−1.0 ± 2.4
Z+jets	2 481.9 ± 85.2
tW	355.8 ± 9.1
tWZ	22.5 ± 0.7
4top	1.0 ± 0.1
tHq	1.4 ± 0.2
Total expected	7 060.4 ± 87.0
Data	7312.0

Table 6.2: Numbers of expected and observed event yields in the lep-had SR broken down by process, obtained with `tHqLoop`. The uncertainties include only statistical uncertainties. Here, the diboson sample is an inclusive sample[2]. The negative number of events in the W+jets is thought to be caused by low statistics.

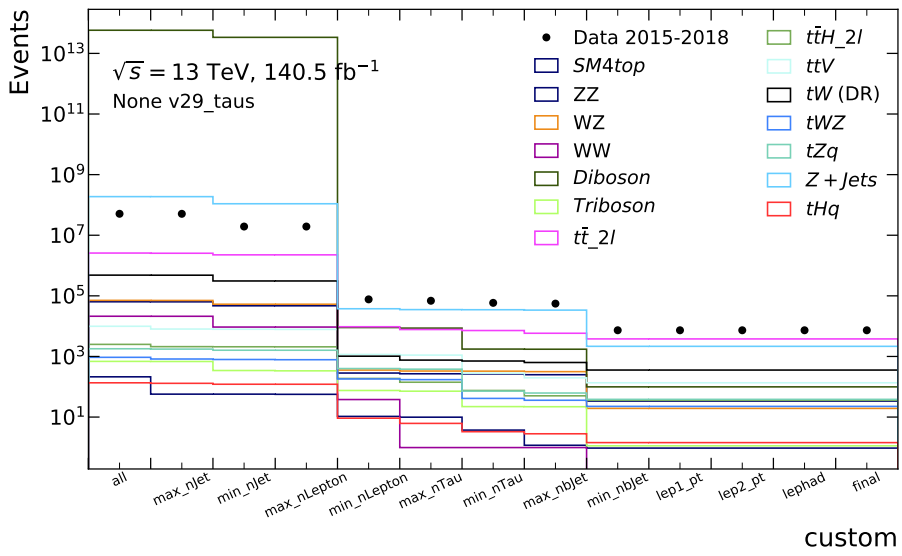


Figure 6.3: The cutflow of the measured event yields after each SR selection requirement on the signal and background processes in the lep-had SR. Here, W+jets is not shown due to a technical issue caused by low statistics.

6.2.2 Signal region yields and plots

The event yields of all simulated processes including the signal and background processes after applying the full selection in the lep-had SR can be found in Table 6.2. The cutflow plot is shown in Fig 6.3. The distributions of lepton- and jet-related variables are shown in Fig 6.4.

6.3 Origin of light leptons

The two light leptons in the final state of the lep-had channel can originate either from the Higgs boson or the top quark. Thus, there are ambiguities regarding the origins of these light leptons, making the reconstruction of Higgs and top quark extremely difficult. Nevertheless, the charge as well as the flavor of these leptons could provide us useful information to probe for their origins. For example, two τ leptons decaying from the Higgs boson must be opposite charge (OS) to each other. Two light leptons must be opposite signed in some backgrounds as well, such as $t\bar{t}$ and Z+jets (involves a Z boson). By requiring two SS light leptons, the large backgrounds can be suppressed. We can also discriminate the signal from other backgrounds which do not involve a Z or a Higgs boson by grouping the leptons by their flavor.

6.3.1 Categorization of light leptons

Based on the idea mentionlight ed above, the two light leptons are classified first into two different lepton pairs depending on their sign: opposite sign (OS) and same sign (SS) pairs. Table 6.3 illustrates all possible configurations of the light lepton pairs with different origins. Here, the two light leptons are labeled as l_1 and l_2 respectively for the discrimination (they are not p_{T}-sorted). Note that, in this table the decay channel 1 and 2 which involve the l_2 and l_1 from the (anti-) top quark respectively are differentiated unlike the case for the Higgs boson (exchange of the l_1 and l_2 of which both originate from the Higgs boson will not give rise to a new light lepton composition in the final state), since the productions of the anti-top quark and the top quark contribute to the number of OS and SS pairs differently[3]. In other words, if the l_1 in the decay channel 1 is from the top quark, then the l_2 in the decay channel 2 must come from the anti-top quark (see Section 6.3.2 for more explanations). Then the light lepton pairs are further categorized by their flavor, i.e. into same flavor (SF) and different flavor (DF) pairs. At the end, there are four different categories of the light lepton pairs: OSSF, OSDF, SSSF, SSDF. The theoretical expectations of the number of events in each category is presented in Section 6.3.2. In order to see the contributions from the backgrounds in each composition, the number of observed events as well as the signal strength are obtained.

decay channel		1	2	3
H	τ	τ_h	τ_h	l_1
	τ	l_1	l_2	l_2
t	W	l_2	l_1	τ_h

Table 6.3: All possible configurations of the origin of the (not p_{T}-sorted) light leptons l_1 and l_2 in the lep-had channel. Invisible decay products are not shown. The charge and flavor of all particles are not specified.

[2] The diboson samples are divided into four different sets of samples: ZZ, WZ, ZZ and "Diboson", and each set contains different DSID samples.

[3] The top quark decays into a positively-charged light lepton, i.e $t \to W^+b \to l^+\nu_l b$, whereas the anti-top quark decays into a negatively-charged light lepton, $\bar{t} \to W^-\bar{b} \to l^-\bar{\nu}_l\bar{b}$.

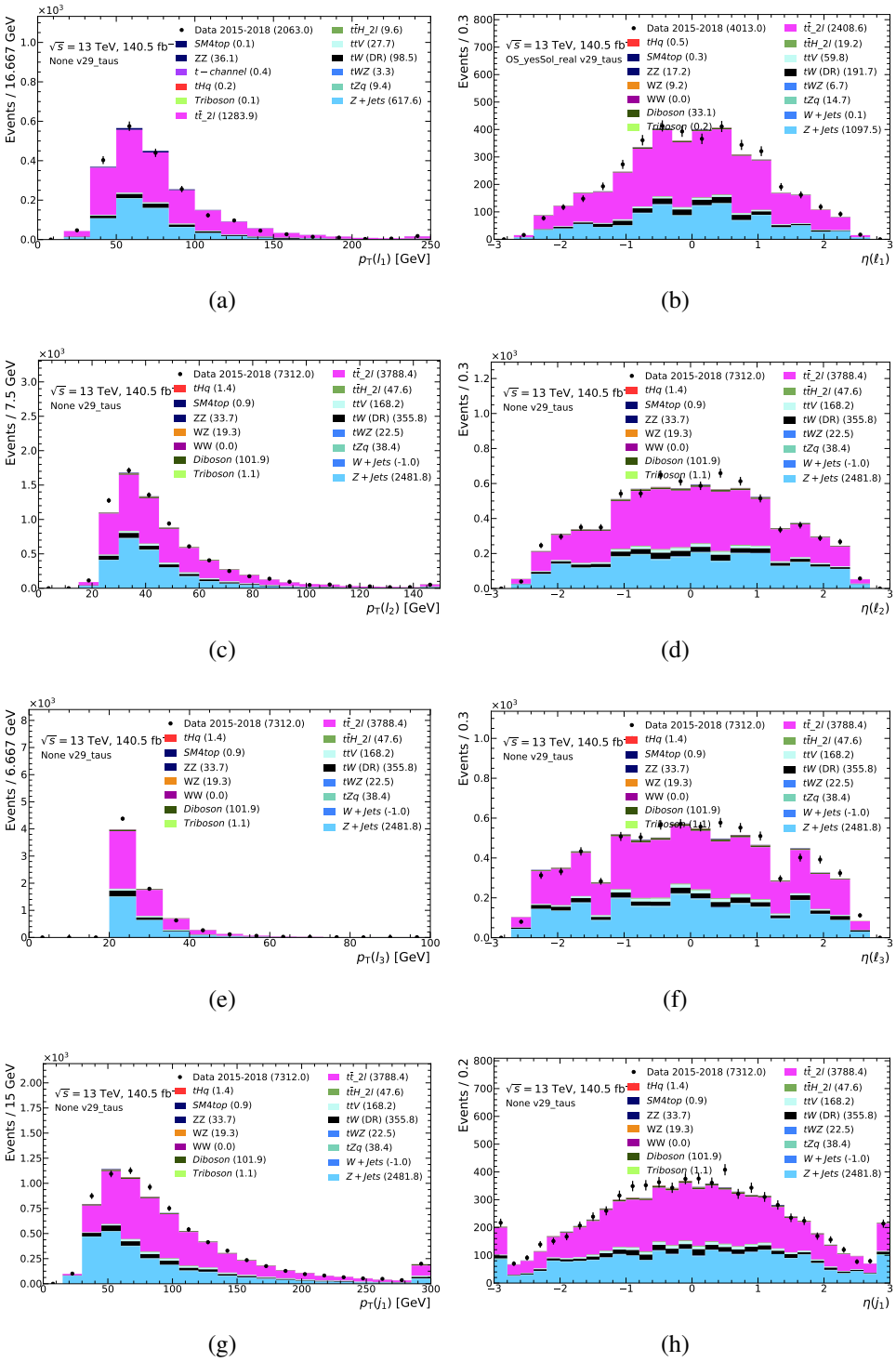


Figure 6.4: Comparison of data and MC predictions for reconstructed lepton-related quantities for events in the lep-had channel. The uncertainty band includes only statistical uncertainties for the signal and backgrounds.

6.3.2 Calculation of the number of events

In this section, the strategy for calculation of theoretically expected number of events will be given. One example of the calculation procedure for a decay mode of $H \to \tau\tau$ is shown.

Number of events

The number of expected events can be calculated with the expression below:

$$N_{\text{event}} = \mathcal{L} \cdot \sigma_{\text{total}} \cdot \text{BR} \cdot f_{t/\bar{t}} \equiv \mathcal{L} \cdot \sigma_{\text{sig}} \,, \tag{6.1}$$

where N_{event} is the expected number of events, $\mathcal{L}$ is the total integrated luminosity during Run 2, and the symbol BR represents the branching ratio (BR) of individual decay channels. The σ_{total} is the total production cross-section of the inclusive tHq process, including $pp \to tHq$ and $pp \to \bar{t}Hq$. $f_{t/\bar{t}}$ is defined in order to account for the contribution of each of these two different productions, respectively. The branching ratio of any decay channels which involve the top or anti-top quark needs to be multiplied by this factor accordingly. We define a quantity denoted as σ_{sig} as the cross-section of individual decay channels for convenience. The total integrated luminosity here is taken as $\mathcal{L} = 139$ fb^{-1} and the total cross-section of the tHq process is taken as $\sigma_{\text{total}} = 72.55$ fb for t-channel at NLO at $\sqrt{s} = 13$ TeV with the combined flavor scheme (4F + 5F) [75]. For the fraction factor $f_{t/\bar{t}}$, we take the production cross-section of the tHq and $\bar{t}Hq$ processes as $\sigma_{tHq} = 47.64$ fb and $\sigma_{\bar{t}Hq} = 24.88$ fb, respectively [75]. The fraction factor f_t and $f_{\bar{t}}$ then can be calculated as below:

$$f_t = \frac{\sigma_{tHq}}{\sigma_{\text{total}}} = \frac{47.64 \text{ fb}}{72.55 \text{ fb}} \cong 0.6567 \,, \tag{6.2}$$

$$f_{\bar{t}} = \frac{\sigma_{\bar{t}Hq}}{\sigma_{\text{total}}} = \frac{24.88 \text{ fb}}{72.55 \text{ fb}} \cong 0.3429 \,. \tag{6.3}$$

Branching ratio

There is a very important point that needs to be taken into account during the calculation of the BR: the light lepton from the leptonic decay of the top quark can either originate from a light lepton which decays directly from the W boson or from the leptonic decay of a τ lepton which decays from the top quark. Thus, in the channel where the top quark decays leptonically, the branching ratios of these two decay channels are summed up:

$$\text{BR}(\text{leptonic decay of } t) = \text{BR}(t \to bW \to bl\nu_l) + \text{BR}(t \to bW \to b\tau\nu_\tau \to bl\nu_\tau\nu_l) \,, \tag{6.4}$$

where l represents the light lepton, i.e. $l = e,\ \mu$. One example of the BR calculation procedure is given below, for one OS case where the Higgs boson decays into one leptonically decaying τ lepton and one hadronically decaying τ lepton, while the top quark decays leptonically. Table 6.4 illustrates the simplified event topology of the final state of this process.

The branching ratio for this process is as follows:

$$\begin{aligned} &\text{BR}(H \to \tau^+\tau^-,\ \text{leptonic decay of } t) \\ &= \text{BR}(H \to \tau^+\tau^-) \cdot \text{BR}(\tau \to \tau_h) \cdot \text{BR}(\tau \to l) \cdot \text{BR}(t \to bW) \cdot \text{BR}(W \to l, \tau_l) \cdot f_t \\ &\cong 0.00227 \,. \end{aligned} \tag{6.5}$$

decay channel		OS lepton pair
H	τ^+	τ_h^+
	τ^-	l^-
t	W^+	l^+, τ_{l^+}

Table 6.4: One possible tHq lep-had decay channel which results in one OS light lepton pair in the final state. The leptonic decay of the top quark here represents the inclusive leptonic decay, denoted as l^+, τ_{l^+} where l stands for the light lepton which decays directly from the top quark and τ_l the leptonically decaying τ lepton from the top quark. τ_h stands for hadronic decay of the τ lepton.

Here, the $W \to l, \tau_l$ represents the inclusive leptonic decay channel of the top quark that is mentioned in the previous section (τ_l represents the leptonically decaying τ lepton).

6.3.3 Theoretical expectations of the number of events in $H \to \tau\tau/WW/ZZ$

In this section, the expected number of events for each of the tHq ML decay channels $H \to \tau\tau/WW/ZZ$ is given. For the rest of this thesis, following conventions will be used for the sign/flavor tables: similar to Table 6.4, the first two columns show the parent particles where the first column represents two major parent particles, the Higgs boson (H) and the top quark (t), and the second column represents their intermediate decay particles $\tau^{+/-}$ and $W^{+/-}$ (b-quark is not shown). The top quark and antitop quark ($\bar{t}$) shown in the first column represent two separate productions associated with the Higgs boson. Next are the columns illustrating the final state particles including only the light leptons ($l^{+/-}$), hadrons ('had'), leptonically decaying τ leptons ($\tau_{l^{+/-}}$) and hadronically decaying τ leptons (τ_h). Invisible decay products are not shown. Each of the decay channels which contain different origin of the light leptons is numbered as decay channel i ($i \in \mathbb{Z}$) in the second top row. The last two columns show the total number of events for both OS and SS lepton pair configurations. In the flavor combination (SF/DF) table that corresponds to each OS/SS configuration in the sign combination table, four different flavor combinations for the two light leptons $l_1 l_2$ are considered: ee, $e\mu$, μe and $\mu\mu$. The pair ee and $\mu\mu$ are grouped into OSSF or SSSF pair, and $e\mu$ and μe are grouped into OSDF or SSDF pair, depending on their sign.

Expected number of events in $H \to \tau\tau$

All possible sign (OS and SS) configurations for the origin of light leptons in $H \to \tau\tau$ are shown in Table 6.5. In this channel, two τ leptons that come from the Higgs boson can decay either leptonically or hadronically while the top quark must decay leptonically (decay channel 3-6); or the Higgs boson can decay into a pair of leptonically decaying τ leptons while the top quark decays into a hadronically decaying τ lepton (decay channel 1-2). Table 6.7 shows all flavor lepton pair combinations for each sign (OS and SS) pair configuration shown in Table 6.5.

Expected number of events in $H \to WW$

In the decay channel $H \to WW$, all three W bosons must decay leptonically, of which two must decay into two light leptons and one must decay into a hadronically decaying τ lepton. Table 6.6 shows all possible event topologies with different origins of the light leptons. The SF and DF tables for Table 6.6 can be found in Appendix C.

lepton pair		OS				SS		OS total	SS total
decay channel		1	2	3	4	5	6	1 - 4	5 - 6
H	τ^+	l^+	l^+	τ_h^+	l^+	τ_h^+	l^+		
	τ^-	l^-	l^-	l^-	τ_h^-	l^-	τ_h^-		
t	W^+	τ_h^+		l^+, τ_{l^+}			l^+, τ_{l^+}		
$\bar{t}$	W^-		τ_h^-		l^-, τ_{l^-}	l^-, τ_{l^-}			
BR(%)		0.04	0.02	0.23	0.13	0.13	0.23		
σ_{sig}[fb]		0.03	0.02	0.17	0.10	0.10	0.17	0.30	0.26
N_{event}		3.67	2.13	23.19	13.50	13.50	23.19	42.50	36.70

Table 6.5: All possible sign (OS and SS) configurations for the origin of light leptons in $H \to \tau\tau$. The BR is expressed in percentage (%) for convenience. The top quark (t) and antitop quark ($\bar{t}$) shown in the first column represent two separate productions associated with the Higgs boson. The OS pairs are from decay channel 1 to 4, and the SS pairs are from the decay channel 5 to 6.

lepton pair		OS				SS		OS total	SS total
decay channel		1	2	3	4	5	6	1 - 4	5 - 6
H	W^+	l^+, τ_{l^+}	l^+, τ_{l^+}	τ_h^+	l^+, τ_{l^+}	τ_h^+	l^+, τ_{l^+}		
	W^-	l^-, τ_{l^-}	l^-, τ_{l^-}	l^-, τ_{l^-}	τ_h^-	l^-, τ_{l^-}	τ_h^-		
t	W^+	τ_h^+		l^+, τ_{l^+}			l^+, τ_{l^+}		
$\bar{t}$	W^-		τ_h^-		l^-, τ_{l^-}	l^-, τ_{l^-}			
BR(%)		0.06	0.04	0.06	0.04	0.04	0.06		
σ_{sig}[fb]		0.05	0.03	0.05	0.03	0.03	0.05	0.15	0.07
N_{event}		6.55	3.81	6.55	3.81	3.81	6.55	20.72	10.36

Table 6.6: All possible sign (OS and SS) configurations for the origin of light leptons in $H \to WW$. The BR is expressed in percentage (%) for convenience. The top quark (t) and antitop quark ($\bar{t}$) shown in the first column represent two separate productions associated with the Higgs boson. The OS pairs are from decay channel 1 to 4, and the SS pairs are from the decay channel 5 to 6.

lepton pair		OS-1	OSSF		OSDF	
decay channel		1 - 4	1	2	3	4
H	τ^+	l^+	e^+	μ^+	e^+	μ^+
	τ^-	l^-	e^-	μ^-	μ^-	e^-
t	W^+	τ_h^+	τ_h^+	τ_h^+	τ_h^+	τ_h^+
BR(%)		0.04	0.01	0.01	0.01	0.01
σ_{sig}[fb]		0.03	0.01	0.01	0.01	0.01
N_{event}		3.67	0.94	0.89	0.92	0.92

lepton pair		OS-2	OSSF		OSDF	
decay channel		1 - 4	1	2	3	4
H	τ^+	l^+	e^+	μ^+	e^+	μ^+
	τ^-	l^-	e^-	μ^-	μ^-	e^-
$\bar{t}$	W^-	τ_h^-	τ_h^-	τ_h^-	τ_h^-	τ_h^-
BR(%)		0.02	0.01	0.01	0.01	0.01
σ_{sig}[fb]		0.02	0.00	0.00	0.00	0.00
N_{event}		2.13	0.55	0.52	0.53	0.53

lepton pair		OS-3	OSSF		OSDF	
decay channel		1 - 4	1	2	3	4
H	τ^+	had	had	had	had	had
	τ^-	l^-	e^-	μ^-	e^-	μ^-
t	W^+	l^+, τ_{l^+}	e^+, τ_{e^+}	μ^+, τ_{μ^+}	μ^+, τ_{μ^+}	e^+, τ_{e^+}
BR(%)		0.23	0.06	0.06	0.06	0.06
σ_{sig}[fb]		0.17	0.04	0.04	0.04	0.04
N_{event}		23.19	5.90	5.70	5.84	5.76

lepton pair		OS-4	OSSF		OSDF	
decay channel		1 - 4	1	2	3	4
H	τ^+	l^+	e^+	μ^+	e^+	μ^+
	τ^-	had	had	had	had	had
$\bar{t}$	W^-	l^-, τ_{l^-}	e^-, τ_{e^-}	μ^-, τ_{μ^-}	μ^-, τ_{μ^-}	e^-, τ_{e^-}
BR(%)		0.13	0.03	0.03	0.03	0.03
σ_{sig}[fb]		0.10	0.02	0.02	0.02	0.02
N_{event}		13.50	3.43	3.40	3.35	3.32

lepton pair		SS-5	SSSF		SSDF	
decay channel		1 - 4	1	2	3	4
H	τ^+	had	had	had	had	had
	τ^-	l^-	e^-	μ^-	e^-	μ^-
t	W^+	l^-, τ_{l^-}	e^-, τ_{e^-}	μ^-, τ_{μ^-}	μ^-, τ_{μ^-}	e^-, τ_{e^-}
BR(%)		0.13	0.03	0.03	0.03	0.03
σ_{sig}[fb]		0.10	0.02	0.02	0.02	0.02
N_{event}		13.50	3.43	3.40	3.35	3.32

lepton pair		SS-6	SSSF		SSDF	
decay channel		1 - 4	1	2	3	4
H	τ^+	l^+	e^+	μ^+	e^+	μ^+
	τ^-	had	had	had	had	had
t	W^+	l^+, τ_{l^+}	e^+, τ_{e^+}	μ^+, τ_{μ^+}	μ^+, τ_{μ^+}	e^+, τ_{e^+}
BR(%)		0.23	0.06	0.06	0.06	0.06
σ_{sig}[fb]		0.17	0.04	0.04	0.04	0.04
N_{event}		23.19	5.90	5.70	5.84	5.76

Table 6.7: All possible flavor combinations (SF and DF) of each OS and SS light lepton pair configuration in Table 6.5. The first column next to the "decay channel" column represents the combined (not classified by flavor) column. From column 1 to 4 the flavor combinations of ee, $\mu\mu$, $e\mu$, μe are shown.

Expected number of events in $H \to ZZ$

In the decay mode $H \to ZZ$, there are more possible combinations than in the previous two channels, as a Z boson can also decay into a pair of ν or quarks. Table 6.8 shows all the decay channels in $H \to ZZ$ that lead to the lep-had final state. Complementary to the explanation given in Table 6.8, one example decay mode where one Z boson decays into a pair of τ leptons and the other Z boson decays into hadrons or a pair of neutrinos ($Z \to \tau\tau$, $Z \to \nu\nu$/had, $t \to l, \tau_l$) will be explained. In this decay mode, there are two further possible decay channels: one of the τ leptons decays into a light lepton and the other τ lepton decays hadronically, while then the top quark must decay leptonically (decay channel 3 from Table 6.8); or both of the τ leptons decay leptonically and the top quark decays into one hadronically decaying τ lepton (decay channel 4 from Table 6.8). All possible sign configurations are illustrated in Table 6.9 and 6.10. Since the contribution from $H \to ZZ$ to the number of OS and SS lepton pairs is relatively lower than those from $H \to \tau\tau$ and $H \to WW$, further categorization into SF and DF lepton pairs is neglected.

Z, Z	$\tau\tau$	ee	$\mu\mu$	had	$\nu\nu$
$\tau\tau$	x	x	x	3) l, τ_l 4) τ_h	5) l, τ_l 6) τ_h
ee	x	x	x	1) τ_h	2) τ_h
$\mu\mu$	x	x	x	1) τ_h	2) τ_h
had	3) l, τ_l 4)τ_h	1) τ_h	1) τ_h	x	x
$\nu\nu$	5) l, τ_l 6) τ_h	2) τ_h	2) τ_h	x	x

3/5) $Z \to \tau_l\tau_h$, $t \to l, \tau_l$, 4/6) $Z \to \tau_l\tau_l$, $t \to \tau_h$

Table 6.8: All possible decay channels in $H \to ZZ$ that lead to the lep-had final state. The first column as well as the first row illustrate the decay modes of each of the two Z bosons respectively. The entry of the table denotes the decay mode of the top quark. The entries marked with 'x' represent the decay mode that do not result in the lep-had final state irrespective of the decay mode of the top quark. The rest that are marked with the same number (from 1 to 6) are the same decay mode. For the decay mode 3 to 6, whether they can lead to the lep-had final state depends on the decay modes of two τ leptons from the Z boson. These decay modes can result in the lep-had final state only if the Z boson and top quark decay as illustrated below the table. The decay modes that are labeled with the same number result in the same final state.

Summary table for expected number of events in $H \to \tau\tau/WW/ZZ$

A summary for the theoretically expected number of events in all considered channels $H \to \tau\tau/WW/ZZ$ is presented in Table 6.11. This table shows that, apart from $H \to \tau\tau$ there are also some contributions from $H \to WW$ and $H \to ZZ$. As can be seen from the table, OS to SS ratio is the smallest for $H \to \tau\tau$, OS/SS $\cong$ 1.158, and is the largest for $H \to ZZ$, OS/SS $\cong$ 3.704. However, the $H \to ZZ$ channel has a very low contribution to the total number of events in OS and SS configuration in comparison to the other two channels. The overall OS/SS ratio is 1.369, which indicates that there is a higher chance for an OS light lepton pair ($\sim$ 57.58%) to be produced than a SS light lepton pair ($\sim$ 42.22%) in the lep-had signal region.

lepton pair		OS							
decay channel		1	2	3	4	5	6	7	8
H	Z	l^+l^-	l^+l^-	l^+l^-	l^+l^-	$\tau_{l^+}\tau_{l^-}$	$\tau_{l^+}\tau_{l^-}$	$\tau_{l^+}\tau_{l^-}$	$\tau_{l^+}\tau_{l^-}$
	Z	had	had	$\nu\nu$	$\nu\nu$	had	had	$\nu\nu$	$\nu\nu$
t	W^+	τ_h^+		τ_h^+		τ_h^+		τ_h^+	
$\bar{t}$	W^-		τ_h^-		τ_h^-		τ_h^-		τ_h^-
BR(10^{-5})		5.74	3.34	1.64	0.96	0.36	0.21	0.10	0.06
σ_{sig}[ab]		4.19	2.44	1.20	0.70	0.26	0.15	0.07	0.004
N_{event}		0.59	0.34	0.17	0.10	0.04	0.02	0.01	0.01

Table 6.9: Possible sign (OS) configurations for the origin of light leptons in $H \to ZZ$. The BR is expressed in 10^{-5} for convenience. The top quark (t) and antitop quark ($\bar{t}$) shown in the first column represent two separate productions associated with the Higgs boson. More possible OS light lepton pair configurations are shown in Table 6.9 complementary to this table. The numbering of the decay channels is irrelevant to Table 6.8.

lepton pair		OS				SS				OS total	SS total
decay channel		9	10	11	12	13	14	15	16	1 - 12	13 - 16
H	Z	$\tau_{l^-}\tau_{h^+}$	$\tau_{l^+}\tau_{h^-}$	$\tau_{l^-}\tau_{h^+}$	$\tau_{l^+}\tau_{h^-}$	$\tau_{l^+}\tau_{h^-}$	$\tau_{l^-}\tau_{h^+}$	$\tau_{l^+}\tau_{h^-}$	$\tau_{l^-}\tau_{h^+}$		
	Z	had	had	$\nu\nu$	$\nu\nu$	had	had	$\nu\nu$	$\nu\nu$		
t	W^+	l^+, τ_{l^+}		l^+, τ_{l^+}		l^+, τ_{l^+}		l^+, τ_{l^+}			
$\bar{t}$	W^-		l^-, τ_{l^-}		l^-, τ_{l^-}		l^-, τ_{l^-}		l^-, τ_{l^-}		
BR(%)		2.26	1.31	0.65	0.38	2.26	1.31	0.65	0.38		
σ_{sig}[ab]		1.65	0.96	0.47	0.27	1.65	0.96	0.47	0.27	12.41	3.35
N_{event}		0.23	0.13	0.07	0.04	0.23	0.13	0.07	0.04	1.74	0.47

Table 6.10: Possible sign (OS and SS) configurations for the origin of light leptons in $H \to ZZ$ complementary to Table 6.9. The BR is expressed in 10^{-5} for convenience. The last two columns present the total number of events in the OS and SS lepton pair configurations including those from Table 6.9. The OS pairs are from decay channel 1 to 12, and the SS pairs are from the decay channel 13 to 16. The numbering of the decay channels is irrelevant to Table 6.8.

decay channel	SSSF [fb]	SSDF [fb]	OSSF [fb]	OSDF [fb]	SS [fb]	OS [fb]	total [fb]	OS/SS
$H \to \tau\tau$	0.13	0.13	0.15	0.15	0.26	0.30	0.57	1.16
$H \to WW$	0.04	0.04	0.07	0.07	0.07	0.15	0.22	2.00
$H \to ZZ$					0.003	0.01	0.02	3.70
total σ_{sig}	0.17	0.17	0.23	0.23	0.34	0.46	0.80	1.37
total N_{event}	23.52	23.52	31.64	31.64	47.46	64.96	112.42	1.37

Table 6.11: A summary table for the theoretically expected number of events in all considered channels $H \to \tau\tau/WW/ZZ$. Except the total number of events N_{event} and OS/SS ratio, all entries are the signal cross section in unit of [fb]. The flavor categorization is neglected in $H \to ZZ$ due to its low contribution.

6.3.4 Observed number of events

In this section, the observed number of events in each sign and flavor composition, i.e. OSSF, OSDF, SSSF, SSDF is shown. The (reweighted) observed number of events for the OS and SS pair configurations including the flavor classifications for the signal and backgrounds can be found in Table 6.12 and 6.13. A summary for the comparison between the expected and observed number of events is given in Table 6.14. The observed number of events of the signal (tHq) in this table is way lower than its theoretical expectation. We assume that the discrepancy might originate from the low statistics of the signal sample itself.

As illustrated in Table 6.12, the major backgrounds to the signal are $t\bar{t}$ and Z+jets. The $t\bar{t}$ accounts for about $\sim 39\%$ of the total number of events of the backgrounds in the SF pair region, whereas it accounts for about $\sim 89\%$ in the DF pair region. This is due to different contributions from the Z+jets in the SF and DF regions. Contrary to the $t\bar{t}$ process where the number of SF and DF leptons does not differ too much (slightly more DF lepton pairs), the contribution of the Z+jets is significantly reduced in OSDF region than in OSSF region, as the light leptons can originate only from the Z boson. By applying a cut requiring at least one DF lepton pair, we can suppress the background from Z+jets. However, a large contamination from $t\bar{t}$ is also expected since there is no restrictions on the flavor of the light leptons in the $t\bar{t}$. Therefore, a further study on these two major backgrounds in terms of the flavor of the light leptons is needed.

The overall background contributions are drastically decreased in the SS regions as can be seen from Table 6.13. Especially the decrease of the number of events in the $t\bar{t}$ is very prominent, as the pair of top and antitop quark is only allowed to decay into an OS lepton pair. Since the contributions of the major backgrounds in the SS regions are fairly small, no big differences can be found between the SF and DF regions. About 38% of the number of events in the signal has been cut away in the SS region compared to that in the OS region, from 0.89 ± 0.18 to 0.55 ± 0.14. However, as can be seen from Table 6.14, the signal strength is much higher in the SS (SSSF and SSDF) regions regardless of the flavor of the light leptons: S/B in the SS regions is larger than in the OS (OSSF and OSDF) regions approximately by a factor of 24 and a factor of 4 for $S/\sqrt{B}$. This effect is graphically represented in Fig 6.5. This indicates that, applying a cut requiring at least one SS lepton pairs might be an efficient background suppression method. The observed number of events of the tHq in the OS regions is roughly 53 times less than the expected value, and is about 117 times less in the SS regions. We presume that this discrepancy might be caused by significantly low statistics of the tHq sample.

	OSSF				OSDF				OS	
$l_1 l_2$	e^-e^+ / e^+e^-		$\mu^-\mu^+$ / $\mu^+\mu^-$		$e^-\mu^+$ / $e^+\mu^-$		μ^-e^+ / μ^+e^-		total	
tZq	12.54	± 0.28	15.46	± 0.29	2.04	± 0.11	1.54	± 0.09	31.57	± 0.43
Diboson	59.33	± 1.72	69.85	± 1.74	5.19	± 0.26	4.89	± 0.54	139.27	± 2.51
Triboson	0.24	± 0.03	0.26	± 0.03	0.19	± 0.03	0.14	± 0.03	0.82	± 0.06
$t\bar{t}$	848.93	± 6.90	995.54	± 7.39	991.85	± 7.53	888.12	± 7.04	3 724.44	± 14.44
$t\bar{t}V$	46.96	± 0.63	49.39	± 0.63	18.88	± 0.47	16.73	± 0.45	131.96	± 1.10
$t\bar{t}H$	8.48	± 0.09	8.67	± 0.09	8.65	± 0.09	7.73	± 0.08	33.53	± 0.18
W+jets	0.12	± 0.12	0.00	± 0.00	0.16	± 0.15	0.00	± 0.00	0.28	± 0.19
Z+jets	1 068.16	± 55.77	1 402.75	± 64.34	2.68	± 1.24	0.72	± 1.64	2 474.32	± 85.18
tW	87.31	± 4.68	95.02	± 4.64	90.57	± 4.67	77.99	± 4.08	35.09	± 9.05
tWZ	7.61	± 0.43	7.62	± 0.40	1.52	± 0.18	1.53	± 0.17	18.28	± 0.64
4top	0.15	± 0.02	0.18	± 0.02	0.20	± 0.02	0.16	± 0.02	0.68	± 0.04
total bkg	2 139.91	± 56.43	2 644.72	± 64.96	1 121.95	± 8.96	999.55	± 8.33	6 906.13	± 86.92
tHq	0.24	± 0.10	0.24	± 0.09	0.21	± 0.09	0.20	± 0.09	0.89	± 0.18
MC	2 140.16	± 56.43	2 644.96	± 64.96	1 122.16	± 8.96	999.75	± 8.33	6 907.02	± 86.91
data	2220.00		2778.00		1136.00		992.00		7126.00	

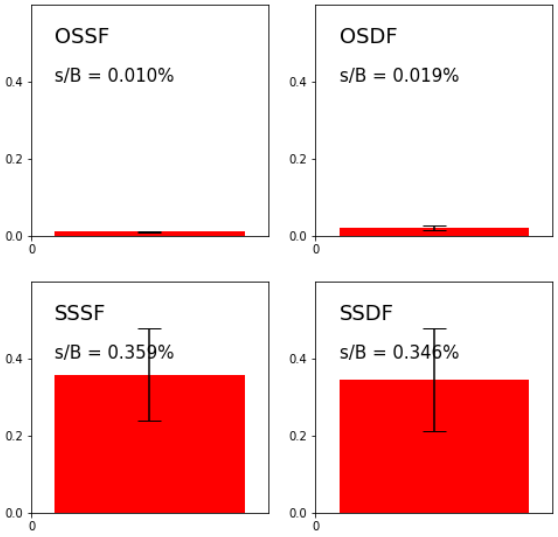
Table 6.12: An overview of the reweighted observed number of events in OS pair configuration including SF and DF pair configurations for all MC samples including the signal and backgrounds. l_1 and l_2 denote each of the two light leptons respectively, but without p_T sorting (they are used solely for discrimination here). The number of events from the tHq process is singled out in the last row in order to be compared with the total number of events from all backgrounds. The row represented as MC includes the signal and all backgrounds.

	SSSF		SSDF		SS
$l_1 l_2$	e^-e^- / e^+e^+	$\mu^-\mu^-$ / $\mu^+\mu^+$	$e^-\mu^-$ / $e^+\mu^+$	μ^-e^- / μ^+e^+	total
tZq	1.94 ± 0.11	1.67 ± 0.09	1.82 ± 0.11	1.44 ± 0.09	6.88 ± 0.20
Diboson	4.69 ± 0.36	3.26 ± 0.21	4.36 ± 0.26	3.37 ± 0.25	15.69 ± 0.55
Triboson	0.08 ± 0.02	0.09 ± 0.02	0.07 ± 0.02	0.07 ± 0.02	0.31 ± 0.04
$t\bar{t}$	26.78 ± 1.21	6.25 ± 0.61	15.50 ± 0.92	15.47 ± 0.86	64.01 ± 1.85
$t\bar{t}V$	9.23 ± 0.28	9.30 ± 0.28	9.85 ± 0.29	7.89 ± 0.26	36.27 ± 0.55
$t\bar{t}H$	3.61 ± 0.07	3.29 ± 0.06	3.77 ± 0.07	3.44 ± 0.07	14.10 ± 0.14
W+jets	0.27 ± 0.18	0.39 ± 0.39	0.08 ± 0.08	−2.11 ± 2.39	−1.99 ± 2.42
Z+jets	6.99 ± 3.37	0.33 ± 0.22	−0.17 ± 0.33	0.38 ± 0.23	7.53 ± 3.40
tW	2.69 ± 0.79	0.40 ± 0.31	1.26 ± 0.15	1.45 ± 0.56	4.91 ± 1.06
tWZ	1.20 ± 0.16	0.90 ± 0.14	1.26 ± 0.15	0.90 ± 0.13	4.26 ± 0.29
4top	0.08 ± 0.01	0.06 ± 0.01	0.06 ± 0.01	0.06 ± 0.01	0.27 ± 0.03
total bkg	57.56 ± 3.71	25.94 ± 0.91	36.97 ± 1.12	32.36 ± 2.64	152.83 ± 4.77
tHq	0.13 ± 0.07	0.17 ± 0.07	0.13 ± 0.07	0.11 ± 0.06	0.55 ± 0.14
MC	57.69 ± 3.71	26.10 ± 0.91	37.11 ± 1.12	32.47 ± 2.65	153.38 ± 4.78
data	64.00	36.00	43.00	43.00	186.00

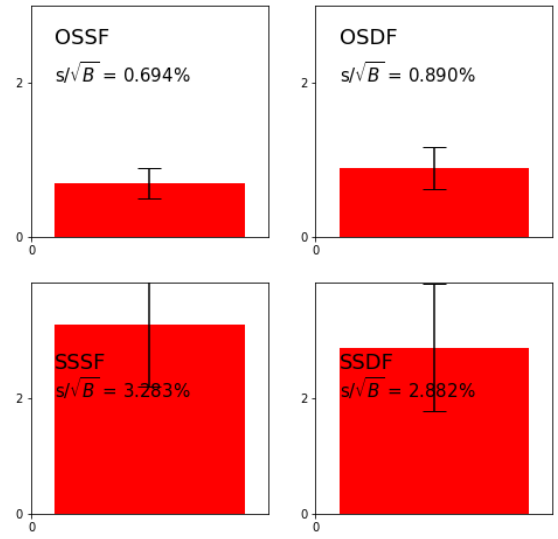
Table 6.13: An overview of the reweighted observed number of events in SS pair configuration including SF and DF pair configurations for all MC samples including the signal and backgrounds. l_1 and l_2 denote each of the two light leptons respectively, but without p_T sorting (they are used solely for discrimination here). The number of events from the tHq process is singled out in the last row in order to be compared with the total number of events from all backgrounds. The row represented as MC includes the signal and all backgrounds. The negative numbers are thought to be caused by technical issues such as low statistics.

	OSSF	OSDF	SSSF	SSDF	total
total bkg	4 784.63 ± 86.05	2 121.50 ± 12.23	83.50 ± 3.82	69.33 ± 2.87	7 058.96 ± 87.04
tHq	0.48 ± 0.13	0.41 ± 0.13	0.30 ± 0.10	0.24 ± 0.09	1.43 ± 0.23
S/B	0.00010 ± 0.00003	0.00019 ± 0.00006	0.00359 ± 0.00120	0.00346 ± 0.00134	
$S/\sqrt{B}$	0.00694 ± 0.00195	0.00890 ± 0.00276	0.03283 ± 0.01086	0.02882 ± 0.01109	
tHq (expected)	23.52	23.52	31.64	31.64	112.42

Table 6.14: A summary for the comparison between the expected and observed number of events in all sign and flavor configurations. The table shows the reweighted observed number of events of the backgrounds and the signal (tHq) in the first two rows respectively. The signal strength is calculated as S/B and $S/\sqrt{B}$. The last row represents the theoretical expectation for the signal.



(a) S/B



(b) $S/\sqrt{B}$

Figure 6.5: The signal strength calculated as S/B (Fig 6.5(a)) and $S/\sqrt{B}$ (Fig 6.5(b)) in all sign and flavor categories. The reweighted observed number of events are used, taken from Table 6.14. The uncertainties are calculated as $\delta(S/B) = \sqrt{(\frac{\delta s}{s})^2 + (\frac{\delta B}{B})^2} \cdot (S/B)$ for S/B and B replaced with $\sqrt{B}$ for $S/\sqrt{B}$.

CHAPTER 7

tHq study part 2: reconstruction of $tHq \to t\tau_l\tau_h q$ channel

In the second part of the tHq study, an overview of the reconstruction of the final state of $tHq \to t\tau_l\tau_h q$ (lep-had) channel is discussed. First, a general strategy regarding the reconstruction of major parent particles in the lep-had channel, i.e. the Higgs boson and top quark will be discussed. Then the reconstruction procedure of these particles will be presented in mainly two steps: the reconstruction of the top quark and the Higgs boson ($\tau\tau$-system). Along with the reconstruction of parent particles, we test the validity of the assumptions we made in the first step by truth-matching studies. Note that due to the low statistics of the tHq MC samples, mostly the tZq MC samples are used throughout the whole section.

7.1 Reconstruction strategy

As introduced in Section 6.1, the Higgs boson and the top quark are the major parent particles in the lep-had channel. Unfortunately, an accurate reconstruction of these particles is extremely challenging due to four neutrinos in the final state, of which three are from the Higgs boson and one is from the top quark, as can be seen in Fig 6.2. One possible strategy to fully reconstruct the final state is to reconstruct the top quark first and then the $\tau\tau$ system. This idea is based on the fact that total missing transverse momentum, $\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}$(total), is the sum of the missing transverse momentum from the di-tau system and that from the top quark as below:

$$\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}(\text{total}) = \overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}(\text{Higgs}) + \overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}(\text{top}). \tag{7.1}$$

As the total $\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}$ can be obtained through the object reconstruction procedure described in Section 3.3.7, if $\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}$(top) is known, then $\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}$(Higgs) can be calculated by subtracting the missing transverse momentum of the top quark from the total $\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}$. Since there is only one neutrino originating from the top quark, reconstructing the top quark system can be done with the analytical method, i.e. $\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}$ from the top quark can be fully reconstructed if the 4-momentum vector of the neutrino is completely known. The invariant mass of the $\tau\tau$ system can be reconstructed with three different methods: 1) partial invariant mass method; 2) collinear approximation; and 3) missing mass calculator (MMC). However, this strategy has a fatal drawback that, the amount of the missing transverse energy that is assigned to the Higgs system

solely depends on that of the reconstructed top quark, $E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{top}) = \sqrt{(p_{\nu,\mathrm{top}}^{x})^2 + (p_{\nu,\mathrm{top}}^{y})^2}$. Therefore, how accurately the neutrino from the top quark is reconstructed will significantly affect obtaining an accurate value for the invariant mass of the Higgs boson.

Furthermore, this reconstruction strategy is only applicable if we know which light lepton comes from the top quark, which is the case for oppositely signed (OS) light lepton pair case. There are two light leptons and one hadronically decaying τ lepton in the final state of the lep-had channel. Here, the charge of a hadronically decaying τ lepton can be known (see Section 3.3.4) and consequently the charge of the light lepton that decays from the other (leptonically decaying) τ lepton is known, which must be OS to each other since the Higgs boson is neutrally charged[1]. This way, once the light lepton from the Higgs boson is fixed, then it is obvious that the remaining light lepton is from the top quark. However, there can be two different cases for the charge of the light lepton from the top quark, either OS or SS to the light lepton from the Higgs boson. We must choose the OS light lepton case, otherwise two SS light leptons give rise to ambiguities in their origin since any of them can originate either from the Higgs boson or from the top quark. For example, assume that we have τ_h^+, e^- and μ^- in the final state (SS light lepton case), then we can either group τ_h^+ and e^-, or τ_h^+ and μ^- to reconstruct the Higgs boson. On the other hand, if we have the muon OS to the electron, i.e. τ_h^+, e^- and μ^+, then we can assure that the light lepton that has to be grouped with τ_h^+ to reconstruct the Higgs boson is e^-.

To deal with the SS light lepton case, we can try to search for some useful features in the OS light lepton case that can be used to distinguish the origin of light leptons without the information regarding the charge of the two light leptons. Then we might be able to build specific decision rules for choosing the light lepton which is more likely to originate from the Higgs boson in the SS lepton case.

7.2 Reconstruction of the top quark

In the following sections, the reconstruction procedure of the neutrino, the W boson, and the top quark will be given. Note that the reconstruction of the top quark is independent of that of any particles involved in the Higgs boson decay, such as a hadronically or a leptonically decaying τ lepton. Therefore, this reconstruction strategy can be applied to other $tHq \to t\tau\tau q$ channels, i.e. $tHq \to t\tau_h\tau_h q$ (had-had channel) and $tHq \to t\tau_l\tau_l q$ (lep-lep channel).

7.2.1 Assumptions

In order to fully reconstruct the top quark, a completely known 4-momentum vector of the missing transverse energy from the top quark is needed. First, we assume that the only source of the missing transverse energy is the neutrino. The 4-momentum vector of the neutrino, denoted as P_ν, can be expressed as:

$$P_\nu = (E_\nu\,, p_\nu^x\,, p_\nu^y\,, p_\nu^z)\,, \tag{7.2}$$

where p_ν^x, p_ν^y and p_ν^z are referred to as the x-, y- and z-component of the neutrino momentum respectively, and $E_\nu = \sqrt{m_\nu^2 + p_\nu^2}$ is the energy of the neutrino (m_ν and p_ν are the rest mass and momentum of the neutrino). Since the rest mass of the neutrino can be neglected due to its small value, there are 3 remaining unknown quantities: p_ν^x, p_ν^y and p_ν^z. This means we need 3 constraints in order to get exact solutions for those unknowns. Therefore, we make assumptions as follows:

[1] Here, we do not consider the hadronically decaying τ lepton from the top quark, even though then the two light leptons must be OS.

- Assumption 1: The top quark as well as the W boson which decays from the top quark are produced *on-shell*, i.e. the invariant masses of the top quark and the W boson are equal to their rest masses respectively. According to the four-momentum conservation law, the invariant mass equations for the top quark and the W boson can be written as below:

$$m_t^2 = (P_b + P_l + P_\nu)^2 = M_t^2 \,, \tag{7.3}$$

$$m_W^2 = (P_l + P_\nu)^2 = M_W^2 \,, \tag{7.4}$$

  where m_t as well as m_W are the invariant masses of the top quark and the W boson respectively, and P_b as well as P_l represent the 4-momentum vectors of the b-quark and the light lepton from the top quark decay. The rest masses of the top quark and the W boson are set to $M_t = 173.0$ GeV and $M_W = 80.4$ GeV. This assumption provides 2 constraints to the kinematic system of the neutrino, making it possible to gain the exact solutions for the x- and y-components when the z-component is known.

- Assumption 2: The z-component of the momentum of the neutrino from the top quark (p_ν^z) linearly depends on that of the light lepton from the top quark with a factor of an arbitrary coefficient α:

$$p_\nu^z = \alpha \cdot p_l^z \,. \tag{7.5}$$

  As mentioned in Section 3.3.7, the total $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed only with the components which lie in the transverse plane. Therefore, the z-component of the total missing energy can be considered as zero. However, this does not mean that the z-component of the $E_{\mathrm{T}}^{\mathrm{miss}}$ from the top quark is also zero, as it is the sum of the z-components from the top quark as well as the Higgs boson that is equal to zero. For this reason, we can set the z-component of the neutrino to be any value, but in a certain relation with that of the light lepton from the same origin. In other words, we assume that the W boson is likely to decay from the top quark within a certain preferred phase space. The coefficient α can be set to an arbitrary number initially, such as 0 (zero z-component) or -1 (the neutrino and the leptons are in back-to-back topology), but it can be optimized to a more probable value which gives the best reconstruction results. However, there is no promising experimental or theoretical basis which supports this assumption. Therefore, truth studies on the kinematics of the neutrino as well as the charged light lepton are mandatory.

7.2.2 Solving two invariant mass equations

Exact solutions for the 4-momentum vector of the neutrino can be obtained by solving a system of two invariant mass equations described in Assumption 1, with the assumed z-component of the neutrino from Assumption 2. Only simplified steps are given in this section (more detailed steps can be found in Appendix D).

First, the invariant mass constraint on the W boson (equation 7.4) can be written as below, assuming the neutrino and the light lepton to be massless ($m_\nu = m_l = 0$):

$$\begin{aligned} m_W^2 &= (P_l + P_\nu)^2 \\ &= m_l^2 + m_\nu^2 + 2P_l \cdot P_\nu \\ &\cong 2(E_l \cdot E_\nu - \vec{p}_l \cdot \vec{p}_l) \,. \end{aligned} \tag{7.6}$$

Similarly, the invariant mass constraint on the top quark (equation 7.3) can be written as below:

$$\begin{aligned} m_t^2 &= (P_b + P_l + P_\nu)^2 \\ &= m_b^2 + m_l^2 + m_\nu^2 + 2(P_b \cdot P_l + P_b \cdot P_\nu + P_l \cdot P_\nu) \\ &\cong 2(P_b \cdot P_l + P_b \cdot P_\nu + P_l \cdot P_\nu)\,, \end{aligned} \tag{7.7}$$

Both equation 7.6 and 7.7 can be simplified into two second-degree polynomials in terms of $x = P_\nu^x$ and $y = P_\nu^y$:

$$0 = a_1 x^2 + b_1 x + c_1 xy + d_1 y^2 + e_1 y + f_1\,, \tag{7.8}$$

for the W boson invariant mass equation, and

$$0 = a_2 x^2 + b_2 x + c_2 xy + d_2 y^2 + e_2 y + f_2\,. \tag{7.9}$$

for the top quark invariant mass equation, respectively. In the followings, all involved coefficients including a_i, b_i, c_i, d_i, e_i and f_i with $i = 1,\ 2$, are described in Appendix D. The reformed invariant mass equations are in the generic formula of the conic section[2]. Depending on the coefficients, there are three types of conic sections: the hyperbola, the parabola and the ellipse (the circle is a special case of the ellipse, though historically it was sometimes called a fourth type) [76]. For convenience, the ellipse shape is taken for describing all possible types of solution sets as shown in Fig 7.1.

In order to solve a system of 2 quadratic equations with 2 unknowns, we first express both equations in terms of only one unknown x (y is also possible) as follows:

$$x_\pm^1 = \frac{-(b_1 + c_1 y) \pm \sqrt{\mathcal{A}_1 y^2 + \mathcal{B}_1 y + C_1}}{2a_1}\,, \tag{7.10}$$

for equation 7.8, and

$$x_\pm^2 = \frac{-(b_2 + c_2 y) \pm \sqrt{\mathcal{A}_2 y^2 + \mathcal{B}_2 y + C_2}}{2a_2}\,, \tag{7.11}$$

for equation 7.9. Then, we set these equations to be equal. After simplification, we get one quartic (4th-order) equation in terms of y only:

$$A \cdot y^4 + B \cdot y^3 + C \cdot y^2 + D \cdot y + E = 0\,. \tag{7.12}$$

For a quartic equation in the form of 7.12, there can be at most 4 roots in total:

$$y_{1,2} = -\frac{B}{4A} - S \pm \frac{1}{2}\sqrt{-4S^2 - 2p + \frac{q}{S}} \tag{7.13}$$

$$y_{3,4} = -\frac{B}{4A} + S \pm \frac{1}{2}\sqrt{-4S^2 - 2p - \frac{q}{S}} \tag{7.14}$$

[2] In mathematics, a conic section (or simply conic) is a curve obtained as the intersection of the surface of a cone with a plane.

with following coefficients:

$$
\begin{aligned}
p &= \frac{8AC - 3B^2}{8A^2} \\
q &= \frac{B^3 - 4ABC + 8A^2D}{8A^3} \\
S &= \frac{1}{2}\sqrt{-\frac{2}{3}p + \frac{1}{3A}(Q + \frac{\Delta_0}{Q})} \\
Q &= \sqrt[3]{\frac{\Delta_1 + \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}} \\
\Delta_0 &= C^2 - 3BD + 12AE \\
\Delta_1 &= 2C^3 - 9BCD + 27B^2E + 27B^2E + 27AD^2 - 72ACE\,.
\end{aligned}
\tag{7.15}
$$

For convenience, the quantities under the square roots in equation 7.13 and 7.14 are denoted as:

$$D_{1,2} = -4S^2 - 2p + \frac{q}{S}\,, \tag{7.16}$$

and

$$D_{3,4} = -4S^2 - 2p - \frac{q}{S}\,, \tag{7.17}$$

respectively. After solving two invariant mass equations in terms of y, we need to reform the y into x, so that we have the x-component of the neutrino 4 momentum-vector. We recall the invariant mass equations that are written in terms of x:

$$x_1^{\pm} = \frac{-(b_1 + c_1 y) \pm \sqrt{\mathcal{A}_1 y^2 + \mathcal{B}_1 y + C_1}}{2a_1}\,, \tag{7.18}$$

for equation 7.10 and

$$x_2^{\pm} = \frac{-(b_2 + c_2 y) \pm \sqrt{\mathcal{A}_2 y^2 + \mathcal{B}_2 y + C_2}}{2a_2}\,. \tag{7.19}$$

for equation 7.11. So there can be 4 x values, x_1^+, x_1^-, x_2^+, x_2^- for each y value. However, not all of these x values are the solutions of corresponding y value, as we are solving a system of two simultaneous equations. One of the x values from the invariant mass equation of the W boson, x_1^+ or x_1^-, must be equal to one of the x values from the invariant mass equation of the top quark, x_2^+ or x_2^-. So only the x values which satisfy the condition:

$$x_1^{\pm} = x_2^{\pm}\,, \tag{7.20}$$

are selected to be paired with corresponding y value. At the end, we have at most 4 exact solutions for the system of two invariant mass equations: $\{x_i, y_i\}$ with $i = 1, 2, 3, 4$. All derivation steps as well as the involved coefficients can be found in Appendix D.

7.2.3 Problems in solutions

We obtained exact solutions for the 4-momentum vector of the neutrino from the top quark by solving two invariant mass equations with an assumption on the z-component. However, there are serious issues regarding the exact solutions, since there can be more than one solution, and the solution itself can be non-real. These are not what we expect to observe in the real world, as only one real solution exists in nature. Here, we bring up the problematic solution cases.

Multiple solutions

As shortly mentioned in the previous section, the two second-order invariant mass equations 7.8 and 7.9 can be regarded as two generic conic sections (the ellipses for convenience). The coincident intersection points of two ellipses may be viewed as the solutions to those invariant mass equations. There can be total 9 possible types of solution sets [77], and the main 5 types are visualized in Fig 7.1. The 2-dimensional plane shown in Fig 7.1 can be either real or imaginary, where the coordinates of the plane are the x- and y-components of the reconstructed neutrino momentum. Note that each intersection point (x, y) must be either real or imaginary in pairs, and therefore the corresponding plane is either real or imaginary only for the one specific intersection point. In other words, for example, all 4 intersection points in Fig 7.1 do not necessarily need to be in the real or imaginary plane at the same time (the graphical representation is to show how many possible types of solution cases available). The 9 possible types of solution sets are:

- No solution (the first diagram in Fig 7.1): two ellipses do not intersect, one shape is inside the other or two ellipses are a distance away from each other.
- One solution (the second diagram in Fig 7.1): two ellipses are tangent to each other, and intersect at exactly one point.
    a. all real (if there is only one solution, then it is always real)
- Two solutions (the third diagram in Fig 7.1): two ellipses intersect at two points.
    a. two imaginary
    b. two real
- Three solutions (the fourth diagram in Fig 7.1): two ellipses intersect at three points.
    a. two imaginary and one real
    b. two real and one imaginary
- Four solutions (the fifth diagram in Fig 7.1): two ellipses intersect at four points.
    a. all imaginary
    b. two real and two imaginary
    c. all real

Imaginary solutions

Apart from the more-than-one-solution problem, there is another issue regarding the solutions themselves: there could be non-real solutions which have no physical meaning. This issue is resulted by negative quantities under the square roots in the coefficients. As you can see from equation 7.13 and 7.14, there are several factors which give rise to non-real p_ν^y as listed below:

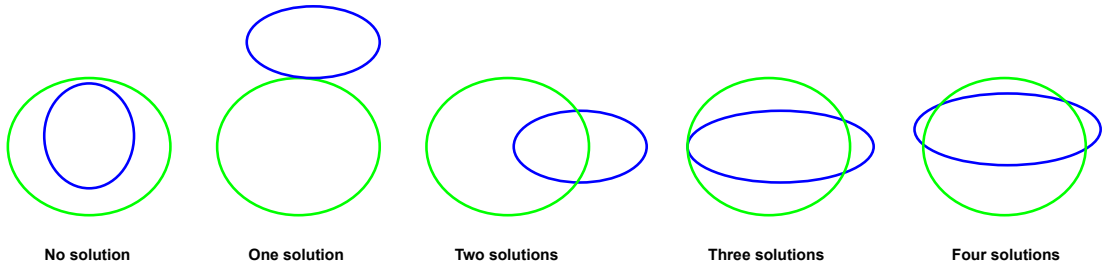


Figure 7.1: Possible types of solution sets for the points of intersection of two ellipses. The 2-dimensional plane can be either real or imaginary, where the coordinates of the plane are the x- and y-components of the reconstructed neutrino momentum. Note that each intersection point (x, y) must be either real or imaginary in pairs.

1. negative discriminants $D_{1,2} < 0$, $D_{3,4} < 0$

2. imaginary discriminants $D_{1,2}$, $D_{3,4}$

First, obviously the negative discriminants ($D_{1,2} < 0$, $D_{3,4} < 0$) in those equations lead to imaginary y values. However, the same problem can occur when $D_{1,2}$ and $D_{3,4}$ themselves are imaginary. Fig 7.2 shows the overall flow of these processes. As can be seen from their definitions in equation 7.16 and 7.17, an imaginary S can contribute to the non-real $D_{1,2}$ and $D_{3,4}$, since unlike p and q, S has a square root $\sqrt{-\frac{2}{3}p + \frac{1}{3A}(Q + \frac{\Delta_0}{Q})}$ that depends on p, A, Q, and Δ_0. Compared to the other coefficients p, A, Δ_0, the Q has a cubic root of a square root of $\Delta_1^2 - 4\Delta_0^3$. Thus, if $\Delta_1^2 - 4\Delta_0^3$ goes negative, then consequently Q, S, $D_{1,2}$, $D_{3,4}$ and y will automatically become imaginary. On the other hand, if $\Delta_1^2 - 4\Delta_0^3$ is positive, then Q will always become real, since a negative quantity under the cubic root does not give an imaginary value. Unfortunately, the positive $\Delta_1^2 - 4\Delta_0^3$ does not guarantee a real S. The S can be imaginary depending on its own under square root. Therefore, in summary, there are mainly three stages (marked as yellow rhombuses in Fig 7.2) that give rise to non-real solutions: 1) negative $\Delta_1^2 - 4\Delta_0^3$ in Q; 2) negative under square root in S; and 3) negative discriminants $D_{1,2}$ and $D_{3,4}$.

7.2.4 Remedy for multiple solutions

As has been brought up in the previous section, only one unique solution for the z-component of the neutrino (p_ν^z) is needed to properly reconstruct the neutrino 4-momentum vector. All other cases that are described in Section 7.2.3, i.e. no-solution or more-than-one-solution cases, need to be excluded. Then there can be following questions that might arise:

1. What causes more than one distinct solution?

2. Can we choose only one solution over the other solutions?

3. If not, how can we find suitable methods for choosing one unique solution?

In this section, we answer the above questions and come up with several remedies to address the multiple-solution case that has been brought up in the previous section.

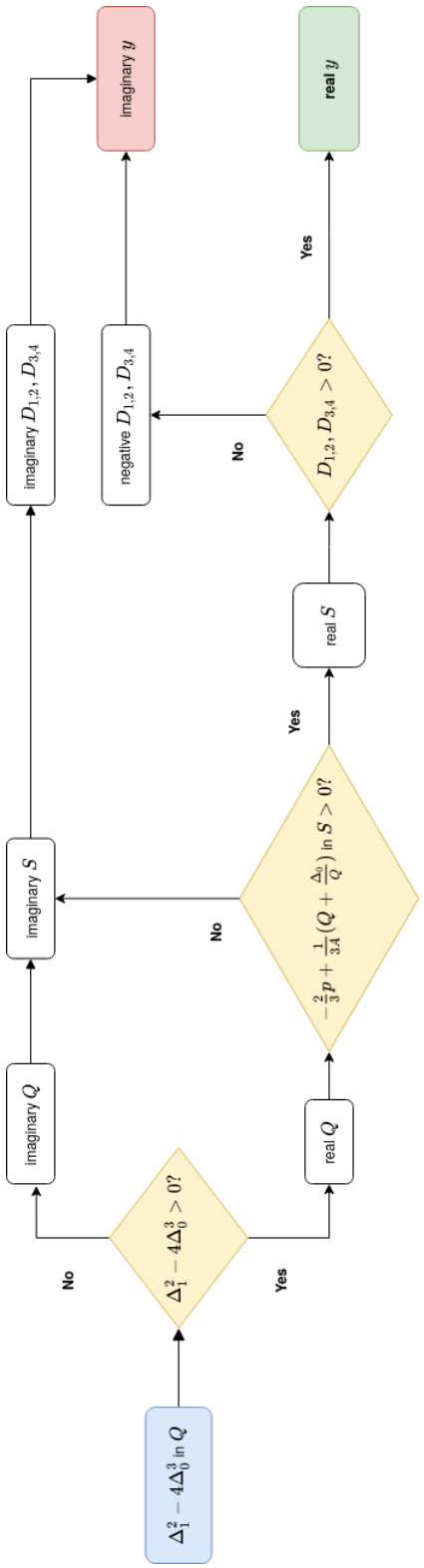
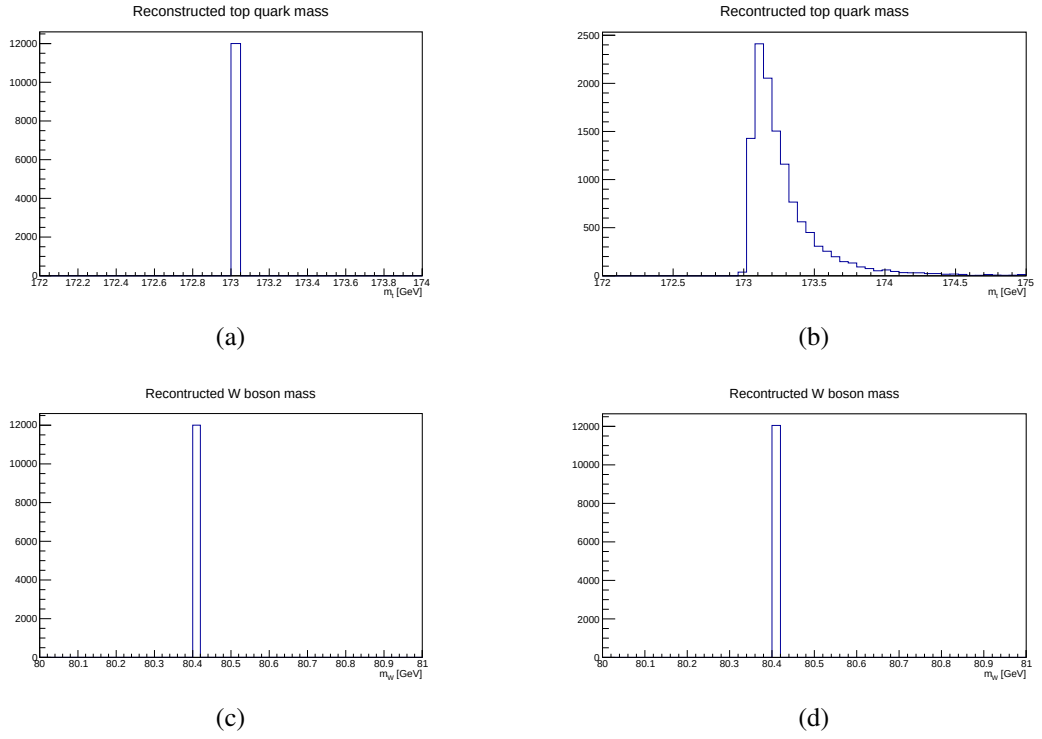


Figure 7.2: A flowchart that shows the process how an imaginary or a real y (p_ν^z) is resulted.

Cause of multiple solutions

To answer the first question above, we need to take a look back to the initial assumptions. As shown in Section 7.2.1, the invariant mass equations are second-order polynominals with 2 variables, i.e. there are quadratic terms that consist of 2 variables in the invariant mass equations. In Assumption 1, we assumed that the invariant masses of the top quark and the W boson are equal to their rest masses respectively. This means that all the solutions we obtained for the x- and y-components must give the exact values that we have assumed, i.e. $M_t = 173.0$ GeV and $M_W = 80.4$ GeV. Otherwise, the x and y values are definitely not the solutions for the invariant mass equations.

To ensure that the obtained x- and y-components are truly the solutions, we need to check if the invariant mass distributions of the top quark as well as the W boson are exactly peaked at the assumed rest masses. Fig 7.3 shows the reconstructed invariant masses of the top quark and the W boson. As can be seen from Fig 7.3(a) and 7.3(c), both invariant mass distributions of the top quark and the W boson have a peak exactly at their assumed mass values. Fig 7.3(b) and 7.3(d) show the case when the b quark is assumed to be massless. The difference between Fig 7.3(a) and 7.3(b) is not huge, since the mass of the b quark is ~ 4 GeV, which is quite small in the scale of the top quark mass (~ 173 GeV). The invariant mass distribution of the W boson stays intact regardless of the b quark mass, since no b-quark is involved in its decay. During the reconstruction procedure, the massless b-quark is assumed for convenience.



(a) (b)

(c) (d)

Figure 7.3: The invariant mass distributions of the top quark and the W boson. The figures in the top row, Fig 7.3(a) and 7.3(b) illustrate the distributions of the top quark, and the bottom row shows those of the W boson. The left and right columns show the distributions when b quark is assumed to be massive and massless, respectively. The used sample is the mc16a tZq sample with $\alpha = 0.5$.

Behavior of multiple solutions

The second question is tricky to answer since there is no reasonable standard that we can use for choosing only one solution among all possible solutions (at least in terms of reconstructing the top quark system). Therefore, we investigate the behavior of the number of solutions in two ways:

1. see how the number of solutions depends on α (an arbitrary constant that can be optimized as mentioned in Section 7.2.1)
2. check whether the kinematic distributions of the reconstructed neutrino (or W boson, top quark) exhibit any different features depending on the number of solutions

Dependency of the number of solutions on α

First, we check how the choice of α affects the results of the reconstruction. Table 7.1 to E.5 (see Appendix E.1) illustrates how frequently each of the main 5 cases (not considering the cases subdivided by the reality of the solutions) described in Section 7.2.3 takes place depending on the α value for the tZq, tHq and $t\bar{t}$ samples. In the tables, each type of the number of solutions is calculated inclusively, i.e. summed up all real and imaginary solutions. The next two columns are the total number of real and imaginary solution cases, which are added up independent of the number of solutions. The ratio of the number of no-solution cases over the total number of events as well as the ratio of the number of real solutions over the number of imaginary solutions are shown in the rightmost side of the table. Table 7.1 to E.2 (see Appendix E.1) show the number of solutions with anti-top removal, and Table E.3 to E.5 (see Appendix E.1) show those without the anti-top removal. Here, the anti-top removal is an algorithm implemented as a part of the reconstruction algorithm which is dedicated to remove the anti-top quark contribution to the lep-had channel[3]. This is done by discarding the events which contain a negatively charged light lepton as the lepton from the top quark.

Both sets of the tables with and without the anti-top removal show that the most of the multiple solution cases are 2-solution cases (only 0 or 1 event for the other solution cases). This means the two invariant ellipses almost intersect at 2 points rather than at 1, 3 or 4 points, and this phenomenon might originate from the topology of two conic sections itself as shown in Fig 7.1. Interestingly, the number of no-solution cases decreases as the α value increases (not always but in terms of a general tendency) in the range of $\alpha \in \{-1.0, 3.0\}$. However, more interestingly, the number of real solutions, or the ratio (real/imag), in all tested samples (tZq, tHq and $t\bar{t}$), reaches a certain maximum value with different α values regardless of the application of the anti-top removal. The ratio (real/imag) is at the maximum value of $\alpha = 0.5, 1.0$ and 0.75 for the tZq, tHq and $t\bar{t}$ samples using the anti-top removal, and $\alpha = 0.25, 1.0$ and 1.0 without using the anti-top removal. Fig 7.4 illustrates this behavior of the ratio (real/imag). There are some fluctuations as you can clearly see from the figure, which implies that the ratio (real/imag) might also fluctuate in a range composed of smaller steps. Additionally, it does not mean that this maximum is the global maximum as we have not checked through a large range of the α. This could be just a local maximum, and therefore we need to investigate where and how many maxima are there in the next studies.

Additionally, the slightly different optimum α value which has the maximum ratio (real/imag) with and without the anti-top removal indicate that the optimum p_ν^z value could be a bit different depending

[3] In the reconstruction algorithm, after the selection of an OS light lepton pair, the light lepton which is SS to the hadronically τ lepton is assigned to be the one from the top quark. Since no charge constraints are applied on this lepton, it could be either positively or negatively charged which means it could come from the top quark or from the anti-top quark. Therefore, in order to ensure that we are performing the truth-matching studies exactly on the top quark (anti-top quark is not considered in the truth-matching studies), we remove the anti-top quark contribution.

α	total	no sol	1 sol	2 sol	3 sol	4 sol	real sol	imag sol	no/total (%)	real/imag (%)
−1.00	99 807	56 179	1	43 626	0	1	22 343	21 285	56.29	51.21
−0.50	99 807	49 721	0	50 085	0	1	27 510	22 576	49.82	54.93
0.00	99 807	30 689	1	69 117	0	0	39 755	29 363	30.75	57.52
0.25	99 807	28 552	0	71 255	0	0	41 160	30 095	28.61	57.76
0.50	99 807	22 137	0	77 670	0	0	45 251	32 419	22.18	58.26
0.75	99 807	19 785	0	80 022	0	0	45 890	34 132	19.82	57.35
1.00	99 807	18 828	1	80 978	0	0	45 536	35 443	18.86	56.23
2.00	99 807	17 452	1	82 354	0	0	41 901	40 454	17.49	50.88
3.00	99 807	18 168	0	81 639	0	0	37 922	43 717	18.20	46.45

Table 7.1: A summary of the frequency of different number of solutions (raw number of events) for a system of two invariant mass equations, for the tZq samples including all campaigns. This is done with the anti-top removal. The uncertainties are not shown, which are the square root of the number of events assuming a Poisson distribution.

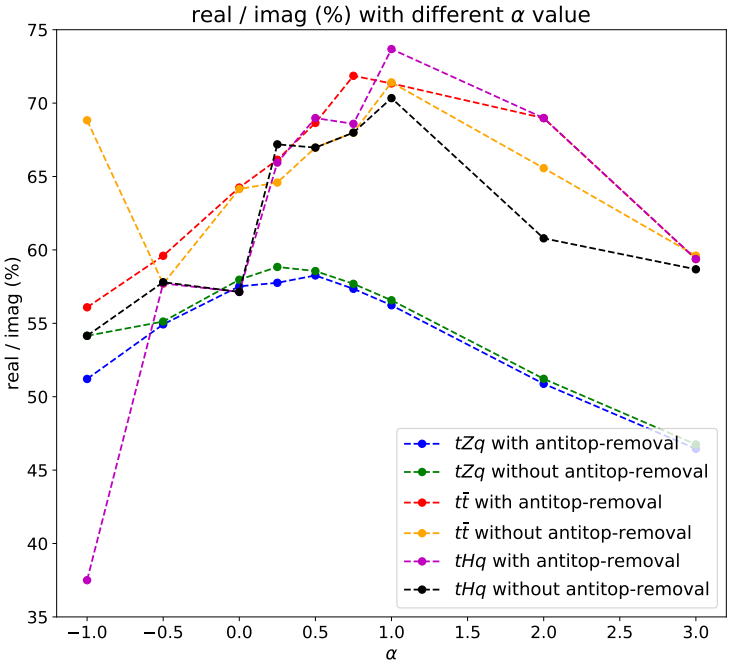
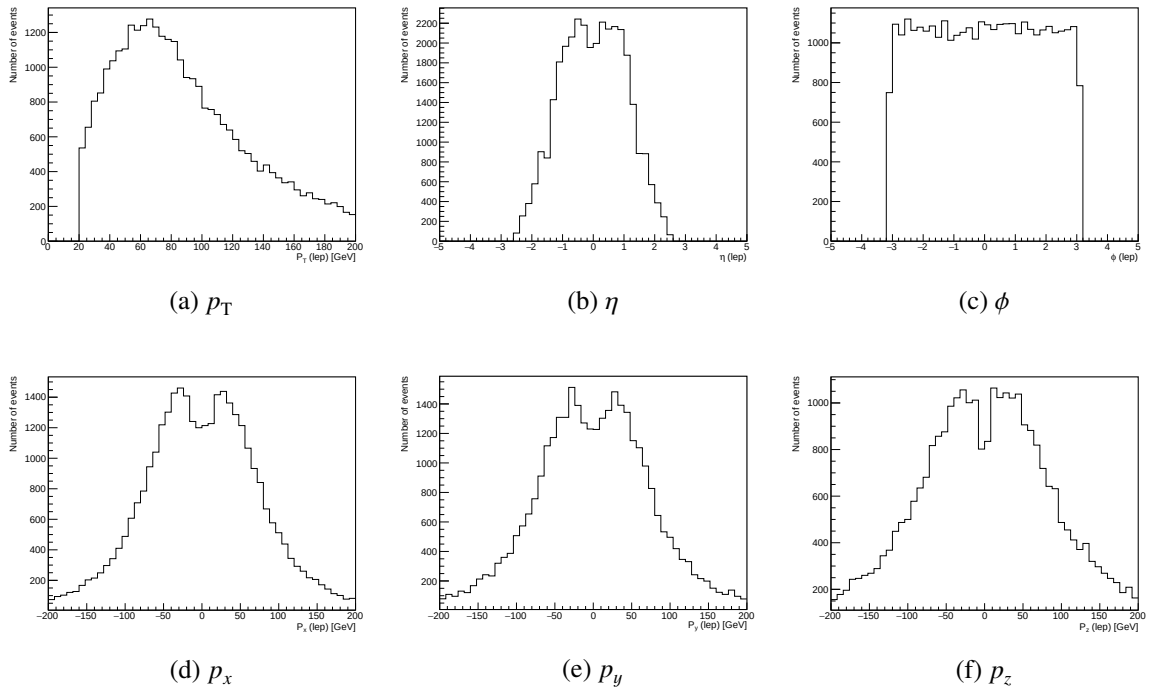


Figure 7.4: The dependency of the ratio of the number of real solutions over that of imaginary solutions (real/imag (%)) with different α value for the tZq, tHq and $t\bar{t}$ samples, with and without the anti-top removal. The uncertainties are not shown.

on the event topology of a process, even if the difference is small (less than $\sim \Delta\alpha = 1.0$). The $t\bar{t}$ is expected to have a larger contribution from the anti-top quark compared to the other two processes, but the difference between with and without the anti-top removal is not very noticeable. Apart from that, the small difference of $\sim \Delta\alpha = 0.25$ between with and without the anti-top removal in all the processes tell us that the contribution from the anti-top quark to the final state does change the dependency of the number of solutions on the α value, but not very significantly. This also can mean that there could be more other factors on which the optimum α value is dependent. For the rest of the studies, the α value is chosen to be 0.5 (the hightest real/imag ratio with the anti-top removal), which is optimized for the tZq samples that are used much more often than the tHq samples (low statistics).

Dependency of the kinematic distributions of the reconstructed particles on the number of solutions

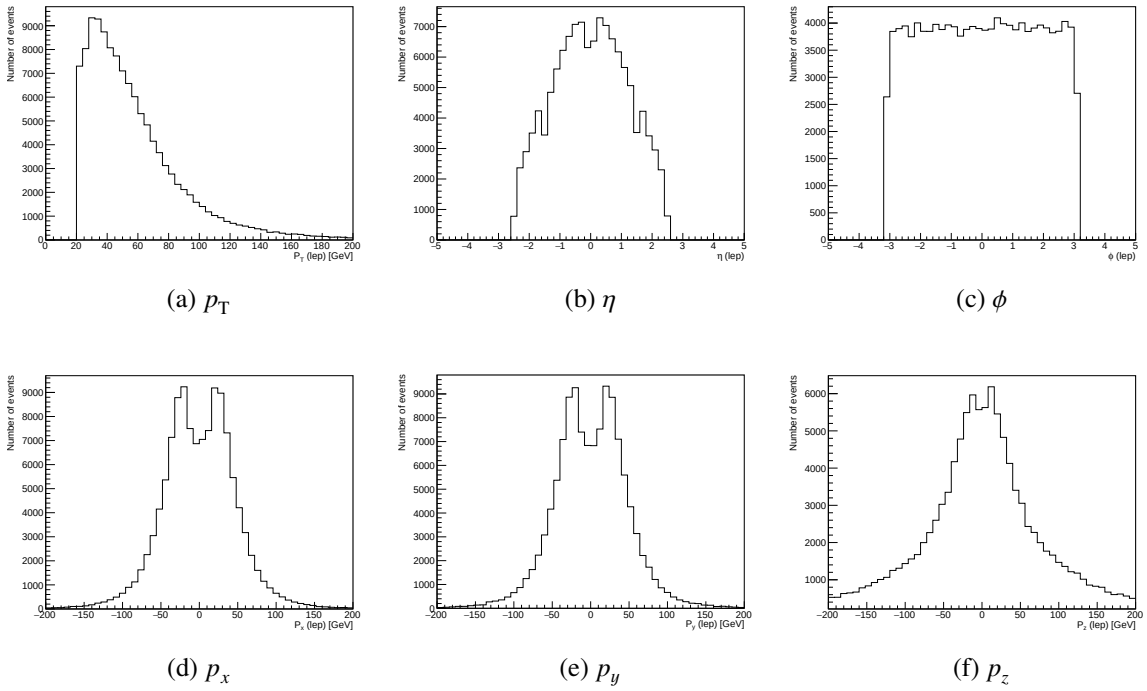


(a) p_{T} (b) η (c) ϕ

(d) p_x (e) p_y (f) p_z

Figure 7.5: The distributions of kinematic variables of the light lepton from the top quark for the 0-solution case. The tZq samples at $\alpha = 0.5$ are used including all campaigns without the anti-top removal.

Secondly, we check whether the kinematic distributions of the reconstructed neutrino present any distinct features depending on the number of solutions. In order to see this dependence, a set of unbiased quantities is needed. For this, the kinematic distributions of the light lepton from the top quark decay are used, since the reconstruction of the leptons is independent of that of the neutrino, W boson, and top quark. Furthermore, there can be more than one solution for the z-component of the reconstructed neutrino momentum, thus the kinematic properties will be biased if only one solution is chosen to be investigated. Fig 7.5 to 7.7 show the distributions of some important kinematic variables of the light lepton from the top quark for the 0-, 2- and 2-(only real) solution cases for the tZq mc16a sample. The behavior of the two solutions in the 2-solution case will be presented in Section 7.2.6 in a probe for a



(a) p_{T} (b) η (c) ϕ

(d) p_x (e) p_y (f) p_z

Figure 7.6: The distributions of kinematic variables of the light lepton from the top quark for the 2-solution case, including real and imaginary solutions (which means that the distributions include the events in which the solutions for p_ν^z are imaginary). The tZq samples at $\alpha = 0.5$ are used including all campaigns without the anti-top removal.

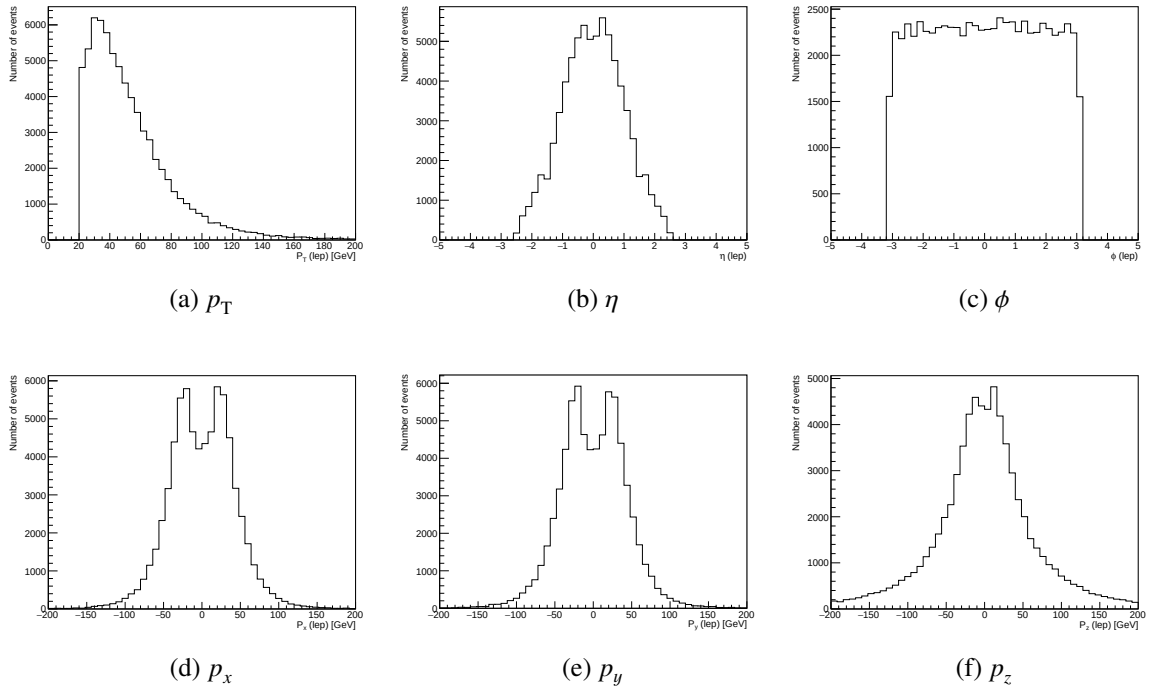
possible standard for choosing one solution.

Possible remedy for multiple solutions

To explore suitable methods for choosing one unique solution, the kinematic properties of the individual solutions in the multiple-solution cases have been examined and presented in Section 7.2.6 along with truth studies. Apart from this, there is still one thing we need to notice. According to the strategy described in Section 7.1, the reconstruction algorithm for the top quark affects the invariant mass of the Higgs boson. Unlike the top quark and the W boson whose invariant masses are constrained to their rest masses respectively, the Higgs boson (or $\tau\tau$ system) can have any value for its invariant mass since we made no assumptions or constraints on it. Based on this, the p_ν^x and p_ν^y solution set which gives the invariant mass of the $\tau\tau$ system closest to its rest mass ($m_H = 125$ GeV) might be potentially chosen as the best solution. Therefore, all obtained (real) solutions are used in parallel for further reconstruction of the W boson, top quark, and eventually the Higgs boson. The distributions of kinematic variables of the reconstructed neutrino, W boson, and top quark will be shown in Section 7.2.6.

7.2.5 Remedy for imaginary solutions

The causes of imaginary solutions are the negative discriminants $D_{1,2}$ and $D_{3,4}$ as well as the negative $\Delta_1^2 - 4\Delta_0^3$ and consequently imaginary under square root in S which cause imaginary $D_{1,2}$ and $D_{3,4}$, as discussed in Section 7.2.3. Possible remedies are listed and explained in the following.



(a) p_{T} (b) η (c) ϕ

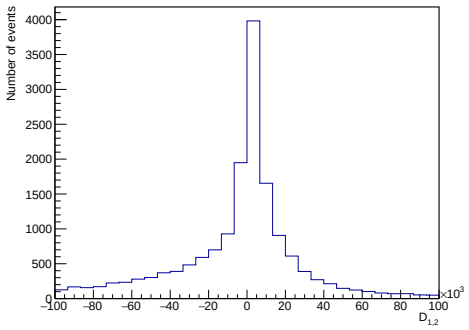
(d) p_x (e) p_y (f) p_z

Figure 7.7: The distributions of kinematic variables of the light lepton from the top quark for the 2-solution case including only real solutions. The tZq samples at $\alpha = 0.5$ are used including all campaigns without the anti-top removal.

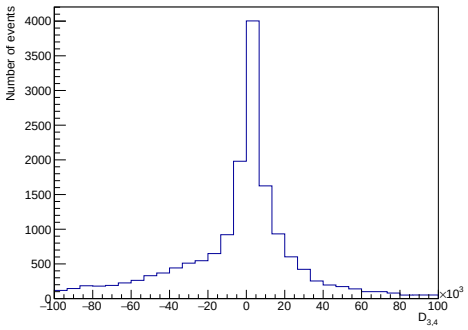
1. set $D_{1,2} = 0$ and $D_{3,4} = 0$
2. set $\Delta_1^2 - 4\Delta_0^3 = 0$
   1) if $-\frac{2}{3}p + \frac{1}{3A}(Q + \frac{\Delta_0}{Q}) < 0$ and $-\frac{2}{3}p < \frac{1}{3A}(Q + \frac{\Delta_0}{Q})$ in S, set $-\frac{2}{3}p = 0$
   2) if $-\frac{2}{3}p + \frac{1}{3A}(Q + \frac{\Delta_0}{Q}) < 0$ and $-\frac{2}{3}p > \frac{1}{3A}(Q + \frac{\Delta_0}{Q})$ in S, set $-\frac{2}{3}p + \frac{1}{3A}(Q + \frac{\Delta_0}{Q}) = 0$

In order to resolve the first case, the negative discriminant case ($D_{1,2} < 0$ or $D_{3,4} < 0$), we check how $D_{1,2}$ and $D_{3,4}$ are distributed. Those are shown in Fig 7.8(a) and 7.8(b) respectively. Both of the distributions have a peak almost near 0. Therefore, setting $D_{1,2}$ and $D_{3,4}$ equal to 0 could be a simple method to address the imaginary solution issue. The second case is more tricky to solve. In principle, if the factor $\Delta_1^2 - 4\Delta_0^3$ gives a negative value and thus result in an imaginary y, one simple remedy is to set it as 0. However, a real Q does not solve the problem so easily. In the case where there is a negative under square root in S which gives a non-real value, we can first check which one is the dominating factor under the square root, $-\frac{2}{3}p$ or $\frac{1}{3A}(Q + \frac{\Delta_0}{Q})$. These two factors possess distinct signs respectively, therefore it would not be a big problem if the dominating one has the positive sign (+), since we can simply ignore the less dominating one. However, if it is not the case, it will be quite problematic. In order to deal with the latter case, the distribution of the under square root of S are plotted, as shown in Fig 7.8(c). Here, only real under square root values are considered ($\Delta_1^2 - 4\Delta_0^3 > 0$). Interestingly, as can be seen from the figure, the under square root of S is always positive if it is real. This means that $\frac{1}{3A}(Q + \frac{\Delta_0}{Q})$ is way more dominating

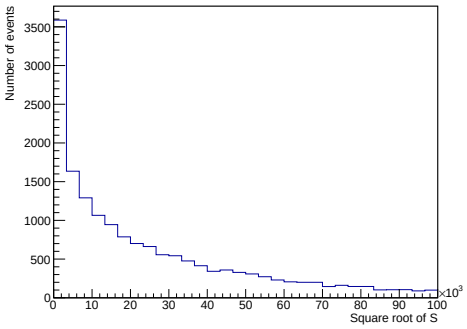
than $-\frac{2}{3}p$. Thus, there is no need to find a remedy to resolve the negative under square root of S, if the imaginary issue of Q is solved by setting $\Delta_1^2 - 4\Delta_0^3 = 0$.



(a) $D_{1,2}$



(b) $D_{3,4}$



(c) Under square root in S

Figure 7.8: The distributions of factors contributing to imaginary solutions. Fig 7.8(a) and 7.8(b) show the distributions of $D_{1,2}$ and $D_{3,4}$, and Fig 7.8(c) shows the distribution of under square root in S. Here, the units of these factors are neglected. The mc16a tZq sample at $\alpha = 0.5$ is used without the anti-top removal.

Nevertheless, setting the negative discriminant to zero constitutes a manipulation of solutions in a nonphysical way. The simplest solution is to discard all events with imaginary solutions. However, as can be seen from Table 7.1 to E.5, imaginary solution cases account for a large fraction (almost half sometimes) of the total events. Thus, it follows that discarding all events with imaginary solutions means loosing a lot of data. A better way of handling such events is strongly required. Unfortunately, all non-real solutions were ignored in our analysis as it seems to be more urgent to focus on resolving the multiple-solution issue to gain a better understanding in obtained solutions and thus choose a more probable solution. As can be seen in the flowchart in Fig 7.9, all events without at least one real root are discarded (note that these cuts are applied after the lep-had SR selection cuts described in Section 6.2.1). Additional to the general event selection procedure presented in Section 6.2.1, for the final state reconstruction this algorithm is used.

On the other hand, several different approaches have been developed by previous neutrino reconstruction researches. One of them is the modification of $E_{\mathrm{T}}^{\mathrm{miss}}$, which is exploited in a search for a light charged Higgs boson [78]. It states that the deviation of measured transverse momentum from true transverse

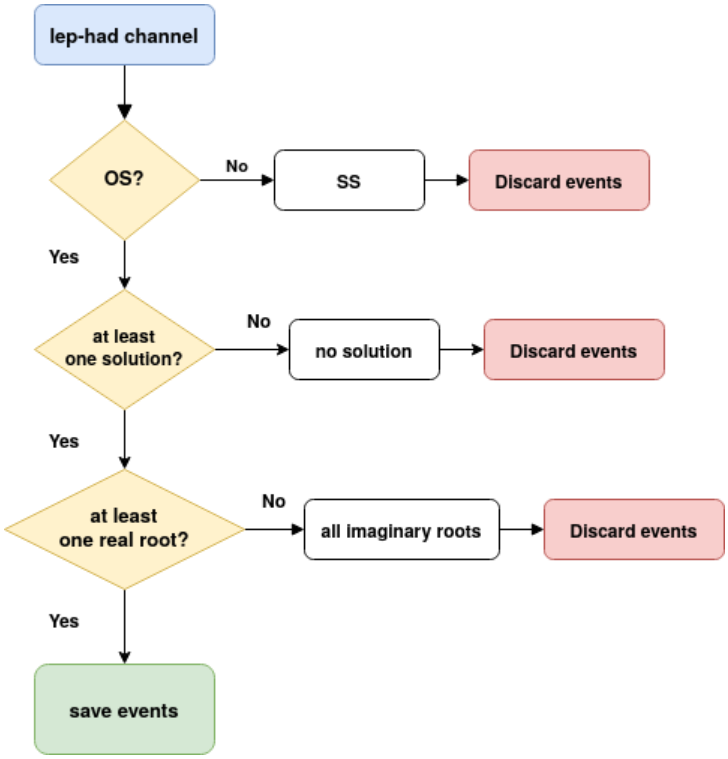


Figure 7.9: A flowchart of the event selection algorithm for the final state reconstruction of the lep-had channel. Additional to the general event selection procedure presented in Section 6.2.1, for the final state reconstruction this algorithm is used.

momentum of neutrino pairs are expected to be the most likely cause of non-real solutions. Thus, the negative discriminant (it is defined differently from this analysis) is set to 0 in the calculation of p_ν^z, while this nonphysical forcing is compensated by giving a new $E_{\mathrm{T}}^{\mathrm{miss}}$ that is calculated from the requirement that the discriminant is 0. This method has been tested and led to physically sensible invariant mass distributions. However, this approach is not completely based on the application of fundamental physical principles. It could be that the total transverse momentum is not measured perfectly, however how much this revision is physically sensitive will depend on the amount of altered transverse momentum. After all, in order to keep our calculations physically meaningful as much as possible, in this analysis we do not alter the measured total missing energy.

7.2.6 Validity of basic assumptions

As discussed previously in Section 7.2.1, the factor which contributes most to the mis-estimation of p_ν^x and p_ν^y is expected to be the assumption that the p_ν^z is linearly dependent on p_l^z, i.e. $p_\nu^z = \alpha \cdot p_l^z$. In principle, any assumption that gives one missing restriction to the remaining unknown in the neutrino 4-momentum vector can be taken. In other words, the missing assumption does not need to be made on the z-component, but can be made on one of the 3 unknowns: the x-, y- and z-components. This is because unlike the total $E_{\mathrm{T}}^{\mathrm{miss}}$, the z-component of the $E_{\mathrm{T}}^{\mathrm{miss}}$ from the top quark can differ from zero[4].

[4] This statement is not valid for the tZq ML channel which involves only one neutrino in the final state, since the total $E_{\mathrm{T}}^{\mathrm{miss}}$ is exactly the transverse momentum of the neutrino, assuming that the only source of missing transverse momentum is the

However, the p_ν^z is chosen in this analysis for the initial assumption in order to make sure that we do not alter the quantities measured independent of this reconstruction algorithm (E_x^{miss} and E_y^{miss} are reconstructed prior to this reconstruction algorithm). To check whether the validity of the assumption $p_\nu^z = \alpha \cdot p_l^z$, a truth-level study has been carried out in the following sections. Here, *truth-level* stands for the measurements of MC simulated stable particles before entering the detector, i.e. no detector smearing effects and deficiencies, trigger selection cuts, etc. The *reconstruction level* (reco-level) quantities are referred to as the kinematic variables of the particles after all detector simulations mentioned above. By comparing truth- and reco-level variables 1:1 (truth-matching), we can see how precisely the neutrino has been reconstructed.

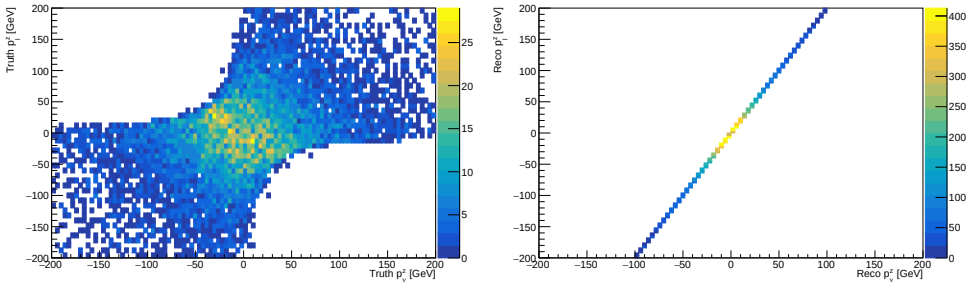
Truth-level study on p_ν^z

In this section, we perform the truth-level study on the p_ν^z and p_l^z on the truth- and reco-levels in order to check how accurately the reconstruction has been carried out. Here, note that we use the truth information of the top quark not of the anti-top quark. First of all, the correlation between p_ν^z and p_l^z on both the truth- and reco-levels is investigated. Fig 7.10(a) and 7.10(b) illustrate the correlation between p_ν^z and p_l^z on the truth- and reco-levels respectively with $\alpha = 0.5$ (which gives the highest ratio of the number of real solutions with respect to the number of imaginary solutions). As can be seen from Fig 7.10(a), there is an overall linear but more specifically a positively increasing hyperbolic correlation between p_ν^z and p_l^z, whereas rather a negative linear correlation can be seen in the central (yellow) region which is enriched with most of the events. In the contrary, an obvious linear correlation with a slope with the assumed α value can be seen in Fig 7.10(b). It is hard to determine if the correlations on the truth- and reco-levels are similar to each other or not, but it seems to be more likely that these two correlations tend to be opposite to each other since most of the events are exhibiting a negative correlation on the truth-level unlike on the reco-level. Therefore, a truth-matching study on $\alpha = -1$ as well as other values has been done and is presented in Appendix E.

To see if there is anything biasing us, we can have a further look into the properties of the light lepton since the light lepton is not involved in the reconstruction procedure of the neutrino. A comparison between the reco-level p_l^z and the truth-level p_l^z has been examined. Fig 7.10(c) shows the relation between the truth-level p_l^z and the reco-level p_l^z. The relation between the truth-level p_ν^z and the reco-level p_ν^z is illustrated in Fig 7.10(d) for the comparison. As you can see, there is a prominent 1:1 dependency of the truth-level to the reco-level for the p_l^z. Off-diagonal events are distributed in an independent way, which exhibit a similar behavior of most of the events in the truth-matching plot for the p_ν^z. The origin of these off-diagonal events needs to be found out and removed away, as this could be one of the factors which contribute to the non-linearity of the p_ν^z reco-truth correlation. We presume that there must be an additional (or double) counting that is going on, and expect that there could be a linear (or similar) tendency in the p_ν^z plot (Fig 7.10(d)) after the removal of those off-diagonal elements as a large fraction of events is concentrated in the horizontal direction (we assumed $\alpha = 0.5$ as well). Apart from this, to see further how the truth-level neutrino and lepton are correlated in a topological way, the pseudo-rapidity (η) distributions of the light lepton and the neutrino, as well as the difference between them are plotted in Fig 7.10(e). The distributions on the reco-level are shown in Fig 7.10(f). As we assumed that $p_\nu^z = 0.5 \cdot p_l^z$, the η distribution of the neutrino is much narrower than that of the light lepton on the reco-level.

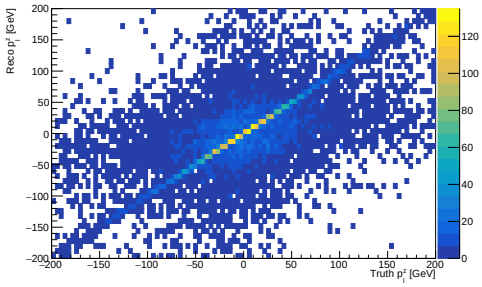
Additionally, in order to examine where the off-diagonal events are coming from, the p_{T} of the electron, muon that originate from the top quark, as well as that of the combination of these two light leptons are investigated. Before this, to ensure that we are reconstructing the lepton from the top quark (we

neutrino.

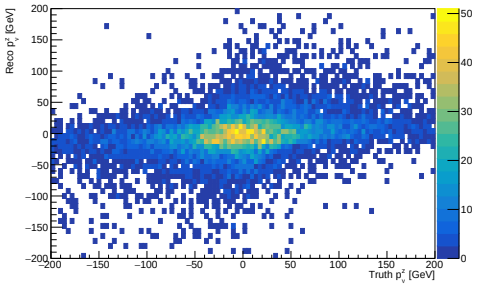


(a) Truth p_l^z vs. Truth p_ν^z

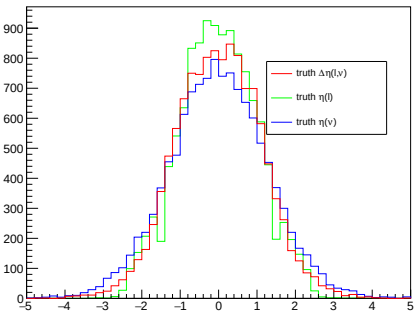
(b) Reco. p_l^z vs. Reco. p_ν^z



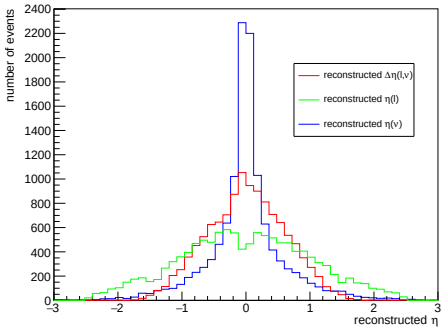
(c) Reco. p_l^z vs. Truth p_l^z



(d) Reco. p_ν^z vs Truth p_ν^z



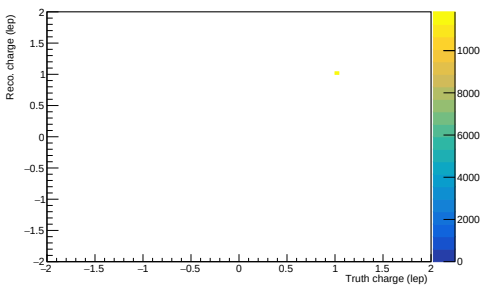
(e) Truth $\eta_l, \eta_\nu, \Delta\eta\ (l, \nu)$



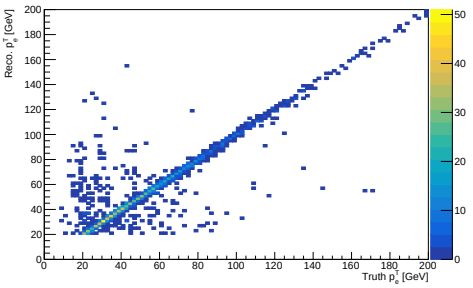
(f) Reco. $\eta_l, \eta_\nu, \Delta\eta\ (l, \nu)$

Figure 7.10: The distributions showing the correlations of important truth-matching quantities between the truth-level and the reco-level, as well as both of the levels themselves. The mc16a tZq sample at $\alpha = 0.5$ with the anti-top removal is used. The removal of the anti-top can be proven by Fig 7.11(a) which shows only the values corresponding to the charge of the light lepton from the top quark.

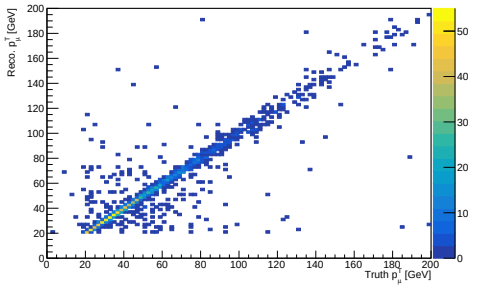
are not reconstructing the anti-top quark in this study), we checked if the charge of the reconstructed lepton matches that of the truth lepton. As shown in Fig 7.11(a), there is only one value (+1 charge for the lepton from the top quark) on each of the two levels, indicating that there is no any anti-top quark reconstructed. Fig 7.11(b) and 7.11(c) show the distribution of p_{T} of the electron and the muon respectively. As you can see, the p_{T} distribution of both of these light leptons draw an almost perfect line, and the combined p_{T} distribution of these light leptons is also showing the same behavior. This clearly indicates that there must be some unnecessary objects or not perfectly assigned leptons are biasing the reconstruction of the light lepton. However, this could not be the hadronically decaying τ lepton as it is identified and reconstructed in a separated way. Therefore, a further and closer look into the origin of this bias is required in the future. Apart from this truth-matching study on the $\alpha = 0.5$, the studies on different α values are performed and are presented in Appendix E.



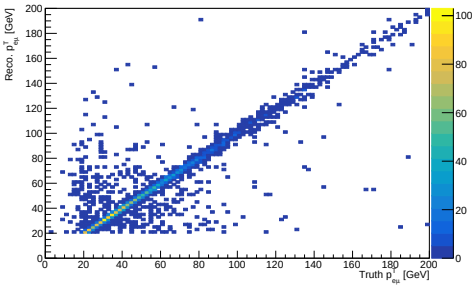
(a) Reco. l charge vs. Truth l charge



(b) Reco. p_e^{T} vs. Truth p_e^{T}



(c) Reco. p_μ^{T} vs. Truth p_μ^{T}



(d) Reco. $p_{(e,\mu)}^{\mathrm{T}}$ vs. Truth $p_{(e,\mu)}^{\mathrm{T}}$

Figure 7.11: The distributions showing the correlations of important truth-matching quantities between the truth-level and the reco-level, as well as both of the levels themselves. The mc16a tZq sample at $\alpha = 0.5$ with the anti-top removal is used. The removal of the anti-top can be proven by Fig 7.11(a) which shows only the values corresponding to the charge of the light lepton from the top quark.

Possible alternative approaches for probing p_ν^z

In order to find a more probable and reasonable value for the p_ν^z, more physically meaningful assumptions or innovative techniques are required. There are several diverse methods that have been used in previous neutrino reconstruction searches. In a search for a light charged Higgs boson ($H^\pm \to \tau^\pm \nu_\tau$), several selection methods for choosing a better p_ν^z of the neutrino from the top quark decay are made based on the maximum and minimum of p_ν^z, the 3-dimensional angle between the reconstructed τ lepton and the

neutrino pair, pseudo-rapidity difference between the reconstructed τ lepton and the neutrino pair. For the reconstruction of the neutrino momentum in the production of a pair of top quarks associated with Higgs boson production ($t\bar{t}H$) [79], instead of using traditional analytic method to calculate the missing neutrino momentum, a novel strategy which exploits a two-stage neural network is used, of which the first stage consists of the Lorentz Boost Network (LBN) and the second stage of a regression network. It states that the Neural Network based results are superior than the analytically obtained solutions, which is shown by the distribution of the neutrino momentum components, the correlation between calculated and simulation truth, as well as the correlation between predicted momentum and simulation truth. These above methods have not tried out in this thesis due to lack of time, however these could be tried out in the future.

Truth-matching study

Along with the truth studies on the p_ν^z, a truth-matching study has been carried out on the kinematic variables of the reconstructed particles, the neutrino, W boson and top quark. From Fig 7.12 to 7.15 show the distributions of p_T, η, ϕ, p_x, p_y and p_z of each reconstructed particle in the 2-solution case, since the 2-solution case accounts for the largest portion of the total number of events as mentioned in Section 7.2.4 (the light lepton distributions have no second solution as the light lepton is not reconstructed from the neutrino reconstruction algorithm).

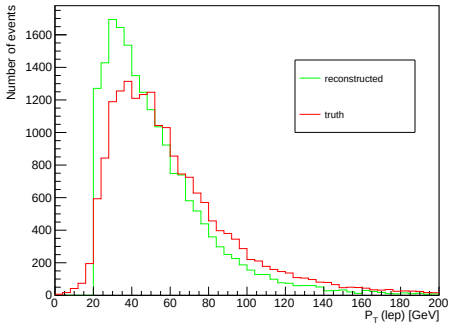
As can be seen from Fig 7.12, the distributions of the truth and reconstructed leptons are overall in a good agreement, except the p_T and p_z distributions. In the p_T distribution, the truth-level peak is more shifted to the right compared to the reco-level peak. Unlike the reconstructed p_T distribution, the truth distribution is not cut at $p_\mathrm{T} > 20$ GeV, since the truth leptons are not sorted by p_T and are not expected to pass lepton trigger selection criteria. Therefore, the events in the lower p_T region are expected to contribute to the shift of the truth distribution. This might explain why the peak of the p_z of truth distribution at zero is more smeared out than the reconstructed one, since high p_T collisions tend to have lower p_z according to the momentum conservation law. The small discrepancies are believed to originate from technical issues as well as detector effects.

Additionally, as illustrated in Fig 7.13 to 7.15, two real solutions exhibit some interesting behaviors. First, one of the two solutions in the ϕ distributions of the reconstructed particles prefers a positive ϕ value while the other one prefers a negative ϕ in a shape of sine waves with oppositely signed amplitude. However, the sum of these two solutions is almost equally distributed similar to the truth distribution. This can be thought to originate from the formula of p_ν^y itself, as shown in equation 7.13 and 7.14:

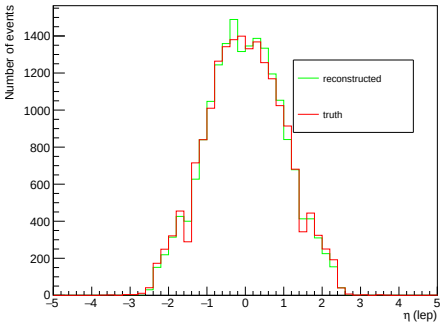
$$y_{1,2} = -\frac{B}{4A} - S \pm \frac{1}{2}\sqrt{-4S^2 - 2p + \frac{q}{S}}$$

$$y_{3,4} = -\frac{B}{4A} + S \pm \frac{1}{2}\sqrt{-4S^2 - 2p - \frac{q}{S}}$$

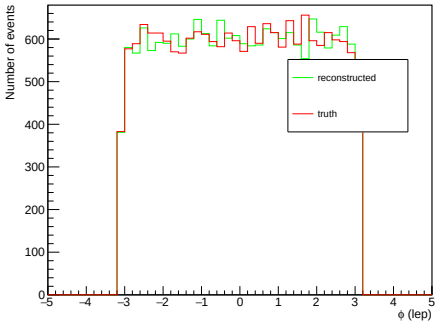
In each of both formulas, there is a square root term which can be either added (+) or subtracted (−), whereas the first two terms stay unaffected. Therefore, the two y values from each equation are expected to be subjected in a certain relation, instead of being totally irrelevant to each other. The distributions of p_x and p_y also exhibit similar behaviors since they are dependent on the ϕ and p_T. In the η distributions, the two reconstructed solutions are not showing a noticeable difference between each other, but they are quite far from the truth values and this might be due to the assumed p_z. Interestingly, all these effect are smeared out as the reconstruction procedure is proceeded, i.e. as the W boson and top quark are reconstructed.



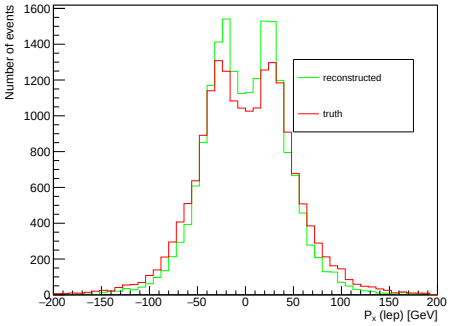
(a) p_{T}



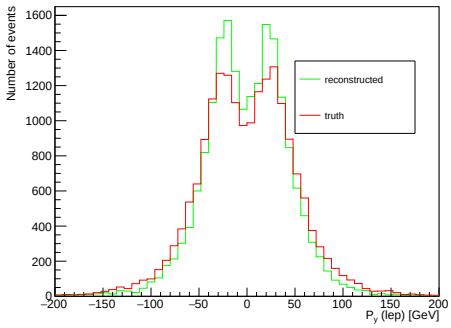
(b) η



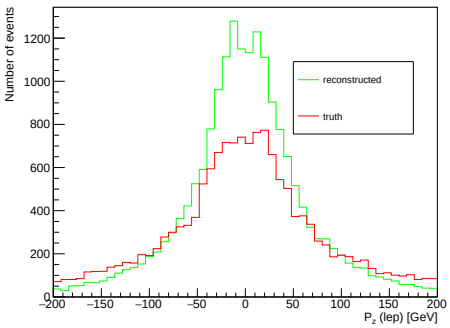
(c) ϕ



(d) p_x

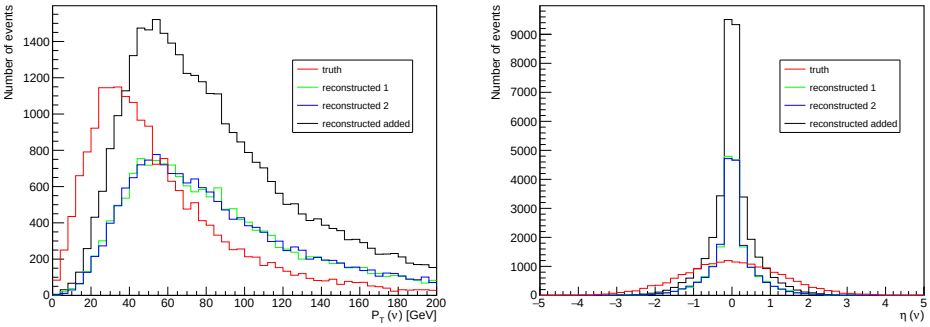


(e) p_y

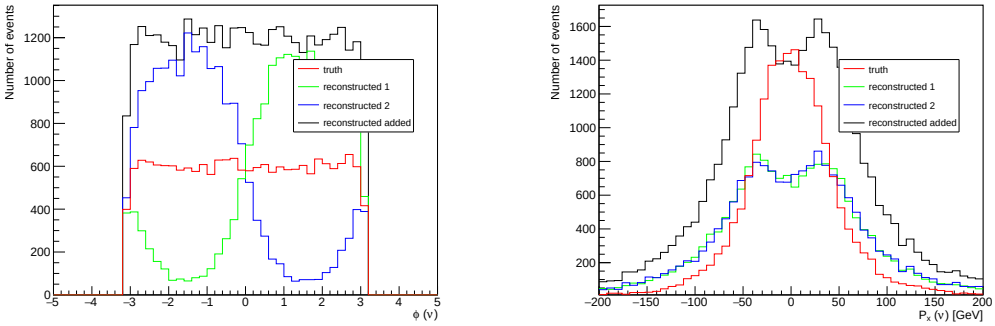


(f) p_z

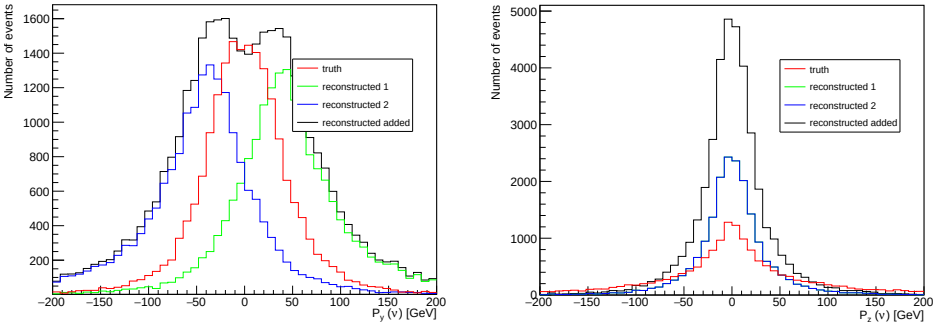
Figure 7.12: The distributions of kinematic variables of the reconstructed light lepton from the top quark. The red line represents the truth-level distributions, and the green line shows the reco-level distributions.



(a) p_{T} (b) η

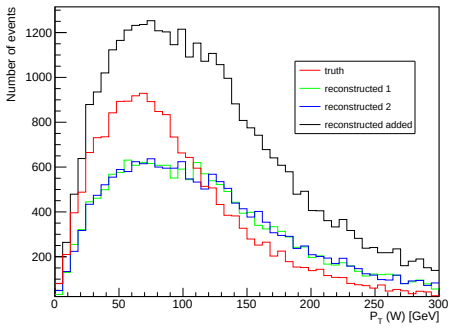


(c) ϕ (d) p_x

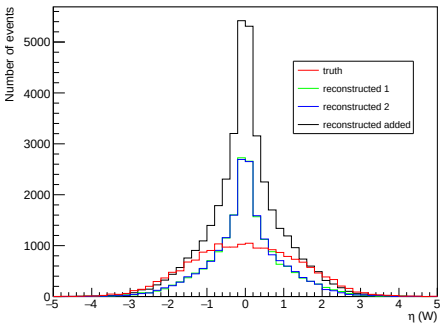


(e) p_y (f) p_z

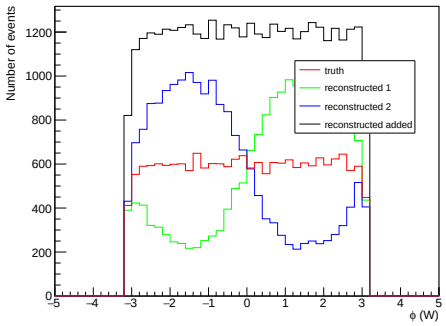
Figure 7.13: The distributions of kinematic variables of the reconstructed neutrino from the top quark with two real solutions. The red line represents the truth-level distribution. Each of the blue and green lines represents each of the two solutions on the reco-level. The black line is the addition of these two solutions.



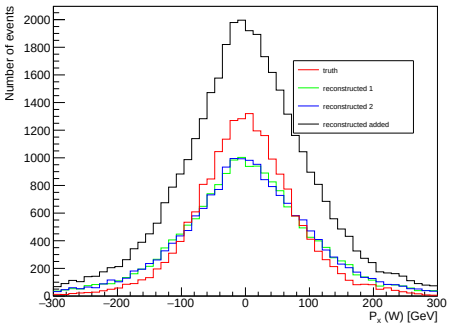
(a) p_{T}



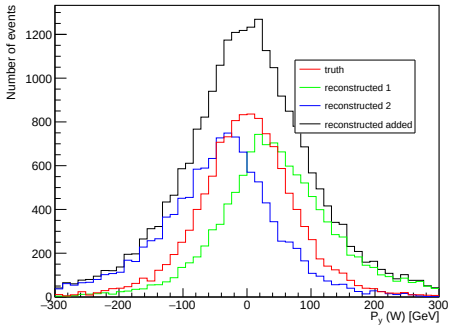
(b) η



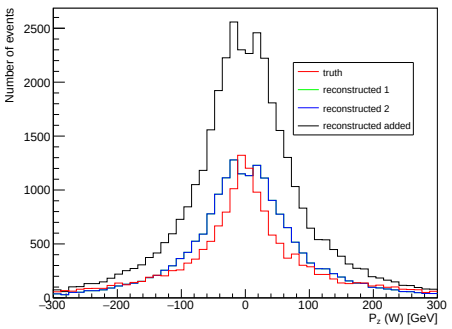
(c) ϕ



(d) p_x

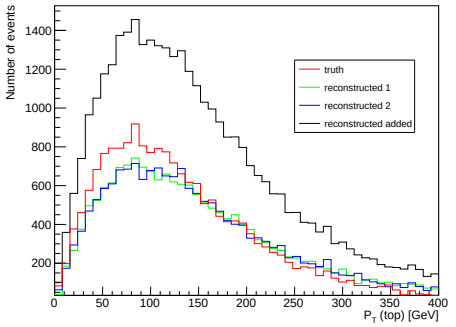


(e) p_y

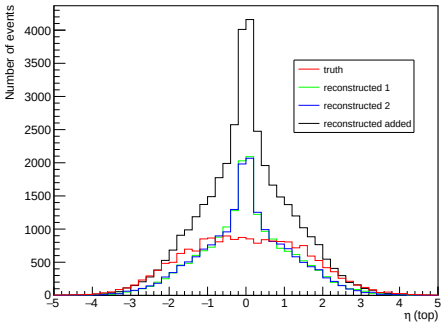


(f) p_z

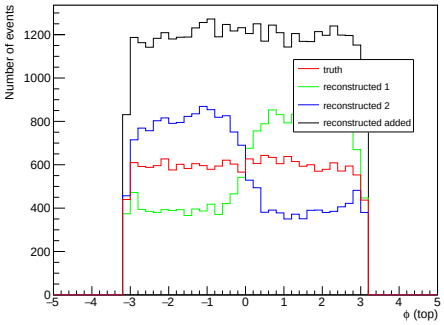
Figure 7.14: The distributions of kinematic variables of the reconstructed W boson from the top quark with two real solutions. The red line represents the truth-level distribution. Each of the blue and green lines represents each of the two solutions on the reco-level. The black line is the addition of these two solutions.



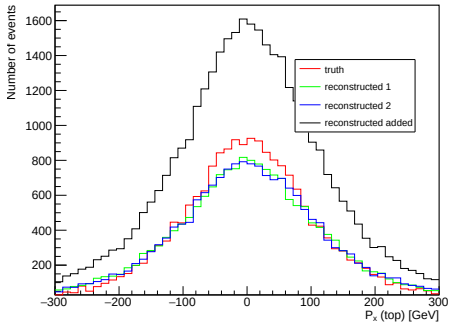
(a) p_{T}



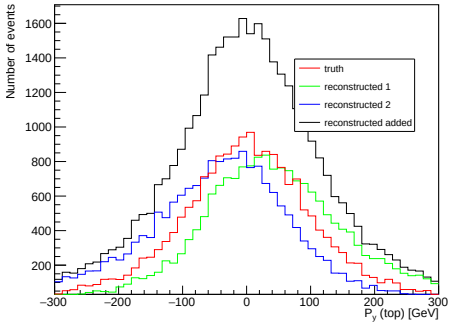
(b) η



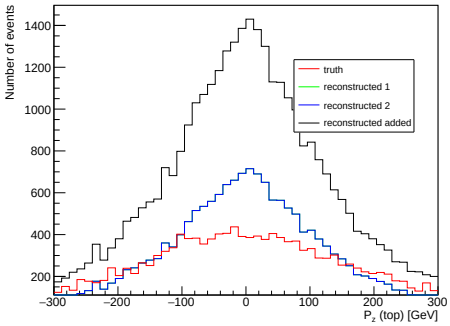
(c) ϕ



(d) p_x



(e) p_y



(f) p_z

Figure 7.15: The distributions of kinematic variables of the reconstructed top quark with two real solutions. The red line represents the truth-level distribution. Each of the blue and green lines represents each of the two solutions on the reco-level. The black line is the addition of these two solutions.

7.3 Reconstruction of the $\tau\tau$ system

In this section, the $\tau\tau$ system is reconstructed with the remaining $E_{\mathrm{T}}^{\mathrm{miss}}$ that is left after the subtraction of the $E_{\mathrm{T}}^{\mathrm{miss}}$ of the top quark from the total $E_{\mathrm{T}}^{\mathrm{miss}}$. Three different techniques that are used for reconstructing the $\tau\tau$ system are presented.

7.3.1 Missing transverse momentum uncertainties in the $\tau\tau$ system

An accurate measurement (or estimation) of the missing energy in the $\tau\tau$ system plays a key role in terms of reconstruction. As mentioned in Section 7.1, the missing energy in the $\tau\tau$ system is calculated as below:

$$\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}(\mathrm{Higgs}) = \overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}(\mathrm{total}) - \overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}(\mathrm{top})\,.$$

From the equation above, we can see that obviously there are two main contributing factors: the total measured missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$, as well as reconstructed $p_{\nu}^{x,y}$. The most dominating factor is surely the latter one, inaccurate assignment of the missing energy to the $\tau\tau$ system. This uncertainty is very tricky to be reduced due to limitations in the analytical method. However, it is also important to evaluate all possible uncertainties that contribute to the former one, since the reconstruction of the top quark system is also based on the measurement of the total $E_{\mathrm{T}}^{\mathrm{miss}}$. The missing energy measurements come with a considerable amount of experimental uncertainties as listed below [78] [80]:

- Non-linear response of the calorimeters
- Different response to electromagnetic and non-electromagnetic components of the hadron shower in the calorimeters
- Minimum energy, momentum thresholds of the calorimeters
- Inefficiencies in the tracker
- Difference between the measured direction of a calorimeter deposit from the original direction of the particle due to the bending of track in the magnetic field
- Underlying event activity
- Pile-up
- Detector noises
- Limited detector acceptance

In our analysis, we assumed that the effects of detector resolution and other experimental limitations are under good control.

7.3.2 The partial invariant mass method

One of the simplest and frequently used methods for reconstructing the $\tau\tau$ system is to reconstruct partial (or reduced) invariant mass. There are two different defined partial invariant masses: 1) the visible mass, which is the invariant mass of visible τ decay products, and 2) the transverse mass, which is the invariant

mass of the visible τ decay products and the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ of the $\tau\tau$ system. The transverse mass, usually denoted as m_{T}, can be expressed as below [80]:

$$m_{\mathrm{T}}^2 = m^2(\tau_{\mathrm{vis}_1}, \tau_{\mathrm{vis}_2}, E_{\mathrm{T}}^{\mathrm{miss}}) = P^{\mu}P_{\mu}\,, \tag{7.21}$$

where P^{μ} is defined as:

$$P^{\mu} = P^{\mu}(\tau_{\mathrm{vis}_1}) + P^{\mu}(\tau_{\mathrm{vis}_2}) + p^{\mu}(E_{\mathrm{T}}^{\mathrm{miss}})\,, \tag{7.22}$$

which is the sum of 4-momentum vector of the visible decay products of τ lepton $i = 1, 2$ and 4-momentum vector of the missing transverse momentum. The transverse mass provides a better separation of the signal from the QCD multi-jet backgrounds with fake τ signatures. Fore example, as stated in the research [81], the W+jets process can be reduced by constraining the transverse mass to satisfy certain requirements such as to be less than 50 GeV. In particular, it is beneficial to reconstruct the transverse mass for the cases with leptonically decaying τ leptons, since the transverse mass can be expressed as:

$$m_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^{l}p_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos(\Delta\phi(l, \nu)))} \tag{7.23}$$

where p_{T}^{l} and $p_{\mathrm{T}}^{\mathrm{miss}}$ are the transverse momentum of the lepton and the invisible decay product from the τ lepton, respectively. The $\Delta\phi(l, \nu)$ is the azimuthal angle between the direction of momentum vector $\vec{p}_{\mathrm{T}}^{\,l}$ and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$.

The visible mass can be obtained by the same method but without taking the missing mass energy into account [80]:

$$m_{\mathrm{vis}}^2 = m^2(\tau_{\mathrm{vis}_1}, \tau_{\mathrm{vis}_2}) = P^{\mu\prime}P'_{\mu},\,, \tag{7.24}$$

where $P^{\mu\prime}$ is defined as:

$$P^{\mu\prime} = P^{\mu}(\tau_{\mathrm{vis}_1}) + P^{\mu}(\tau_{\mathrm{vis}_2})\,. \tag{7.25}$$

The main advantage of reconstructing the visible mass is that it can be applied to all signal event topologies for the $H \to \tau\tau$ channel, i.e. the $\tau_h\tau_h$, $\tau_l\tau_h$ and $\tau_l\tau_l$ channels. Moreover, the visible decay products have better measured parameters, providing the mass resonance with a high accuracy. However, since it does not fully account for the momentum bias of the invisible decay products, the partially reconstructed invariant mass distribution is significantly broadened, resulting in a severely decreased sensitivity. This issue can be easily seen in the low mass region of the invariant mass of $\tau\tau$, where the $Z \to \tau\tau$ background is predominant with its larger and broader mass distribution (see [80]).

7.3.3 Collinear apporoximation method

The collinear approximation is also known as one of the most commonly used technique for the reconstruction of the invariant mass of the $\tau\tau$ system with the presence of invisible decay products. It is mainly based on two assumptions:

1. The τ lepton and all its decay products are collinear, i.e. $\phi_{\nu} = \phi_{\mathrm{vis}}$ and $\theta_{\nu} = \theta_{\mathrm{vis}}$.

2. The source of missing transverse momenta is due to the neutrinos only.

The visible decay products in each τ decay carry some fraction of the τ lepton momenta, which is denoted in x_i in formula 1 [82], and f_i in formula 2 [80], with τ lepton $i = 1, 2$. These unknown fractions can be determined using missing transverse momenta as well as measured transverse momenta constraints, and as a consequence, full reconstruction of the di-tau system is possible.

However, the collinear approximation has some significant shortcomings. This approximation is only valid for events where the $\tau\tau$ system is boosted, i.e. produced in association with hard jet(s) with large $E_{\mathrm{T}}^{\mathrm{miss}}$, such that the visible τ decay products are not back-to-back in the plane transverse to the beam line. This is because when two τ leptons are produced back-to-back, the missing momentum associated with their neutrinos partially cancels out, making the invariant mass of a resonance cannot be directly reconstructed from the $E_{\mathrm{T}}^{\mathrm{miss}}$. The majority of Higgs boson production in LHC is done via $gg \to HH \to \tau\tau$ channel, in which τ leptons are nearly produced in the back-to-back topology, thus only a relatively small fraction of events are available for this technique. Fortunately, the signal event in this analysis, tHq, has additional jets which are produced along with the Higgs boson, i.e. the b-jet(s) as well as forward jet(s), allowing this technique to be applicable for this analysis. However, the collinear approximation method is very sensitive to the resolution of $E_{\mathrm{T}}^{\mathrm{miss}}$ and is likely to over-estimate the mass resonance of the $\tau\tau$ system, leaving a long tail in the reconstructed mass distribution. This long-tail effect will considerably affect accurate measurements in low-mass Higgs boson searches, where the tails of a much larger $Z \to \tau\tau$ background is completely dominant in the expected Higgs mass peak region [80, 82].

In the following, two different formulas are exploited for the comparison of reconstruction accuracy of two collinear approximation methods. Nevertheless, the main idea of the two formulas is the same, as both of them start from the same assumptions stated before.

Formula 1

The first formula starts from two equations as below:

$$E_{\mathrm{T}_x}^{\mathrm{miss}} = p_{\mathrm{mis1}} \sin\theta_{\mathrm{vis1}} \cos\phi_{\mathrm{vis1}} + p_{\mathrm{mis2}} \sin\theta_{\mathrm{vis2}} \cos\phi_{\mathrm{vis2}} \,, \tag{7.26}$$

$$E_{\mathrm{T}_y}^{\mathrm{miss}} = p_{\mathrm{mis1}} \sin\theta_{\mathrm{vis1}} \sin\phi_{\mathrm{vis1}} + p_{\mathrm{mis2}} \sin\theta_{\mathrm{vis2}} \sin\phi_{\mathrm{vis2}} \,, \tag{7.27}$$

where $E_{\mathrm{T}_x}^{\mathrm{miss}}$ and $E_{\mathrm{T}_y}^{\mathrm{miss}}$ are the x- and y-components of a 2-dimension vector $\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}} = (\overrightarrow{E_{\mathrm{T}_x}^{\mathrm{miss}}}, \overrightarrow{E_{\mathrm{T}_y}^{\mathrm{miss}}})$ (vectorial expressions for some variables will be neglected for convenience); p_{mis1} and p_{mis2} are the inclusive missing momenta (there can be two neutrinos in one τ decay) from each τ decay; $\theta_{\mathrm{vis1,2}}$ and $\phi_{\mathrm{vis1,2}}$ are the polar and azimuthal angles of the visible products from each τ decay. Then the invariant mass of the $\tau\tau$ system, denoted as $M_{\tau\tau}$, can be calculated as follows:

$$M_{\tau\tau} = \frac{m_{\mathrm{vis}}}{\sqrt{x_1 x_2}} \,, \tag{7.28}$$

where m_{vis} is referred to the invariant mass of visible decay products in the $\tau\tau$ system. The $x_{1,2}$ here is defined as:

$$x_{1,2} = p_{\mathrm{vis1,2}} / (p_{\mathrm{vis1,2}} + p_{\mathrm{mis1,2}}) \,, \tag{7.29}$$

which is the momentum fraction that is carried away by visible products in each τ decay.

The $p_{\mathrm{mis1,2}}$ can be extracted, as we know the total missing transverse momenta assigned to the $\tau\tau$ system after full reconstruction of the top quark, by the assumption $\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}(\mathrm{Higgs}) = \overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}(\mathrm{total}) - \overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}(\mathrm{top})$. We can rewrite the equation 7.26 and equation 7.27 in terms of p_{mis1} and p_{mis1}:

$$p_{\mathrm{mis1}} = \frac{E_{\mathrm{T}_y}^{\mathrm{miss}} \cos\phi_{\mathrm{vis2}} - E_{\mathrm{T}_x}^{\mathrm{miss}} \sin\phi_{\mathrm{vis2}}}{\sin\theta_{\mathrm{vis1}} (\sin\phi_{\mathrm{vis2}} \cos\phi_{\mathrm{vis1}} - \cos\phi_{\mathrm{vis1}} \sin\phi_{\mathrm{vis2}})} \,, \tag{7.30}$$

$$p_{\text{mis2}} = \frac{E^{\text{miss}}_{\text{T}_x} \sin\phi_{\text{vis1}} - E^{\text{miss}}_{\text{T}_y} \cos\phi_{\text{vis1}}}{\sin\theta_{\text{vis2}}(\sin\phi_{\text{vis2}}\cos\phi_{\text{vis1}} - \cos\phi_{\text{vis1}}\sin\phi_{\text{vis2}})}\,. \tag{7.31}$$

As we can see from these equations, the variables which directly affect the $p_{\text{mis1,2}}$ and thus the momentum fraction $x_{1/2}$, are the missing transverse momentum components, $E^{\text{miss}}_{\text{T}_x}$ and $E^{\text{miss}}_{\text{T}_y}$. Thus, how close the missing transverse momentum $\tau\tau$ system is assigned to its true value will directly determine the resolution of its mass resonance. The reconstruction efficiency of the visible decay products from each τ decay contributes relatively less to the resolution of the mass peak [80].

Formula 2

Starting from the same idea that was used for the first formula, the neutrino momentum can be also written in terms of the visible particle momentum as below:

$$\overrightarrow{q_i} = \overrightarrow{p_{\tau_i}} - \overrightarrow{p_i} = F_i\overrightarrow{p_i}\,, \quad F_i = \frac{1}{f_i} - 1\,, \tag{7.32}$$

which can be generally written as:

$$\overrightarrow{E^{\text{miss}}_{\text{T}}} = \sum_i F_i\overrightarrow{p_{i\text{T}}}\,, \tag{7.33}$$

where $\vec{q_i}$ and $\vec{p_i}$ are the momenta of invisible and visible decay products from each τ_i decay respectively; $\vec{p_{\tau_i}}$ is the total momenta of τ_i; and F_i is the momentum fraction carried by the visible product in the τ_i decay. The unknown fractions denoted as f_i is defined as:

$$f_1 = \frac{1}{1+r_2}\,, \quad f_2 = \frac{1}{1+r_1}\,, \tag{7.34}$$

where

$$r_i = \left|\frac{E^{\text{miss}}_{\text{T}_y} p_i^x - E^{\text{miss}}_{\text{T}_x} p_i^y}{p_1^y p_2^x - p_1^x p_2^y}\right|. \tag{7.35}$$

The invariant mass of the $\tau\tau$ system is then calculated with the momentum fraction f_1 and f_2 as follow:

$$M_{\tau\tau} = \frac{m_{\text{vis}}}{\sqrt{f_1 f_2}}\,. \tag{7.36}$$

7.3.4 The Missing Mass Calculator technique

A new reconstruction technique named Missing Mass Calculator (MMC) has been recently proposed in a search for resonance decaying to $\tau\tau$ [80]. As mentioned in the previous section, the partial mass reconstruction method and collinear approximation method have significant drawbacks. The transverse mass reconstruction gives a very broad mass spectrum that decreases sensitivity in the low mass region, and the collinear approximation is limited only to events in which the $\tau\tau$ system is boosted. As compared to the two methods above, the MMC technique does not suffer from any of these limitations, and can be applied to any $\tau\tau$ event topologies and allows for a complete reconstruction of the event kinematics of the $\tau\tau$ final states with significantly improved resolution of the invariant mass and missing transverse momentum [80].

The theoretical concept of MMC

Like other reconstruction methods, MMC begins with assuming that there are no other sources for missing transverse momentum apart from the neutrinos from τ lepton decays. Under such an assumption of a perfect detector resolution, the event topology can be fully reconstructed by solving from 6 to 8 unknown variables depending on the number of leptonically decaying τ leptons: x-, y- and z-components of the invisible momentum carried away by neutrino(s) in each τ decay. Applying the mass constraints of τ leptons with the previously mentioned assumption that $E_{\mathrm{T}}^{\mathrm{miss}}$ originates only from neutrinos (the equation 7.26 and 7.27), there are four equations that can be used for restricting unknowns:

$$E_{\mathrm{T}_x}^{\mathrm{miss}} = p_{\mathrm{mis1}} \sin\theta_{\mathrm{vis1}} \cos\phi_{\mathrm{vis1}} + p_{\mathrm{mis2}} \sin\theta_{\mathrm{vis2}} \cos\phi_{\mathrm{vis2}} \,,$$

$$E_{\mathrm{T}_y}^{\mathrm{miss}} = p_{\mathrm{mis1}} \sin\theta_{\mathrm{vis1}} \sin\phi_{\mathrm{vis1}} + p_{\mathrm{mis2}} \sin\theta_{\mathrm{vis2}} \sin\phi_{\mathrm{vis2}} \,,$$

$$M_{\tau_1}^2 = m_{\mathrm{mis1}}^2 + m_{\mathrm{mis1}}^2 + 2\sqrt{p_{\mathrm{vis1}}^2 + m_{\mathrm{vis1}}^2}\sqrt{p_{\mathrm{mis1}}^2 + m_{\mathrm{mis1}}^2} - 2p_{\mathrm{vis1}}p_{\mathrm{mis1}}\cos\theta_{\nu m_1} \,, \tag{7.37}$$

$$M_{\tau_2}^2 = m_{\mathrm{mis2}}^2 + m_{\mathrm{mis2}}^2 + 2\sqrt{p_{\mathrm{vis2}}^2 + m_{\mathrm{vis2}}^2}\sqrt{p_{\mathrm{mis2}}^2 + m_{\mathrm{mis2}}^2} - 2p_{\mathrm{vis2}}p_{\mathrm{mis2}}\cos\theta_{\nu m_2} \,, \tag{7.38}$$

where $E_{\mathrm{T}_x}^{\mathrm{miss}}$ and $E_{\mathrm{T}_y}^{\mathrm{miss}}$ are the x- and y-components of the $\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}$ vector. Here, $p_{\mathrm{mis1,2}}$, $m_{\mathrm{mis1,2}}$, $\theta_{\mathrm{mis1,2}}$ and $\phi_{\mathrm{mis1,2}}$ are the momenta (vector), invariant masses, polar and azimuthal angles of the invisible τ decay products in τ lepton $i = 1, 2$. Similarly, $p_{\mathrm{vis1,2}}$, $m_{\mathrm{vis1,2}}$, $\theta_{\mathrm{vis1,2}}$ and $\phi_{\mathrm{vis1,2}}$ are those of the visible τ decay products in each τ decay. The τ_i lepton invariant mass is set to $M_{\tau_{1,2}} = 1.77\ \mathrm{GeV}/c^2$. $\Delta\theta_{\nu m_{1,2}}$ is the angle between the vectors $\vec{p}_{\mathrm{mis}}$ and $\vec{p}_{\mathrm{vis}}$ (it can be expressed in terms of other variables) for each τ decay. For the hadronically decaying τ leptons, the number of unknowns can be reduced by one for each of these decays, as only one neutrino is involved in the hadronic τ decay, and thus the variable m_{mis} can be set to be 0.

The number of unknowns need to be solved is from 6 to 8 depending on the number of leptonic τ decays, whereas the maximum number of available constraints is only 4. This means the existing information is not sufficient to obtain the exact solution. However, not all solutions in this under-constrained system seem to be equally probable in a certain event topology. Some additional information such as decay kinematics can provide extra constraints and allows better estimation of the mass resonance by choosing more likely solutions over less likely ones. In the $H \to \tau\tau$ event topology, the τ decay kinematics can provide additional information about $M_{\tau\tau}$ to extract more probable solutions. In order to evaluate the probability of such a decay topology, the decay kinematics need to be incorporated as probability functions in a so-called global event probability, denoted as $\mathcal{P}_{\mathrm{event}}$. The main idea of the MMC technique is to find the most probable $M_{\tau\tau}$ solution by producing the $M_{\tau\tau}$ distribution for all scanned points in a particular parameter phase (grid) that is weighed by its corresponding global event probability $\mathcal{P}_{\mathrm{event}}$. The maximum point of the $M_{\tau\tau}$ distribution is taken as the final estimator for a given event. One example of such additional constraint, $\Delta R_{1,2}$, the expected angular distance between the invisible and visible τ decay products is described in [80], and will be shortly explained in the next section.

The description of the MMC method

A parameter $\Delta R = \sqrt{(\eta_{\mathrm{vis}} - \eta_{\mathrm{mis}})^2 + (\theta_{\mathrm{vis}} - \theta_{\mathrm{mis}})^2}$ is chosen for the validation procedure of MMC in [80]. Although choosing a 3-dimensional angle between the neutrino(s) and visible decay products might make more sense, a 2-dimensional angle is chosen for simplicity. For hadronically decaying τ lepton cases, the system of equations 7.26, 7.27, 7.37 and 7.38 can be solved exactly for any point in, for

example, a 2-dimensional parameter space of a pair of parameters (θ_{mis_1},θ_{mis_2}). In other words, all possible solutions of the missing transverse momentum $\vec{p}_{\text{mis}_{1,2}}$ from these four equations, can be broken down into a uniquely and fully defined point on the grid. Therefore, for each point on the grid of (θ_{mis_1}, θ_{mis_2}), the parameter $\Delta R_{1,2}$, the distance between $\vec{p}_{\text{vis}_{1,2}}$ and $\vec{p}_{\text{mis}_{1,2}}$, is calculable. The next step is to evaluate the probability of such a decay topology using this parameter $\Delta R_{1,2}$. To achieve it, a probability density function based on the dependency of this chosen kinematic variable needs to be defined first. Ignoring the polarization of the τ lepton, the ΔR distribution depends only on the initial momentum p_τ and the decay type of τ lepton, but not on its source. Therefore, the ΔR distribution is scanned in the initial τ momentum in the range of 10 GeV $< p_\tau <$ 100 GeV with a bin size of 5 GeV, in simulated $Z/\gamma* \to \tau\tau$ events. It is further parameterized by being fitted with a linear combination of Gaussian and Landau functions, yielding a fully parameterized p_τ-dependent probability distribution $\mathcal{P}(\Delta R, p_\tau)$. To integrate this information as an additional constraint, the global event probability is defined as follows:

$$\mathcal{P}_{\text{event}} = \mathcal{P}(\Delta R_1, p_{\tau_1}) \times \mathcal{P}(\Delta R_2, p_{\tau_1}) \tag{7.39}$$

where the functions $\mathcal{P}$ are chosen according to one of the decay types, and uniquely defined for each point on the parameter space of (θ_{mis_1}, θ_{mis_2}). In order to find the best estimator of the invariant mass of $\tau\tau$, an $M_{\tau\tau}$ value is produced for each of all points on the grid, and then the obtained $M_{\tau\tau}$ distribution is weighed by a corresponding probability $\mathcal{P}_{\text{event}}$. The maximum point of the $M_{\tau\tau}$ distribution is the final estimator of the $\tau\tau$ invariant mass. Apart from the initial τ lepton momentum, other kinematic variables can also be used depending on one's desire.

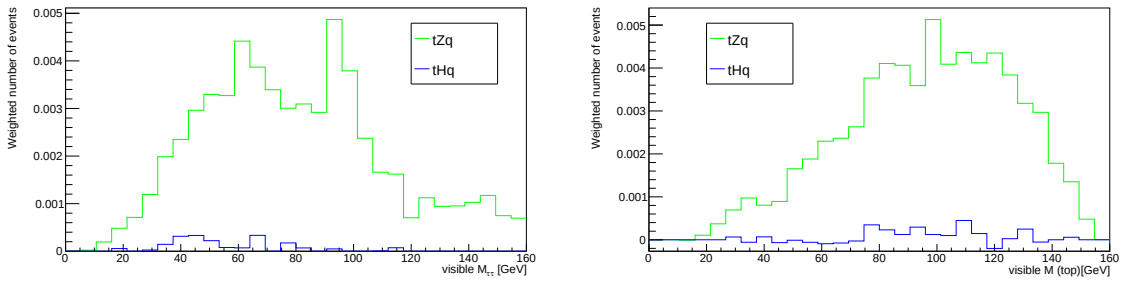
In the case of the event topology involving at least one leptonically deaying τ lepton (lep-had or lep-lep channel), the previous procedure needs to be adjusted to take the invariant mass m_{mis} into account, as m_{mis} is unknown if there are more than one neutrino produced in one τ decay. A phase space with higher dimensionality needs to be scanned, for example, (θ_{mis_1}, θ_{mis_2}, m_{mis_1}) when one τ lepton decays leptonically, and (θ_{mis_1}, θ_{mis_2}, m_{mis_1}, m_{mis_2}) if both of the τ leptons decay leptonically. One can assume an uniform probability for parameter m_{mis} for convenience, but scanning according to the m_{mis} probability function will improve the estimation performance.

7.3.5 Comparison of three reconstruction techniques

In the following, the results of the invariant mass of the $\tau\tau$ system reconstructed with the previously mentioned three different reconstruction techniques will be shown.

Reconstructed partial mass

The transverse mass of the $\tau\tau$ system is not reconstructed in this analysis due to the limitations on the decay channel of the $\tau\tau$ system. This is because, in order to reconstruct the transverse mass, leptonically decaying τ leptons are required as can be seen from equation 7.23 (this case is limited to the tHq process). The invariant mass of the visible decay products in the $\tau\tau$ system as well as the top quark system are shown in Fig 7.16. As can be seen from Fig 7.16(a), a little bump around 90 GeV can be found, which is expected to be a statistical fluctuation. The visible mass distribution is roughly peaked at 60 GeV. Fig 7.16(b) shows the reconstructed visible top quark mass. Compared to the visible $\tau\tau$ mass distribution, that of the top quark has a more stable shape without noticeable fluctuations. Unfortunately, the tHq samples suffer from significant low statistics.



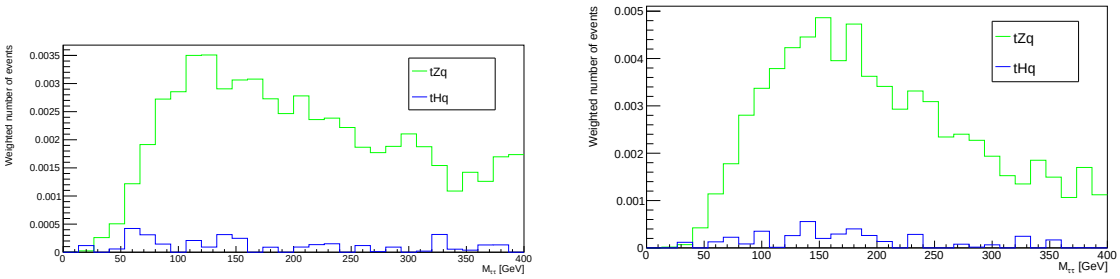
(a) (b)

Figure 7.16: The reconstructed invariant mass of visible decay products of the $\tau\tau$ system as well as of the top quark system for one (solution 1 here) of the two solutions in the 2-real-solution case. The green line shows the distribution of the tZq sample and the blue line shows the tHq sample.

Reconstructed invariant mass with the collinear approximation method

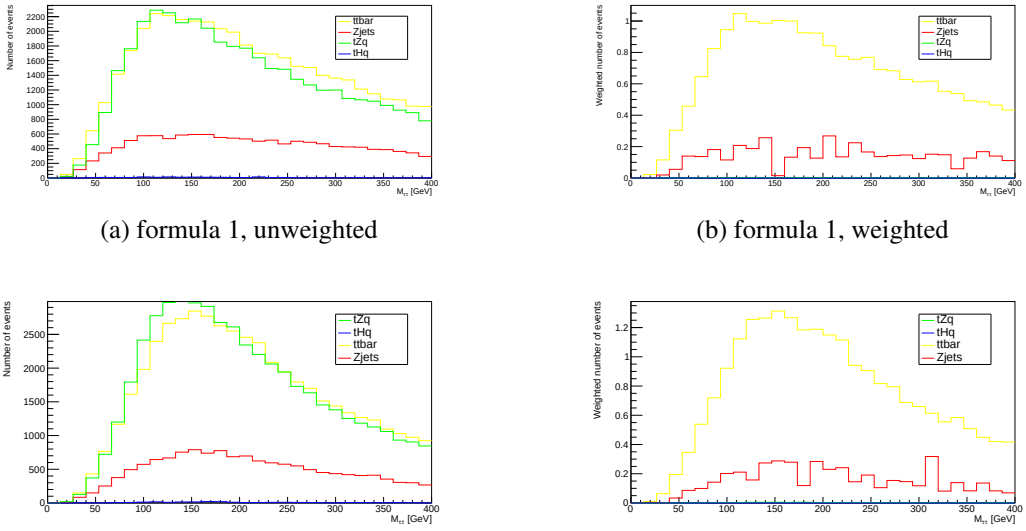
The reconstructed invariant mass distributions of the $\tau\tau$ system with the collinear method are shown in this section. Fig 7.17 shows the invariant mass of the $\tau\tau$ system reconstructed with two different formulas used in the collinear method for the tZq and tHq samples. As you can see from both plots, the reconstructed invariant mass in the tZq sample has a long tail with a peak at around 120 GeV for the formula 1 and 150 GeV for the formula 2. The invariant mass peak is expected to be located near the actual Z boson mass which is around $\sim$ 91 GeV (for the tHq sample it is $\sim$ 125 GeV), but the obtained peaks seem to be off from this value. As mentioned in Section 7.3.3, the collinear method is very sensitive to the resolution of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and tends to over-estimate the mass resonance. Therefore, we can conclude that the assigned $E_{\mathrm{T}}^{\mathrm{miss}}$ to the Higgs system is not perfect, and again a new method of feeding an accurate $E_{\mathrm{T}}^{\mathrm{miss}}$ is strongly needed. The tHq samples suffer from the low statistics and the mass resonance is barely recognizable in these samples. The difference between these two formulas is also about $\sim$ 30 GeV which is quite huge, and we can interpret this that a different way of calculating the momentum fraction carried by the visible products in the τ leptons can lead to a different invariant mass result. The big difference might originate from the imperfect assignment of the $E_{\mathrm{T}}^{\mathrm{miss}}$ to the Higgs system.

Fig 7.18 shows both the unweighted and weighted $\tau\tau$ invariant mass distributions with the tZq, tHq, Z+jets as well as $t\bar{t}$ samples. As you can see in all the plots in Fig 7.18, apart from the tHq sample, the Z+jets and $t\bar{t}$ samples also have a peak at around the same value that is mentioned previously (around 120 GeV for the formula 1 and 150 GeV for the formula 2), which indicates that the reconstruction algorithm is somehow forcing the $\tau\tau$ invariant mass to have a similar value for all these samples. Here, note that we are not picking up one solution of which the invariant mass is close to the top quark mass out of two solutions in the 2-solution cases (most of the multiple solution cases are the 2-solution cases), but we are reconstructing both of them in parallel and thus there is no biases in terms of the forced invariant mass (actually no big difference can be found between the solution 1 and 2 as shown in Fig 7.19). Therefore, a further investigation on how the reconstruction algorithm should be improved to discriminate different event topology is required.



(a) (b)

Figure 7.17: The reconstructed invariant mass of the $\tau\tau$ system using two different collinear formulas for the tZq and tHq samples including all campaigns. Fig 7.17(a) shows the invariant mass obtained with formula 1, and Fig 7.17(b) shows the one with formula 2.



(a) formula 1, unweighted (b) formula 1, weighted

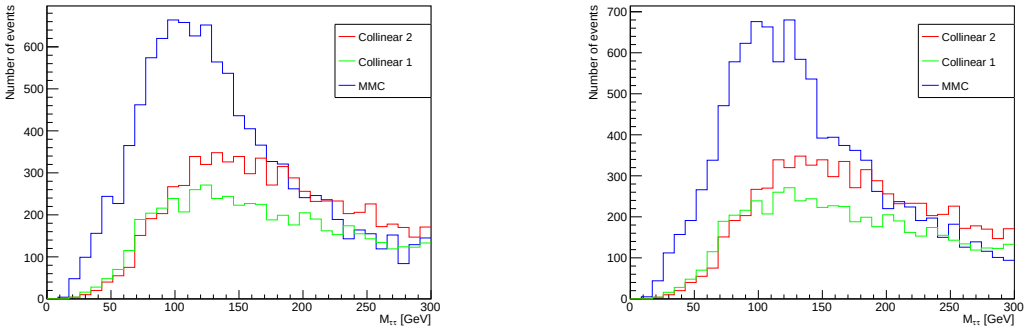
(c) formula 2, unweighted (d) formula 2, weighted

Figure 7.18: The reconstructed invariant mass of the $\tau\tau$ system using two different collinear formulas for the tZq, tHq, Z+jets as well as $t\bar{t}$ samples including all campaigns. The plots on the left side show the unweighted distributions and those on the right side show the weighted distributions. The tHq samples have too low statistics too be recognizable, and the tZq samples can be barely seen in the weighted plots.

Reconstructed invariant mass with the MMC

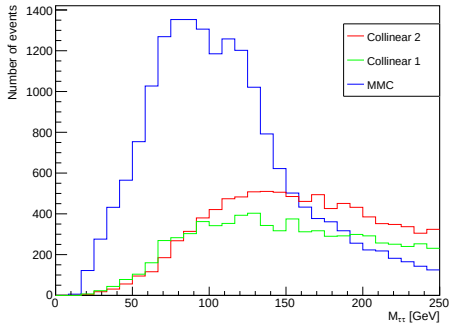
The reconstructed invariant mass distributions of the $\tau\tau$ system with the MMC technique as well as a comparison with the colinear method are shown in this section. Fig 7.19 shows the reconstructed $\tau\tau$ mass of both solutions (solution 1 and 2) in the 2-real-solution case using the MMC and two different collinear formulas for the tZq sample. Interestingly, as you can see, there is no noticeable but only small difference between the two solutions, and the peak in both of the plots is located around $\sim$ 100 GeV. This indicates that each of the two solutions in the 2-real-solution case can have a different value of $E_{\mathrm{T}}^{\mathrm{miss}} = \sqrt{(p_\nu^x)^2 + (p_\nu^y)^2}$ and $\overrightarrow{E_{\mathrm{T}}^{\mathrm{miss}}}$, but this difference is not big enough to see the different contributions of the two distinct solutions to the $\tau\tau$ invariant mass. Apart from this, as can be seen from the figures, obviously the MMC technique provides a much better and accurate mass resonance than the other collinear methods (formula 1 and 2). The peak of the MMC distribution is closer to the real Z mass ($\sim$ 91 GeV), whereas the other two distributions are rather broader than the MMC one and are shifted to a higher mass region. This might be related to the techniques themselves, as the collinear approximation is not a robust method for estimating an accurate invariant mass when the $E_{\mathrm{T}}^{\mathrm{miss}}$ is not very reliable.

Other than the peak location, the resolution of the mass resonance is much more improved in the MMC distribution, which has a higher and narrower peak than the two collinear approximation distributions. Furthermore, as you can see from Fig 7.20, the MMC distribution has a way shorter tail throughout the shown mass ranges. The figure shows the weighted and unweighted reconstructed $\tau\tau$ invariant mass for the tZq and tHq samples, and some fluctuations or an underflow ($\sim$ 30 GeV) of the tHq samples due to its low statistics can be noticed. In summary, we can conclude that the MMC method is a much more efficient and remarkable way for the reconstruction of the invariant mass of the Higgs boson than the other two methods, partial reconstruction method and the collinear approximation method. However, a more accurate calculation and estimation of the $E_{\mathrm{T}}^{\mathrm{miss}}$ of the Higgs system is strongly required in order to make a good use of this excellent MMC technique.

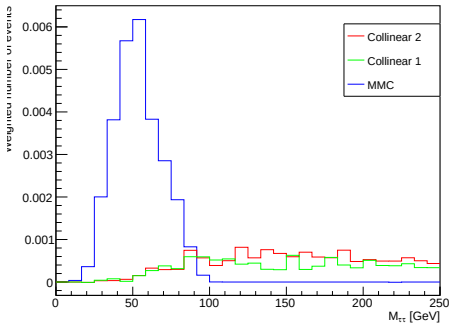


(a) (b)

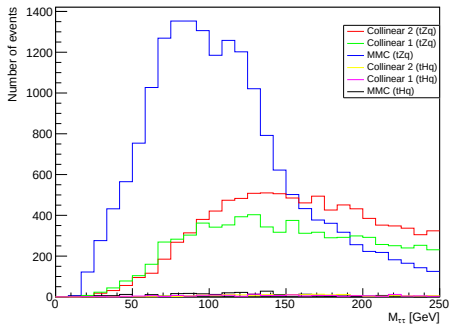
Figure 7.19: The reconstructed $\tau\tau$ invariant mass of both solutions (solution 1 and 2) in the 2-real-solution case for the tZq sample. Fig 7.19(a) shows the distribution for the solution 1, and Fig 7.19(b) shows the solution 2. The mc16a tZq sample is used.



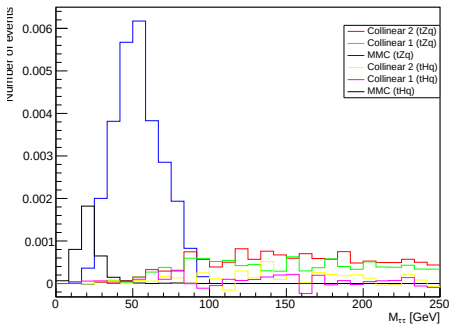
(a) Unweighted $M_{\tau\tau}$ for tZq



(b) Weighted $M_{\tau\tau}$ for tZq



(c) Unweighted $M_{\tau\tau}$ for tZq, tHq



(d) Weighted $M_{\tau\tau}$ for tZq, tHq

Figure 7.20: The weighted and unweighted reconstructed $\tau\tau$ invariant mass for one of two solutions (solution 1) in the 2-real-solution case for the tZq and tHq samples.

CHAPTER 8

Conclusion

In this thesis, a study on the production of a single top quark in association with a Higgs boson (tHq) focusing on the Higgs decay channel $H \to \tau_l\tau_h$ (lep-had channel) is presented. The study is based on the data of pp collisions at $\sqrt{s}$ = 13 GeV recorded by the ATLAS detector during Run 2 (2015 - 2018), which corresponds to an integrated luminosity of 139 fb^{-1}.

Unlike other top quark productions associated with a Higgs boson such as $t\bar{t}H$, whose sensitivity is only limited to the magnitude of the Higgs coupling constants y_t (with the top quark) and g_{HVV} (with the W bosons), tHq is also sensitive to the sign of the coupling constants. In the lep-had channel, the Higgs boson decays into two τ leptons, of which one decays hadronically and the other one decays leptonically, while the top quark decays leptonically. There can be different origin of the light leptons in the lep-had final state, triggering ambiguities in terms of accurate final state reconstruction. Four neutrinos in the final state also throws a great challenge to full reconstruction of the top quark and the Higgs boson, which can be used as discriminating variables for signal-background separation. Therefore, this thesis has searched for some possible solutions to resolve these problems.

Prior to the tHq study, the multilepton (ML) channel of the production of a single top quark associated with a Z boson (tZq) has been investigated. This study is generally useful for the tHq study, since it not only constitutes one of the backgrounds to the tHq process and thus allows for direct background estimation, but also can be used for the validation of the framework (`tHqLoop`) which is used for the tHq study. For this, the event yields as well as the distributions of kinematic variables obtained in two signal regions, 2j1bSR and 3j1bSR, are compared to those from the previous tZq study. The obtained results show a great agreement with those from the previous study, confirming the robustness of the framework.

The first part of the tHq study is mainly about the origin of the light leptons. In order to assign the light leptons to their correct origin, a categorization by their sign (opposite sign or same sign) as well as their flavor (different flavor or same flavor) resulting in four different combinations: OSSF, OSDF, SSSF, SSDF, has been carried out. For each category, the theoretical expected number of events is calculated in all tHq ML channels $H \to \tau\tau/WW/ZZ$. The calculation indicates that there are approximately 1.37 more OS lepton pairs than SS lepton pairs are in the signal region, but since the major background $t\bar{t}$ possesses a large number of OS lepton pairs, the signal sensitivity ($S/\sqrt{B}$) in the OS (OSSF and OSDF) regions is lower than that in the SS (SSSF and SSDF) regions by a factor of 4. Therefore, it would be a better idea to seek for the signal in the SS regions, but an efficient suppression method for the Z+jets and $t\bar{t}$ still needs to be discussed. Then the theoretical expectation has been compared to the observed number of events. However, we clearly confirmed that there is sever low statistic issue for the tHq samples from the discrepancies between them.

The second part of the tHq study is the final state reconstruction. We first reconstructed the top quark

with the analytical method and then the Higgs boson with commonly used $\tau\tau$-system reconstruction tools, since the total transverse momentum is known and fixed. The exact solutions are obtained by applying 3 constrains to the neutrino four-momentum vector based on the assumptions that the top quark and the W boson are on-shell and the z-component of the neutrino momentum (p_ν^z) is dependent on that of the light lepton from the same origin ($p_\nu^z = \alpha \cdot p_l^z$ where α is an arbitrary constant). However, there are multiple-solution and non-real solution issues, which are contradictory to the natural observation where there is only one real solution. To resolve these issues, some remedies which convert the imaginary solutions to real ones have been proposed, and an investigation on individual solutions in the case of more-than-one solution have been carried out. Since the number of solutions depends on the initial assumption on α, optimization as well as validation using a truth-matching method have been performed. The α value is optimized to the value (0.5) which gives the highest ratio of the number of real solutions over that of imaginary solutions. The truth study indicates that there is a mismatching between the reco- and truth-level p_ν^z, throwing an unsolved task for finding a better assumption on the p_ν^z. For this, NN might be helpful apart from the analytical approach. Nevertheless, the reconstructed invariant mass of the di-tau system shows that this method worked out, and presents that the Missing Mass Calculator (MMC) is a better tool than relatively older methods, collinear approximation and partial mass reconstruction. Furthermore, some new techniques are being developed currently for the improvement of the MMC performance, and one can try out these techniques to have a better reconstructed Higgs system. In one word, this analysis shows that a full reconstruction of the lep-had final state has been quite successful by using the analytic method as well as the advanced technique MMC even the reconstruction is not perfect. We expect a much better outcome if the assumption issues can be solved, and then this reconstruction algorithm can be used by other $tHq \rightarrow t\tau\tau q$ channels in the future.

Bibliography

[1] J. J. T. M. F.R.S., *XL. Cathode Rays*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **44** (1897) 293,
eprint: https://doi.org/10.1080/14786449708621070,
URL: https://doi.org/10.1080/14786449708621070 (cit. on p. 1).

[2] G. Aad et al., *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** (2012) 1, arXiv: 1207.7214 [hep-ex] (cit. on pp. 1, 17).

[3] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 9781107034266 (cit. on pp. 3–6, 10).

[4] Wikibooks, *A-level Physics (Advancing Physics)/The Standard Model — Wikibooks, The Free Textbook Project*, [Online; accessed 28-April-2020], 2017,
URL: https://en.wikibooks.org/w/index.php?title=A-level_Physics_(Advancing_Physics)/The_Standard_Model&oldid=3249657 (cit. on p. 4).

[5] The Nobel Prize in Physics 1959, *NobelPrize.org. Nobel Media AB 2020*, Sun. 9 Feb 2020,
URL: https://www.nobelprize.org/prizes/physics/1959/summary/ (cit. on p. 4).

[6] T. L. Collaboration et al., *A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model*, 2001, arXiv: hep-ex/0112021 [hep-ex] (cit. on p. 5).

[7] A. Lenz, *Constraints on a Fourth Generation of Fermions from Higgs Boson Searches*, Advances in High Energy Physics **2013** (2013) 13 (cit. on p. 5).

[8] J. Feltesse, *Introduction to Parton Distribution Functions*, Scholarpedia **5** (2010) 10160, revision #186761 (cit. on p. 10).

[9] H. rauenfelder and E. M. Henley, *Subatomic Physics, 2nd ed.*, Scholarpedia **5** (1991) 10160, revision #186761 (cit. on p. 11).

[10] F. Abe et al.,
Observation of Top Quark Production inp¯pCollisions with the Collider Detector at Fermilab, Physical Review Letters **74** (1995) 2626, ISSN: 1079-7114,
URL: http://dx.doi.org/10.1103/PhysRevLett.74.2626 (cit. on p. 12).

[11] S. Abachi et al., *Observation of the Top Quark*, Physical Review Letters **74** (1995) 2632,
ISSN: 1079-7114, URL: http://dx.doi.org/10.1103/PhysRevLett.74.2632 (cit. on p. 12).

[12] M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. D **98** (3 2018) 030001,
URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001 (cit. on p. 13).

[13] *Measurement of the $t\bar{t}$ production cross-section and lepton differential distributions in $e\mu$ dilepton events from pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, tech. rep. ATLAS-CONF-2019-041, CERN, 2019, URL: https://cds.cern.ch/record/2686255 (cit. on p. 13).

[14] Atlas Collaboration et al., *Fiducial, total and differential cross-section measurements of t-channel single top-quark production in pp collisions at 8 TeV using data collected by the ATLAS detector*, English, European Physical Journal C **77** (2017), ISSN: 1434-6044 (cit. on p. 14).

[15] M. Aaboud et al., *Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **04** (2017) 086, arXiv: 1609.03920 [hep-ex] (cit. on p. 14).

[16] G. Aad et al., *Measurement of the production cross-section of a single top quark in association with a W boson at 8 TeV with the ATLAS experiment*, Journal of High Energy Physics **2016** (2016) (cit. on p. 14).

[17] M. Aaboud et al., *Measurement of the cross-section for producing a W boson in association with a single top quark in pp collisions at $\sqrt{s}$ = 13 TeV with ATLAS*, JHEP **01** (2018) 063, arXiv: 1612.07231 [hep-ex] (cit. on p. 14).

[18] T. A. collaboration, *Evidence for single top-quark production in the s-channel in proton-proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector using the Matrix Element Method*, (2015) (cit. on p. 14).

[19] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. **D81** (2010) 054028, arXiv: 1001.5034 [hep-ph] (cit. on p. 14).

[20] B. M. M. Allbrooke et al., *Measurement of the tt?W and tt?Z production cross sections in multilepton final states using 3.2 fb?? of pp collisions at s $\sqrt{}$ = 13 TeV with the ATLAS detector*, European Physical Journal C: Particles and Fields **77** (2017), URL: http://sro.sussex.ac.uk/id/eprint/66370/ (cit. on p. 15).

[21] M. Aaboud et al., *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. **B784** (2018) 173, arXiv: 1806.00425 [hep-ex] (cit. on pp. 15, 17).

[22] A. M. Sirunyan et al., *Observation of $t\bar{t}H$ production*, Phys. Rev. Lett. **120** (2018) 231801, arXiv: 1804.02610 [hep-ex] (cit. on pp. 15, 17).

[23] A. M. Sirunyan et al., *Search for associated production of a Higgs boson and a single top quark in proton-proton collisions at $\sqrt{s}$ = TeV*, Phys. Rev. **D99** (2019) 092005, arXiv: 1811.09696 [hep-ex] (cit. on pp. 15, 53).

[24] M. Alhroob et al., *Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, tech. rep. ATL-COM-PHYS-2019-933, CERN, 2019, URL: https://cds.cern.ch/record/2683901 (cit. on pp. 16, 17, 46).

[25] A. M. Sirunyan et al., *Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s}$ =13 TeV*, Phys. Rev. Lett. **122** (2019) 132003, arXiv: 1812.05900 [hep-ex] (cit. on p. 17).

[26] H. P. Künzle, *Galilei and Lorentz structures on space-time : comparison of the corresponding geometry and physics*, en, Annales de l'I.H.P. Physique théorique **17** (1972) 337, URL: http://www.numdam.org/item/AIHPA_1972__17_4_337_0 (cit. on p. 17).

[27] A. M. Sirunyan et al., *Observation of Higgs boson decay to bottom quarks*, Phys. Rev. Lett. **121** (2018) 121801, arXiv: 1808.08242 [hep-ex] (cit. on p. 17).

[28] V. Khachatryan et al., *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV*, Eur. Phys. J. **C75** (2015) 212, arXiv: 1412.8662 [hep-ex] (cit. on p. 17).

[29] J. Ellis and T. You, *Updated Global Analysis of Higgs Couplings*, JHEP **06** (2013) 103, arXiv: 1303.3879 [hep-ph] (cit. on p. 17).

[30] *Standard Model Total Production Cross Section Measurements*, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/ (cit. on p. 18).

[31] *Large Hardon Collider*, URL: http://www.lhc-facts.ch/ (cit. on p. 19).

[32] Wikipedia contributors, *Large Hadron Collider — Wikipedia, The Free Encyclopedia*, [Online; accessed 10-February-2020], 2020, URL: https://en.wikipedia.org/w/index.php?title=Large_Hadron_Collider&oldid=939876548 (cit. on p. 20).

[33] G. Ripellino, *The alignment of the ATLAS Inner Detector in Run 2*, tech. rep. ATL-INDET-PROC-2016-003, CERN, 2016, URL: https://cds.cern.ch/record/2213441 (cit. on pp. 20, 22).

[34] *How does CMS measure luminosity?*, URL: https://cms.cern/news/how-does-cms-measure-luminosity (cit. on p. 20).

[35] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001, URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08001 (cit. on pp. 20, 21).

[36] *Total Integrated Luminosity and Data Quality in 2018*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (cit. on p. 20).

[37] G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 21–25, 30).

[38] S. Chatrchyan et al., *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004 (cit. on p. 21).

[39] K. Aamodt et al., *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002 (cit. on p. 21).

[40] A. A. Alves Jr. et al., *The LHCb Detector at the LHC*, JINST **3** (2008) S08005 (cit. on p. 21).

[41] A. L. Rosa, *The ATLAS Insertable B-Layer: from construction to operation*, Journal of Instrumentation **11** (2016) C12036, URL: https://doi.org/10.1088%2F1748-0221%2F11%2F12%2Fc12036 (cit. on p. 23).

[42] J. R. Pater, *The ATLAS SemiConductor Tracker operation and performance*, Journal of Instrumentation **7** (2012) C04001, URL: https://doi.org/10.1088%2F1748-0221%2F7%2F04%2Fc04001 (cit. on p. 23).

[43] A. Vogel, *ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates*, tech. rep. ATL-INDET-PROC-2013-005, CERN, 2013, URL: https://cds.cern.ch/record/1537991 (cit. on p. 23).

[44] H. Kolanoski and N. Wermes, *Teilchendetektoren*, Springer, 2016, ISBN: 9783662453490, 9783662453506 (cit. on pp. 23–25).

[45] *Reconstruction of physics objects in ATLAS*, URL: http://dlnp.jinr.ru/en/news/dlnp-news/122-cern-director-general-fabiola-gianotti-holds-a-special-press-conference-today (cit. on p. 26).

[46] M. Aaboud et al., *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*, Eur. Phys. J. **C77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on p. 26).

[47] *Number of Interactions per Crossing*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Pileup_Interactions_and_Data_Tak (cit. on p. 27).

[48] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, tech. rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, 2008, URL: https://cds.cern.ch/record/1099735 (cit. on p. 28).

[49] *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*, tech. rep. ATL-PHYS-PUB-2017-022, CERN, 2017, URL: https://cds.cern.ch/record/2298955 (cit. on p. 28).

[50] M. Aaboud et al., *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at* $\sqrt{s}$ *= 13 TeV*, Eur. Phys. J. **C79** (2019) 639, arXiv: 1902.04655 [physics.ins-det] (cit. on p. 28).

[51] *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at* $\sqrt{s}$ *= 13 TeV*, tech. rep. ATLAS-CONF-2017-029, CERN, 2017, URL: https://cds.cern.ch/record/2261772 (cit. on p. 29).

[52] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 29).

[53] *Tagging and suppression of pileup jets with the ATLAS detector*, tech. rep. ATLAS-CONF-2014-018, CERN, 2014, URL: https://cds.cern.ch/record/1700870 (cit. on p. 30).

[54] M. Aaboud et al., *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. **C77** (2017) 580, [Erratum: Eur. Phys. J.C77,no.10,712(2017)], arXiv: 1705.02211 [hep-ex] (cit. on p. 30).

[55] M. Hodgkinson, D. Tovey and R. Duxfield, *Energy Flow Reconstruction with the eflowRec Combined Reconstruction Software in Athena 15.6.9.8*, tech. rep. ATL-PHYS-INT-2011-031, CERN, 2011, URL: https://cds.cern.ch/record/1342125 (cit. on p. 31).

[56] *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, tech. rep. ATL-PHYS-PUB-2016-012, CERN, 2016, URL: https://cds.cern.ch/record/2160731 (cit. on p. 31).

[57] *Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at at* $\sqrt{s}$*= 13 TeV*, tech. rep. ATL-PHYS-PUB-2015-027, CERN, 2015, URL: https://cds.cern.ch/record/2037904 (cit. on p. 31).

[58] G. Aad et al., *Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, (2020), arXiv: 2002.07546 [hep-ex] (cit. on pp. 33–35, 39, 44–46).

[59] M. Alhroob et al., *Observation of the associated production of a top quark and a Z boson at 13 TeV with ATLAS*, tech. rep. ATL-COM-PHYS-2018-140, CERN, 2018, URL: https://cds.cern.ch/record/2304824 (cit. on pp. 33, 39, 44–47, 51, 111).

[60] R. D. Field, *The Underlying Event in Hard Scattering Processes*, eConf **C010630** (2001) P501, ed. by N. Graf, arXiv: hep-ph/0201192 (cit. on p. 34).

[61] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph] (cit. on p. 34).

[62] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, ISSN: 0168-9002, URL: http://www.sciencedirect.com/science/article/pii/S0168900203013688 (cit. on p. 34).

[63] S. Höche, "Introduction to parton-shower event generators", *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, 2015 235, arXiv: 1411.4085 [hep-ph] (cit. on p. 34).

[64] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on p. 34).

[65] F. Demartin et al., *Higgs production in association with a single top quark at the LHC*, 2015, arXiv: 1504.00611 [hep-ph] (cit. on p. 34).

[66] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on p. 34).

[67] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on p. 34).

[68] *TopLoop Documentation*, URL: https://atlas-toploop.web.cern.ch/atlas-toploop/docs/index.html (cit. on p. 37).

[69] *Event weights*, URL: http://ipnp.cz/scheirich/?page_id=292 (cit. on p. 37).

[70] I. A. Cioară, *Associated Production of a Top Quark and a Z Boson in pp Collisions at $\sqrt{s}$ = 13 TeV Using the ATLAS Detector*, PhD Thesis: University of Bonn, 2018, URL: http://hss.ulb.uni-bonn.de/diss_online (cit. on p. 40).

[71] O. Bessidskaia Bylund, "Modelling Wt and tWZ production at NLO for ATLAS analyses", *Proceedings, 9th International Workshop on Top Quark Physics (TOP 2016): Olomouc, Czech Republic, September 19-23, 2016*, 2016, arXiv: 1612.00440 [hep-ph] (cit. on p. 42).

[72] A. Calandri, *Flavour tagging algorithms and performance at the ATLAS experiment*, PoS **LHCP2016** (2016) 180, ed. by J. Bijnens, A. Hoecker and J. Olsen (cit. on p. 44).

[73] V. Khachatryan et al., *Search for the associated production of a Higgs boson with a single top quark in proton-proton collisions at* $\sqrt{s} = 8$ *TeV*, JHEP **06** (2016) 177, arXiv: 1509.08159 [hep-ex] (cit. on p. 53).

[74] *Search for production of a Higgs boson and a single top quark in multilepton final states in proton collisions at* $\sqrt{s} = 13$ TeV, tech. rep. CMS-PAS-HIG-17-005, CERN, 2017, URL: https://cds.cern.ch/record/2264553 (cit. on p. 53).

[75] F. Demartin et al., *Higgs production in association with a single top quark at the LHC*, Eur. Phys. J. **C75** (2015) 267, arXiv: 1504.00611 [hep-ph] (cit. on p. 59).

[76] Wikipedia contributors, *Conic section — Wikipedia, The Free Encyclopedia*, [Online; accessed 8-March-2020], 2020, URL: https://en.wikipedia.org/w/index.php?title=Conic_section&oldid=944475360 (cit. on p. 72).

[77] M. Alberich-Carramiñana, B. Elizalde and F. Thomas, *New algebraic conditions for the identification of the relative position of two coplanar ellipses*, Computer Aided Geometric Design **54** (2017) 35, ISSN: 0167-8396, URL: http://www.sciencedirect.com/science/article/pii/S0167839617300791 (cit. on p. 74).

[78] S. Richter, *Invariant mass reconstruction in a search for light charged Higgs bosons in pp collisions at* $\sqrt{s} = 7$ *TeV*, MA thesis: Helsinki U., 2013 (cit. on pp. 83, 93).

[79] B. Idaszek, *RECONSTRUCTION OF THE NEUTRINO MOMENTUM IN TOPPAIR ASSOCIATEDHIGGS BOSON PRODUCTION USINGDEEPLEARNING*, Helsinki U., 2019, URL: https://www.institut3a.physik.rwth-aachen.de/cms/institut3a/Das-Institut/Profil/Prof-Erdmann/~onvv/Abschlussarbeiten/lidx/1/ (cit. on p. 88).

[80] A. Elagin et al., *A new mass reconstruction technique for resonances decaying to di-tau*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **654** (2011) 481, ISSN: 0168-9002, URL: http://www.sciencedirect.com/science/article/pii/S0168900211014112 (cit. on pp. 93–97).

[81] A. M. Sirunyan et al., *Observation of the Higgs boson decay to a pair of* τ *leptons with the CMS detector*, Phys. Lett. **B779** (2018) 283, arXiv: 1708.00373 [hep-ex] (cit. on p. 94).

[82] P. Konar and A. K. Swain, *Reconstructing semi-invisible events in resonant tau pair production from Higgs*, Physics Letters B **757** (2016) 211, ISSN: 0370-2693, URL: http://www.sciencedirect.com/science/article/pii/S0370269316300417 (cit. on pp. 94, 95).

APPENDIX A

List of Monte Carlo samples

An overview of the MC simulated samples that were used for this analysis is given. The systematic uncertainties were not covered, thus only nominal samples are contained in this list.

Table A.1: An overview of MC samples for signal and background processes (full simulation). The samples correspond to version 29 ntuples (with τ). The table is taken from [59].

Process	Sample ID	Generator	σ [pb]	k-factor	N-generated
tHq	346229	MadGraphPy8Ev–A14–tHjb–125-4fl-bb-NLO	0.0400	1.00	mc16a: 0.30 M mc16d: 0.38 M mc16e: 0.50 M
	346230	MadGraphPy8Ev–A14–tHjb–125-4fl-lep-NLO	0.0200	1.00	mc16a: 0.30 M mc16d: 0.38 M mc16e: 0.50 M
tZq	412063	MadGraphPy8Ev–A14–tllq–NLO	0.0300	1.00	mc16a: 4.99 M mc16d: 6.23 M mc16e: 8.26 M
$t\bar{t}$	410472	PowPy8Ev–A14–ttbar–hdamp258p75–dil	77.0	1.14	mc16a: 79.83 M mc16d: 44.88 M * mc16e: 99.25 M
tW	410648	PowPy8Ev–A14–Wt–t–dil	4.00	0.94	mc16a: 1.00 M mc16d: 1.25 M mc16e: 1.66 M
	410649	PowPy8Ev–A14–Wt–tbar–dil	3.99	0.94	mc16a: 1.00 M mc16d: 1.25 M mc16e: 1.65 M
ttH	346343	PowPy8Ev–A14–ttH125–allhad	0.0534	1.00	mc16a: 4.98 M mc16d: 6.49 M mc16e: 8.25 M
	346344	PowPy8Ev–A14–ttH125–sl	0.223	1.00	mc16a: 4.99 M mc16d: 6.50 M mc16e: 8.26 M
	346345	PowPy8Ev–A14–ttH125–dilep	0.231	1.00	mc16a: 4.99 M mc16d: 6.49 M

Process	Sample ID	Generator	σ [pb]	k-factor	N-generated
					mc16e: 8.28 M
ttV	410155	aMcAtNloPy8Ev–A14–ttW	0.548	1.10	mc16a: 7.50 M mc16d: 7.50 M mc16e: 12.04 M
	410156	aMcAtNloPy8Ev–A14–ttZnunu	0.155	1.11	mc16a: 1.50 M mc16d: 1.50 M mc16e: 2.00 M
	410157	aMcAtNloPy8Ev–A14–ttZqq	0.528	1.11	mc16a: 3.00 M mc16d: 3.00 M mc16e: 3.59 M
	410218	aMcAtNloPy8Ev–A14–ttee	0.0369	1.12	mc16a: 1.41 M mc16d: 1.34 M mc16e: 2.17 M
	410219	aMcAtNloPy8Ev–A14–ttmumu	0.0369	1.12	mc16a: 1.41 M mc16d: 1.34 M mc16e: 2.17 M
	410220	aMcAtNloPy8Ev–A14–tttautau	0.0365	1.12	mc16a: 0.94 M mc16d: 0.90 M mc16e: 0.96 M
tWZ	410408	aMcAtNloPy8Ev–A14–tWZ–Ztoll–DR1	0.0200	1.00	mc16a: 0.10 M mc16d: 0.12 M mc16e: 0.16 M
Diboson	363356	Sherpa221–ZqqZll	15.6	0.14	mc16a: 5.40 M mc16d: 5.40 M mc16e: 8.95 M
	363358	Sherpa221–WqqZll	3.44	1.00	mc16a: 5.40 M mc16d: 26.91 M mc16e: 8.96 M
	364250	Sherpa222–llll	1.25	1.00	mc16a: 17.84 M mc16d: 36.00 M mc16e: 25.68 M
	364253	Sherpa222–lllv	4.58	1.00	mc16a: 15.54 M mc16d: 32.11 M mc16e: 26.79 M
	364254	Sherpa222–llvv	12.5	1.00	mc16a: 15.00 M mc16d: 29.98 M mc16e: 24.89 M
Z+jets	364114	Sherpa221–Zee–maxHtPtV0_70–L	1 630	0.98	mc16a: 8.00 M mc16d: 10.00 M mc16e: 13.27 M
	364115	Sherpa221–Zee–maxHtPtV0_70–C	224	0.98	mc16a: 5.00 M mc16d: 6.24 M mc16e: 8.31 M
	364116	Sherpa221–Zee–maxHtPtV0_70–B	126	0.98	mc16a: 8.00 M mc16d: 9.99 M mc16e: 13.28 M
	364117	Sherpa221–Zee–maxHtPtV70_140–L	76.3	0.98	mc16a: 5.96 M mc16d: 7.38 M

Process	Sample ID	Generator	σ [pb]	k-factor	N-generated
	364118	Sherpa221–Zee–maxHtPtV70_140–C	20.3	0.98	mc16e: 9.95 M mc16a: 2.00 M mc16d: 2.50 M
	364119	Sherpa221–Zee–maxHtPtV70_140–B	12.6	0.98	mc16e: 3.33 M mc16a: 5.97 M mc16d: 7.49 M
	364120	Sherpa221–Zee–maxHtPtV140_280–L	25.0	0.98	mc16e: 9.91 M mc16a: 5.00 M mc16d: 6.25 M
	364121	Sherpa221–Zee–maxHtPtV140_280–C	9.37	0.98	mc16e: 8.36 M mc16a: 3.00 M mc16d: 3.75 M
	364122	Sherpa221–Zee–maxHtPtV140_280–B	6.08	0.98	mc16e: 4.99 M mc16a: 12.44 M mc16d: 15.66 M
	364123	Sherpa221–Zee–maxHtPtV280_500–L	4.87	0.98	mc16e: 20.74 M mc16a: 2.00 M mc16d: 2.50 M
	364124	Sherpa221–Zee–maxHtPtV280_500–C	2.28	0.98	mc16e: 2.78 M mc16a: 1.00 M mc16d: 1.25 M
	364125	Sherpa221–Zee–maxHtPtV280_500–B	1.49	0.98	mc16e: 1.74 M mc16a: 2.00 M mc16d: 2.50 M
	364126	Sherpa221–Zee–maxHtPtV500_1000	1.81	0.98	mc16e: 3.33 M mc16a: 3.00 M mc16d: 3.71 M
	364127	Sherpa221–Zee–maxHtPtV1000_Ecms	0.150	0.98	mc16e: 4.98 M mc16a: 1.00 M mc16d: 1.25 M
	364100	Sherpa221–Zmumu–maxHtPtV0_70–L	1 630	0.98	mc16e: 1.67 M mc16a: 7.97 M mc16d: 9.91 M
	364101	Sherpa221–Zmumu–maxHtPtV0_70–C	224	0.98	mc16e: 13.26 M mc16a: 4.98 M mc16d: 6.20 M
	364102	Sherpa221–Zmumu–maxHtPtV0_70–B	127	0.98	mc16e: 8.28 M mc16a: 7.98 M mc16d: 9.26 M
	364103	Sherpa221–Zmumu–maxHtPtV70_140–L	75.0	0.98	mc16e: 13.24 M mc16a: 5.98 M mc16d: 7.48 M
	364104	Sherpa221–Zmumu–maxHtPtV70_140–C	20.4	0.98	mc16e: 9.94 M mc16a: 2.00 M mc16d: 2.49 M
	364105	Sherpa221–Zmumu–maxHtPtV70_140–B	12.4	0.98	mc16e: 3.31 M mc16a: 5.98 M mc16d: 7.47 M

Process	Sample ID	Generator	σ [pb]	k-factor	N-generated
	364106	Sherpa221–Zmumu–maxHtPtV140_280–L	24.3	0.98	mc16e: 9.94 M mc16a: 5.00 M mc16d: 6.24 M
	364107	Sherpa221–Zmumu–maxHtPtV140_280–C	9.28	0.98	mc16e: 8.29 M mc16a: 3.00 M mc16d: 3.75 M
	364108	Sherpa221–Zmumu–maxHtPtV140_280–B	6.01	0.98	mc16e: 4.99 M mc16a: 12.46 M mc16d: 15.63 M
	364109	Sherpa221–Zmumu–maxHtPtV280_500–L	4.77	0.98	mc16e: 20.74 M mc16a: 2.00 M mc16d: 2.46 M
	364110	Sherpa221–Zmumu–maxHtPtV280_500–C	2.27	0.98	mc16e: 3.32 M mc16a: 1.00 M mc16d: 1.25 M
	364111	Sherpa221–Zmumu–maxHtPtV280_500–B	1.49	0.98	mc16e: 1.67 M mc16a: 2.00 M mc16d: 2.50 M
	364112	Sherpa221–Zmumu–maxHtPtV500_1000	1.79	0.98	mc16e: 3.33 M mc16a: 3.00 M mc16d: 3.75 M
	364113	Sherpa221–Zmumu–maxHtPtV1000_Ecms	0.150	0.98	mc16e: 5.09 M mc16a: 1.00 M mc16d: 1.25 M
	364128	Sherpa221–Ztautau–maxHtPtV0_70–L	1 630	0.98	mc16e: 1.67 M mc16a: 7.99 M mc16d: 10.00 M
	364129	Sherpa221–Ztautau–maxHtPtV0_70–C	224	0.98	mc16e: 13.27 M mc16a: 4.98 M mc16d: 6.14 M
	364130	Sherpa221–Ztautau–maxHtPtV0_70–B	128	0.98	mc16e: 8.27 M mc16a: 8.00 M mc16d: 9.99 M
	364131	Sherpa221–Ztautau–maxHtPtV70_140–L	76.0	0.98	mc16e: 13.28 M mc16a: 6.00 M mc16d: 7.50 M
	364132	Sherpa221–Ztautau–maxHtPtV70_140–C	20.2	0.98	mc16e: 9.97 M mc16a: 2.00 M mc16d: 2.50 M
	364133	Sherpa221–Ztautau–maxHtPtV70_140–B	12.3	0.98	mc16e: 3.33 M mc16a: 5.97 M mc16d: 7.50 M
	364134	Sherpa221–Ztautau–maxHtPtV140_280–L	24.8	0.98	mc16e: 9.96 M mc16a: 4.94 M mc16d: 6.23 M
	364135	Sherpa221–Ztautau–maxHtPtV140_280–C	9.33	0.98	mc16e: 8.30 M mc16a: 3.00 M mc16d: 3.75 M

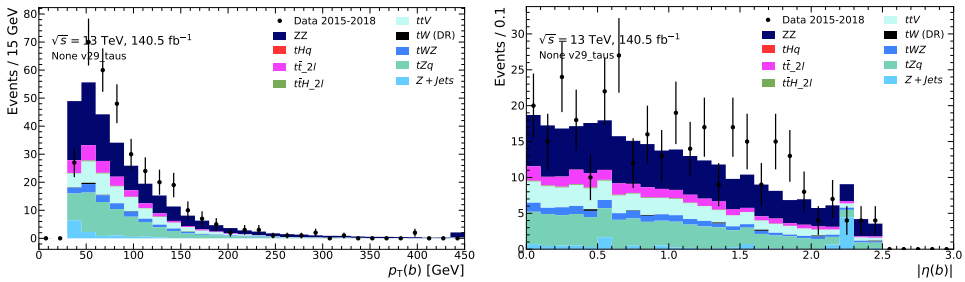
Process	Sample ID	Generator	σ [pb]	k-factor	N-generated
	364136	Sherpa221–Ztautau–maxHtPtV140_280–B	5.48	0.98	mc16e: 4.99 M mc16a: 4.99 M mc16d: 6.21 M
	364137	Sherpa221–Ztautau–maxHtPtV280_500–L	4.79	0.98	mc16e: 8.29 M mc16a: 2.00 M mc16d: 2.50 M
	364138	Sherpa221–Ztautau–maxHtPtV280_500–C	2.28	0.98	mc16e: 3.32 M mc16a: 1.00 M mc16d: 1.23 M
	364139	Sherpa221–Ztautau–maxHtPtV280_500–B	1.50	0.98	mc16e: 1.67 M mc16a: 1.98 M mc16d: 2.50 M
	364140	Sherpa221–Ztautau–maxHtPtV500_1000	1.81	0.98	mc16e: 3.32 M mc16a: 3.00 M mc16d: 3.75 M
	364141	Sherpa221–Ztautau–maxHtPtV1000_Ecms	0.150	0.98	mc16e: 4.96 M mc16a: 1.00 M mc16d: 1.25 M mc16e: 1.67 M

APPENDIX B

Additional SR plots

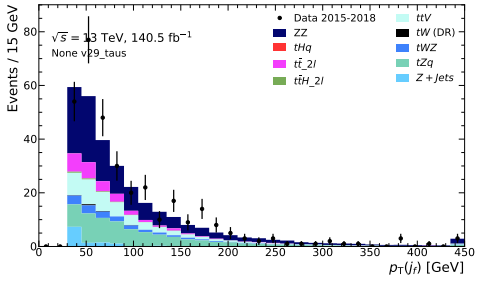
In this appendix, additional plots for the 2j1b and 3j1b SRs (the tZq ML channel selection) that are produced with `tHqLoop` are shown.

B.1 Basic kinematic variables in the tZq 2j1b SR

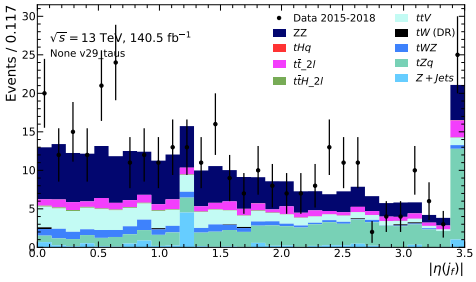


(a)

(b)

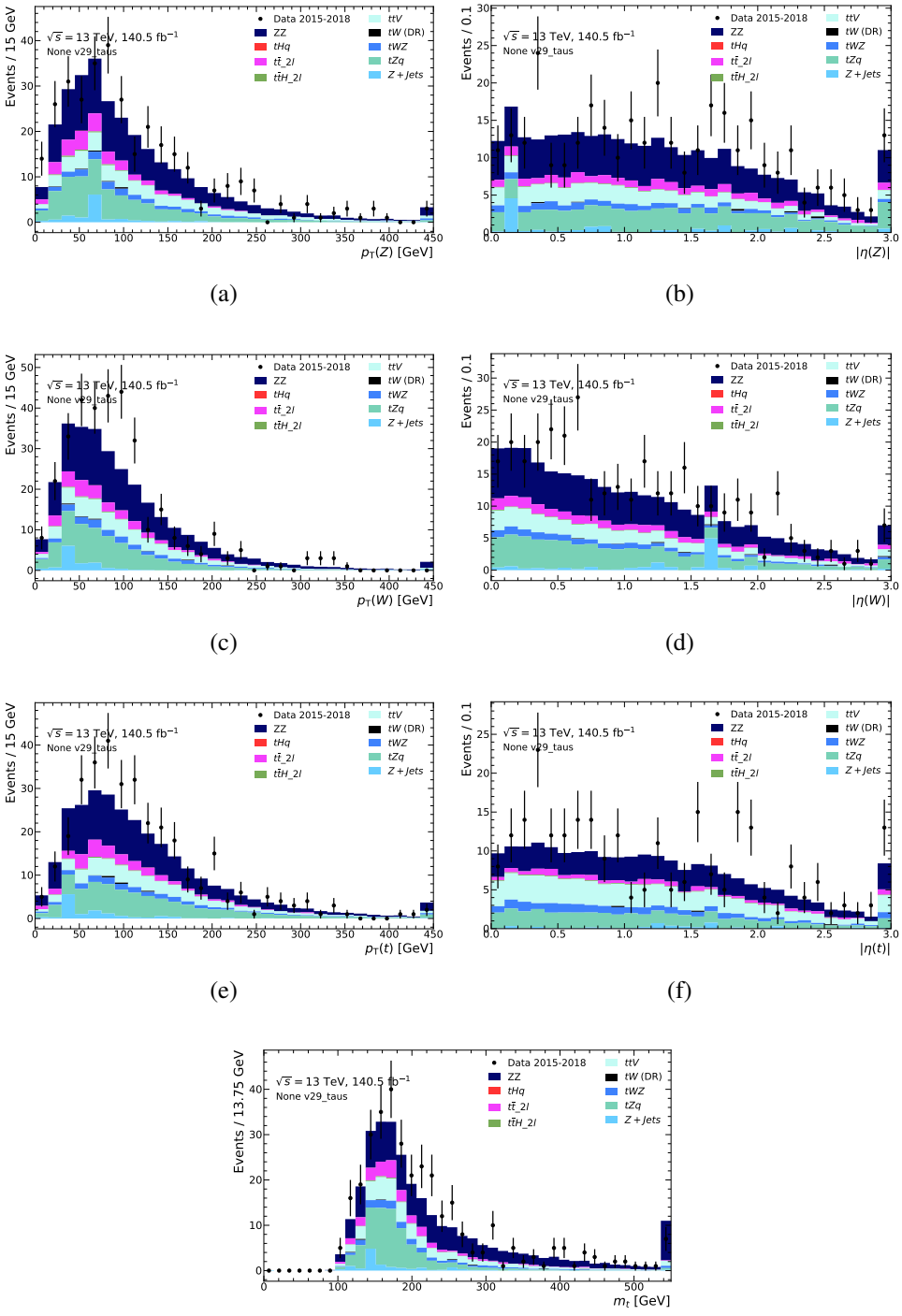


(c)



(d)

Figure B.1: Comparison of data and MC predictions for reconstructed jet-related quantities for events in the 2j1b SR. The uncertainty band includes only statistical uncertainties for the signal and backgrounds.



(a) (b)

(c) (d)

(e) (f)

(g)

Figure B.2: Comparison of data and MC predictions for reconstructed event-related quantities for events in the 2j1b SR. The uncertainty band includes only statistical uncertainties for the signal and backgrounds.

B.2 Basic kinematic variables in the tZq 3j1b SR

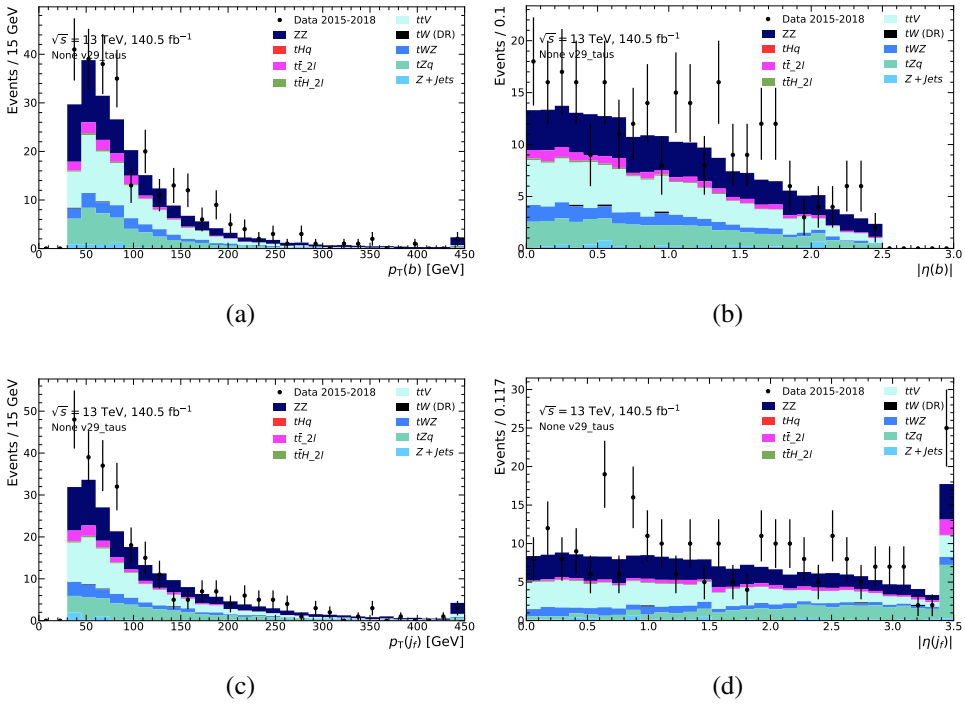


Figure B.3: Comparison of data and MC predictions for reconstructed jet-related quantities for events in the 3j1b SR. The uncertainty band includes only statistical uncertainties for the signal and backgrounds.

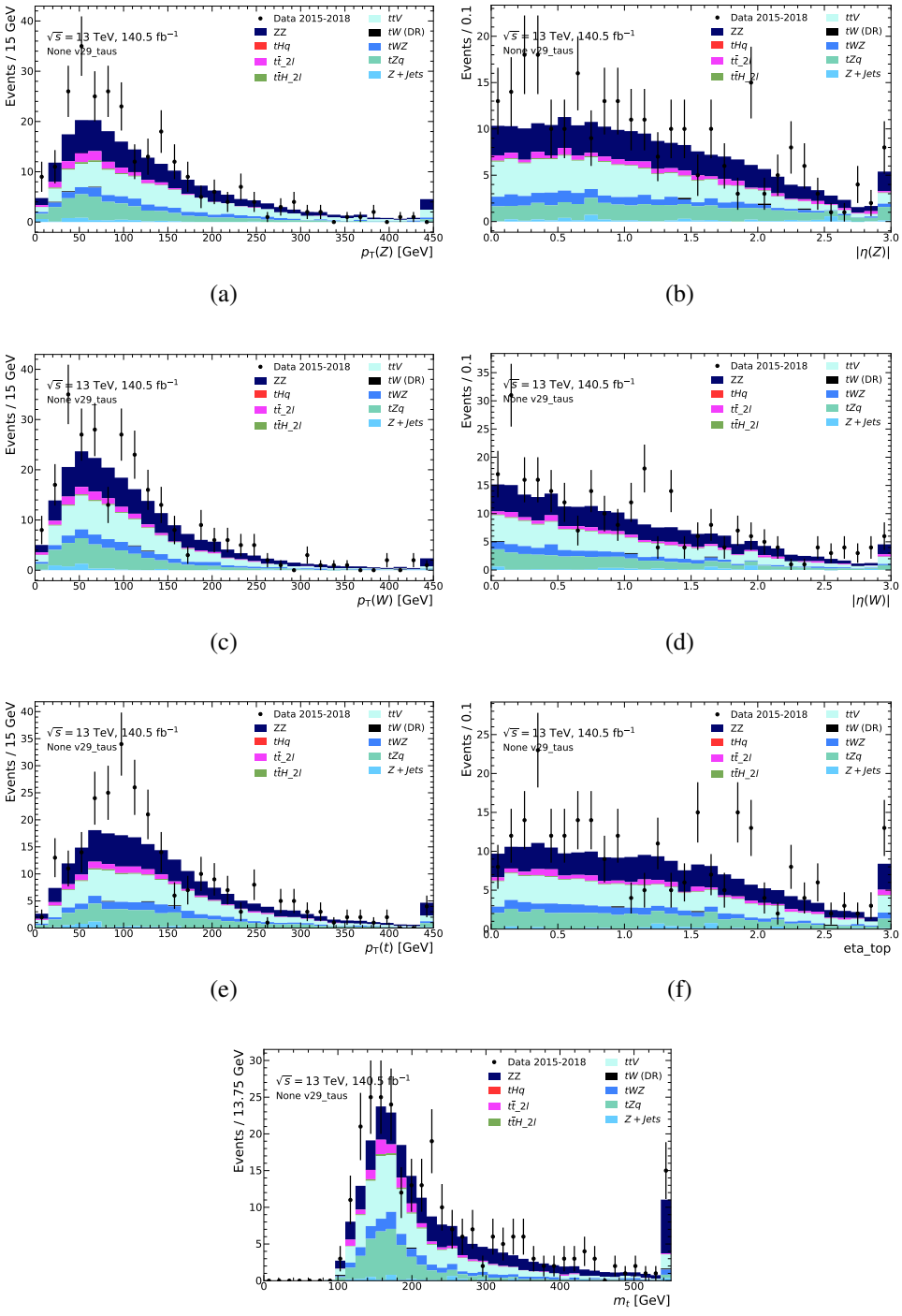


Figure B.4: Comparison of data and MC predictions for reconstructed event-related quantities for events in the 3j1b SR. The uncertainty band includes only statistical uncertainties for the signal and backgrounds.

APPENDIX C

Calculation of the number of events in $tHq \rightarrow t\tau_l\tau_h q$ channel

lepton pair		OS-1	OSSF		OSDF	
decay channel		1 - 4	1	2	3	4
H	W^+	l^+/τ_{l^+}	e^+/τ_{e^+}	μ^+/τ_{μ^+}	e^+/τ_{e^+}	μ^+/τ_{μ^+}
	W^-	l^-/τ_{l^-}	e^-/τ_{e^-}	μ^-/τ_{μ^-}	μ^-/τ_{μ^-}	e^-/τ_{e^-}
t	W	τ_h^+	τ_h^+	τ_h^+	τ_h^+	τ_h^+
BR(%)		0.06	0.02	0.02	0.02	0.02
σ_{sig}[fb]		0.05	0.01	0.01	0.01	0.01
N_{event}		6.55	1.65	1.62	1.64	1.64

lepton pair		OS-2	OSSF		OSDF	
decay channel		1 - 4	1	2	3	4
H	W^+	l^+/τ_{l^+}	e^+/τ_{e^+}	μ^+/τ_{μ^+}	e^+/τ_{e^+}	μ^+/τ_{μ^+}
	W^-	l^-/τ_{l^-}	e^-/τ_{e^-}	μ^-/τ_{μ^-}	μ^-/τ_{μ^-}	e^-/τ_{e^-}
$\bar{t}$	W	τ_h^-	τ_h^-	τ_h^-	τ_h^-	τ_h^-
BR(%)		0.04	0.01	0.01	0.01	0.01
σ_{sig}[fb]		0.03	0.01	0.01	0.01	0.01
N_{event}		3.81	0.96	0.94	0.95	0.95

lepton pair		OS-3	OSSF		OSDF	
decay channel		1 - 4	1	2	3	4
H	W^+	τ_h^+	τ_h^+	τ_h^+	τ_h^+	τ_h^+
	W^-	l^-/τ_{l^-}	e^-/τ_{e^-}	μ^-/τ_{μ^-}	e^-/τ_{e^-}	μ^-/τ_{μ^-}
t	W	l^+/τ_{l^+}	e^+/τ_{e^+}	μ^+/τ_{μ^+}	μ^+/τ_{μ^+}	e^+/τ_{e^+}
BR(%)		0.06	0.02	0.02	0.02	0.02
σ_{sig}[fb]		0.05	0.01	0.01	0.01	0.01
N_{event}		6.55	1.65	1.64	1.64	1.62

lepton pair		OS-4	OSSF		OSDF	
decay channel		1 - 4	1	2	3	4
H	W^+	l^+/τ_{l^+}	e^+/τ_{e^+}	μ^+/τ_{μ^+}	e^+/τ_{e^+}	μ^+/τ_{μ^+}
	W^-	τ_h^-	τ_h^-	τ_h^-	τ_h^-	τ_h^-
$\bar{t}$	W	l^-/τ_{l^-}	e^-/τ_{e^-}	μ^-/τ_{μ^-}	μ^-/τ_{μ^-}	e^-/τ_{e^-}
BR(%)		0.04	0.01	0.01	0.09	0.09
σ_{sig}[fb]		0.03	0.01	0.01	0.01	0.01
N_{event}		3.81	0.96	0.95	0.95	0.94

lepton pair		SS-5	SSSF		SSDF	
decay channel		1 - 4	1	2	3	4
H	W^+	τ_h^+	τ_h^+	τ_h^+	τ_h^+	τ_h^+
	W^-	l^-/τ_{l^-}	e^-/τ_{e^-}	μ^-/τ_{μ^-}	e^-/τ_{e^-}	μ^-/τ_{μ^-}
$\bar{t}$	W	l^-/τ_{l^-}	e^-/τ_{e^-}	μ^-/τ_{μ^-}	μ^-/τ_{μ^-}	e^-/τ_{e^-}
BR(%)		0.04	0.01	0.01	0.09	0.09
σ_{sig}[fb]		0.03	0.01	0.01	0.01	0.01
N_{event}		3.81	0.96	0.95	0.95	0.94

lepton pair		SS-6	SSSF		SSDF	
decay channel		1 - 4	1	2	3	4
H	W^+	l^+/τ_{l^+}	e^+/τ_{e^+}	μ^+/τ_{μ^+}	e^+/τ_{e^+}	μ^+/τ_{μ^+}
	W^-	τ_h^-	τ_h^-	τ_h^-	τ_h^-	τ_h^-
t	W	l^+/τ_{l^+}	e^+/τ_{e^+}	μ^+/τ_{μ^+}	μ^+/τ_{μ^+}	e^+/τ_{e^+}
BR(%)		0.06	0.02	0.02	0.02	0.02
σ_{sig}[fb]		0.05	0.01	0.01	0.01	0.01
N_{event}		6.55	1.65	1.64	1.64	1.62

Table C.1: All possible flavor combinations (SF and DF) of each OS and SS light lepton pair configuration in Table 6.6.

APPENDIX D

Analytical solutions for the neutrino 4-momentum vector in the decay of a top quark

In this section, detailed steps of analytically solving the 4-momentum vector of the neutrino from the top quark is given.

D.1 Invariant mass equation of the W boson

The invariant mass constraint on the W boson, equation 7.4 assuming $m_\nu = m_l = 0$, can be written as below:

$$\begin{aligned} m_W^2 &= (P_l + P_\nu)^2 \\ &= m_l^2 + m_\nu^2 + 2P_l \cdot P_\nu \\ &\cong 2(E_l \cdot E_\nu - \vec{p_l} \cdot \vec{p_l})\,. \end{aligned} \tag{D.1}$$

If we expand the equation D.1, and make the left-hand side equal to 0, then it can simplified into an equation that can be expressed in terms of x and y:

$$\begin{aligned} 0 &= -\frac{1}{2}m_W^2 + E_l \cdot E_\nu - \vec{p_l^x} \cdot \vec{p_\nu^x} - \vec{p_l^y} \cdot \vec{p_\nu^y} - \vec{p_l^z} \cdot \vec{p_\nu^z} \\ &= -\frac{1}{2}\frac{m_W^2}{E_l} + \sqrt{(p_\nu^x)^2 + (p_\nu^y)^2 + (p_\nu^z)^2} - \frac{\vec{p_l^x}}{E_l} \cdot \vec{p_\nu^x} - \frac{\vec{p_l^y}}{E_l} \cdot \vec{p_\nu^y} - \frac{\vec{p_l^z}}{E_l} \cdot \vec{p_\nu^z} \\ &= -\frac{1}{2}\frac{m_W^2}{E_l} - \frac{\vec{p_l^z}}{E_l} \cdot \vec{p_\nu^z} + \sqrt{x^2 + y^2 + (p_\nu^z)^2} - \frac{\vec{p_l^x}}{E_l} \cdot x - \frac{\vec{p_l^y}}{E_l} \cdot y \quad (x \equiv \vec{p_l^x}\,, \quad y \equiv \vec{p_l^y}) \\ &= R + \sqrt{x^2 + y^2 + P} - a \cdot x - b \cdot y\,. \end{aligned} \tag{D.2}$$

with following coefficients that are defined as:

$$\begin{aligned} R &\equiv -\frac{1}{2}\frac{m_W^2}{E_l} - \frac{\vec{p_l^z}}{E_l} \cdot \vec{p_\nu^z}\,, \\ a &\equiv \frac{\vec{p_l^x}}{E_l}\,, \\ b &\equiv \frac{\vec{p_l^y}}{E_l}\,, \\ P &\equiv (p_\nu^z)^2\,. \end{aligned} \tag{D.3}$$

Then we rewrite equation D.2 and take the square of it:

$$\begin{aligned} (ax + by - R)^2 &= x^2 + y^2 + P \\ 0 &= a^2x^2 + b^2y^2 + R^2 + (2ab)xy - (2aR)x - (2bR)y - x^2 - y^2 - P \\ &= (a^2 - 1)x^2 - (2aR)x + (2ab)xy + (b^2 - 1)y^2 - (2bR)y + R^2 - P \\ &= a_1x^2 + b_1x + c_1xy + d_1y^2 + e_1y + f_1\,, \end{aligned} \tag{D.4}$$

where the coefficients are:

$$\begin{aligned} a_1 &\equiv a^2 - 1\,, \\ b_1 &\equiv -2aR\,, \\ c_1 &\equiv 2ab\,, \\ d_1 &\equiv b^2 - 1\,, \\ e_1 &\equiv -2bR\,, \\ f_1 &\equiv R^2 - P\,. \end{aligned} \tag{D.5}$$

If we solve the quadratic equation D.4 with respect to x, then there are two possible solutions:

$$x = \frac{-(b_1 + c_1y) \pm \sqrt{\mathcal{A}_1y^2 + \mathcal{B}_1y + C_1}}{2a_1} \tag{D.6}$$

with coefficients:

$$\begin{aligned} \mathcal{A}_1 &= c_1^2 - 4a_1d_1\,, \\ \mathcal{B}_1 &= 2b_1c_1 - 4a_1e_1\,, \\ C_1 &= b_1^2 - 4a_1f_1\,. \end{aligned} \tag{D.7}$$

D.2 Invariant mass equation of the top-quark

We solve the second constraint, the invariant mass equation of the top-quark, equation 7.3:

$$\begin{aligned} m_t^2 &= (P_b + P_l + P_\nu)^2 \\ &= m_b^2 + m_l^2 + m_\nu^2 + 2(P_b \cdot P_l + P_b \cdot P_\nu + P_l \cdot P_\nu) \\ &\cong 2(P_b \cdot P_l + P_b \cdot P_\nu + P_l \cdot P_\nu)\,. \end{aligned} \tag{D.8}$$

We expand the equation D.8 as we did for W boson:

$$\begin{aligned} 0 &= -\frac{1}{2}m_t^2 + (E_b \cdot E_l - \vec{p_b} \cdot \vec{p_l}) + (E_b + E_l)E_\nu - (\vec{p_b} + \vec{p_l}) \cdot \vec{p_\nu} \\ &= R_2 + (E_b + E_l)E_\nu - (\vec{p_b} + \vec{p_l}) \cdot \vec{p_\nu} \\ &= \frac{R_2}{E_b + E_l} + E_\nu - \frac{\vec{p_b^x} + \vec{p_l^x}}{E_b + E_l} \cdot \vec{p_\nu^x} - \frac{\vec{p_b^y} + \vec{p_l^y}}{E_b + E_l} \cdot \vec{p_\nu^y} - \frac{\vec{p_b^z} + \vec{p_l^z}}{E_b + E_l} \cdot \vec{p_\nu^z} \\ &= \frac{R_2 - (\vec{p_b^z} + \vec{p_l^z}) \cdot \vec{p_\nu^z}}{E_b + E_l} + \sqrt{x^2 + y^2 + (p_\nu^z)^2} - a'x - b'y \\ &= R_2' + \sqrt{x^2 + y^2 + P} - a'x - b'y\,, \end{aligned} \tag{D.9}$$

where coefficients are defined as below:

$$\begin{aligned} R_2 &\equiv -\frac{1}{2}m_t^2 + (E_b \cdot E_l - \vec{p_b} \cdot \vec{p_l})\,, \\ R_2' &\equiv \frac{R_2 - (\vec{p_b^z} + \vec{p_l^z}) \cdot \vec{p_\nu^z}}{E_b + E_l}\,, \\ a' &\equiv \frac{\vec{p_b^x} + \vec{p_l^x}}{E_b + E_l}\,, \\ b' &\equiv \frac{\vec{p_b^y} + \vec{p_l^y}}{E_b + E_l}\,. \end{aligned} \tag{D.10}$$

If we take the square of the equation D.9 and simplify, then we get:

$$0 = a_2x^2 + b_2x + c_2xy + d_2y^2 + e_2y + f_2\,, \tag{D.11}$$

where

$$\begin{aligned} a_2 &\equiv a'^2 - 1\,, \\ b_2 &\equiv -2a'R_2'\,, \\ c_2 &\equiv 2a'b'\,, \\ d_2 &\equiv b'^2 - 1\,, \\ e_2 &\equiv -2b'R_2'\,, \\ f_2 &\equiv R_2'^2 - P\,. \end{aligned} \tag{D.12}$$

Similar to equation D.6, there are two solutions for x in equation D.11,

$$x = \frac{-(b_2 + c_2 y) \pm \sqrt{\mathcal{A}_2 y^2 + \mathcal{B}_2 y + C_2}}{2a_2} \tag{D.13}$$

with coefficients:

$$\begin{aligned} \mathcal{A}_2 &= c_2^2 - 4a_2 d_2\,, \\ \mathcal{B}_2 &= 2b_2 c_2 - 4a_2 e_2\,, \\ C_2 &= b_2^2 - 4a_2 f_2\,. \end{aligned} \tag{D.14}$$

D.3 Solving the y-component of the neutrino 4-momentum vector

Since we need to find solutions which satisfy both invariant mass equations, we set the equation D.6 and equation D.13 as equal:

$$\begin{aligned} \left(-b_1 - c_1 y \pm \sqrt{\mathcal{A}_1 y^2 + \mathcal{B}_1 y + C_1}\right) a_2 &= \left(-b_2 - c_2 y \pm \sqrt{\mathcal{A}_2 y^2 + \mathcal{B}_2 y + C_2}\right) a_1 \\ (c_2 a_1 - c_1 a_2) y + (b_2 a_1 - b_1 a_2) &= \pm\left(a_1 \sqrt{\mathcal{A}_2 y^2 + \mathcal{B}_2 y + C_2} - a_2 \sqrt{\mathcal{A}_1 y^2 + \mathcal{B}_1 y + C_1}\right) \\ \beta y + \alpha &= \pm(a_1 \sqrt{\lambda_2} - a_2 \sqrt{\lambda_1})\,, \end{aligned} \tag{D.15}$$

which is expressed with coefficients of

$$\begin{aligned} \alpha &= b_1 a_2 - b_2 a_1 \\ \beta &= c_1 a_2 - c_2 a_1 \\ \lambda_1 &= \mathcal{A}_1 y^2 + \mathcal{B}_1 y + C_1 \\ \lambda_2 &= \mathcal{A}_2 y^2 + \mathcal{B}_2 y + C_2\,, \end{aligned} \tag{D.16}$$

We square the both sides of equation D.15 and rearrange:

$$\begin{aligned} (\beta^2 - \mathcal{A}_2 a_1^2 - \mathcal{A}_1 a_2^2) \cdot y^2 \\ (2\alpha\beta - \mathcal{B}_2 a_1^2 - \mathcal{B}_1 a_2^2) \cdot y \\ (\alpha^2 - C_2 a_1^2 - C_1 a_2^2) = -2a_1 a_2 \sqrt{\lambda_1 \cdot \lambda_2} \\ \psi_1 y^2 + \psi_2 y + \psi_3 = -2a_1 a_2 \sqrt{\lambda_1 \cdot \lambda_2} \end{aligned} \tag{D.17}$$

with coefficients:

$$\begin{aligned} \psi_1 &= \beta^2 - \mathcal{A}_2 a_1^2 - \mathcal{A}_1 a_2^2 \\ \psi_2 &= 2\alpha\beta - \mathcal{B}_2 a_1^2 - \mathcal{B}_1 a_2^2 \\ \psi_3 &= \alpha^2 - C_2 a_1^2 - C_1 a_2^2\,. \end{aligned} \tag{D.18}$$

We take the square of both sides again to remove the square root, and after the simplification we get:

$$\begin{aligned}
&\psi_1 y^4 + 2\psi_1\psi_2 y^3 + (2\psi_1\psi_3 + \psi_2^2)y^2 + 2\psi_2\psi_3 y + \psi_3^2 \\
&\qquad = 4a_1^2 a_2^2 \lambda_1 \lambda_2 \\
&\qquad = 4a_1^2 a_2^2 (\mathcal{A}_1 y^2 + \mathcal{B}_1 y + C_1)(\mathcal{A}_2 y^2 + \mathcal{B}_2 y + C_2) \\
&\qquad = \gamma[\mathcal{A}_1\mathcal{A}_2 y^4 + (\mathcal{A}_1\mathcal{B}_2 + \mathcal{B}_1\mathcal{A}_2)y^3 \\
&\qquad\quad + (\mathcal{A}_1 C_2 + \mathcal{B}_1\mathcal{B}_2 + C_1\mathcal{A}_2)y^2 + (\mathcal{B}_1 C_2 + C_1\mathcal{B}_2)y + C_1 C_2]\,,
\end{aligned} \tag{D.19}$$

and finally we get a quartic equation in terms of y:

$$\begin{aligned}
0 = &(\psi_1 - \gamma\mathcal{A}_1\mathcal{A}_2)\cdot y^4 \\
&+ [2\psi_1\psi_2 - \gamma(\mathcal{A}_1\mathcal{B}_2 + \mathcal{B}_1\mathcal{A}_2)]\cdot y^3 \\
&+ [(2\psi_1\psi_3 + \psi_2^2) - \gamma(\mathcal{A}_1 C_2 + \mathcal{B}_1\mathcal{B}_2 + C_1\mathcal{A}_2)]\cdot y^2 \\
&+ [2\psi_2\psi_3 - \gamma(\mathcal{B}_1 C_2 + C_1\mathcal{B}_2)]\cdot y \\
&+ (\psi_3^2 - \gamma C_1 C_2)\,.
\end{aligned} \tag{D.20}$$

Basically, for a quartic equation which is expressed as:

$$A\cdot y^4 + B\cdot y^3 + C\cdot y^2 + D\cdot y + E = 0\,. \tag{D.21}$$

there are 4 roots:

$$y_{1,2} = -\frac{B}{4A} - S \pm \frac{1}{2}\sqrt{-4S^2 - 2p + \frac{q}{S}} \tag{D.22}$$

$$y_{3,4} = -\frac{B}{4A} + S \pm \frac{1}{2}\sqrt{-4S^2 - 2p - \frac{q}{S}} \tag{D.23}$$

with coefficients:

$$\begin{aligned}
p &= \frac{8AC - 3B^2}{8A^2} \\
q &= \frac{B^3 - 4ABC + 8A^2 D}{8A^3} \\
S &= \frac{1}{2}\sqrt{-\frac{2}{3}P + \frac{1}{3A}(Q + \frac{\Delta_0}{Q})} \\
Q &= \sqrt[3]{\frac{\Delta_1 + \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}} \\
\Delta_0 &= C^2 - 3BD + 12AE \\
\Delta_1 &= 2C^3 - 9BCD + 27B^2E + 27B^2E + 27AD^2 - 72ACE
\end{aligned} \tag{D.24}$$

Thus, we can solve the y in equation D.20 as equation D.21 with coefficient replaced as:

$$\begin{aligned} A &= \psi_1 - \gamma \mathcal{A}_1 \mathcal{A}_2 \\ B &= 2\psi_1\psi_2 - \gamma(\mathcal{A}_1\mathcal{B}_2 + \mathcal{B}_1\mathcal{A}_2) \\ C &= (2\psi_1\psi_3 + \psi_2^2) - \gamma(\mathcal{A}_1 C_2 + \mathcal{B}_1\mathcal{B}_2 + C_1\mathcal{A}_2) \\ D &= 2\psi_2\psi_3 - \gamma(\mathcal{B}_1 C_2 + C_1\mathcal{B}_2) \\ E &= \psi_3^2 - \gamma C_1 C_2 \end{aligned} \tag{D.25}$$

D.4 Solving the x-component of the neutrino 4-momentum vector

With the y value known, we can obtain x values by recalling the equation for x values from each invariant mass equation:

$$x_1^{\pm} = \frac{-(b_1 + c_1 y) \pm \sqrt{\mathcal{A}_1 y^2 + \mathcal{B}_1 y + C_1}}{2a_1},$$

for the equation D.6 and

$$x_2^{\pm} = \frac{-(b_2 + c_2 y) \pm \sqrt{\mathcal{A}_2 y^2 + \mathcal{B}_2 y + C_2}}{2a_2}.$$

for equation D.13. So there are 4 x solutions, x_1^+, x_1^-, x_2^+, x_2^- for each y value. However, not all of these x values are the solution for corresponding y value, as we are solving a system of two simultaneous equations. One of the x values from the invariant mass equation of the W boson, x_1^+ or x_1^-, must be equal to one of the x values from the invariant mass equation of the top quark, x_2^+ or x_2^-. So only the solution which satisfies the condition:

$$x_1^{\pm} = x_2^{\pm}, \tag{D.26}$$

is selected to be the solution set, i.e. $\{x_i, y_i\}$ with $i = 1, 2, 3, 4$ for the x - and y-components of the neutrino four-momentum vector.

APPENDIX E

Additional tables and plots for the reconstruction of $tHq \rightarrow t\tau_l\tau_h$ channel

In this appendix, additional tables and plots for the reconstruction of the $tHq \rightarrow t\tau_l\tau_h$ channel are presented.

E.1 Additional tables for the dependency of the number of solutions on α

α	total	no sol	1 sol	2 sol	3 sol	4 sol	real sol	imag sol	no/total (%)	real/imag (%)
−1.00	212	132	0	80	0	0	30	50	62.26	37.50
−0.50	212	89	0	123	0	0	71	52	41.98	57.72
0.00	212	37	0	175	0	0	100	75	17.45	57.14
0.25	212	27	0	185	0	0	122	63	12.74	65.95
0.50	212	25	0	187	0	0	129	58	11.79	68.98
0.75	212	21	0	191	0	0	131	60	9.91	68.59
1.00	212	22	0	190	0	0	140	50	10.38	73.68
2.00	212	25	0	187	0	0	129	58	11.79	68.98
3.00	212	30	0	182	0	0	105	77	14.15	59.38

Table E.1: A summary of the frequency of different number of solutions (raw number of events) for a system of two invariant mass equations, for the tHq samples including all campaigns. This is done with the anti-top removal. The uncertainties are not shown, which are the square root of the number of events assuming a Poisson distribution.

α	total	no sol	1 sol	2 sol	3 sol	4 sol	real sol	imag sol	no/total (%)	real/imag (%)
−1.00	44 483	21 837	0	22 646	0	0	12 702	9 944	49.09	56.09
−0.50	44 483	17 756	0	26 727	0	0	15 929	10 798	39.92	59.60
0.00	44 483	7 313	0	37 170	0	0	23 881	13 289	16.44	64.25
0.25	44 483	6 572	0	37 911	0	0	25 076	12 835	14.77	66.14
0.50	44 483	5 551	0	38 932	0	0	26 728	12 204	12.48	68.65
0.75	44 483	3 975	0	40 508	0	0	29 110	11 398	8.94	71.86
1.00	44 483	3 769	2	40 712	0	0	29 047	11 667	8.47	71.34
2.00	44 483	5 551	0	38 932	0	0	26 728	12 204	12.48	68.98
3.00	44 483	4 358	1	40 124	0	0	23 827	16 298	9.80	59.38

Table E.2: A summary of the frequency of different number of solutions (raw number of events) for a system of two invariant mass equations, for the $t\bar{t}$ samples including all campaigns. This is done with the anti-top removal. The uncertainties are not shown, which are the square root of the number of events assuming a Poisson distribution.

α	total	no sol	1 sol	2 sol	3 sol	4 sol	real sol	imag sol	no/total (%)	real/imag (%)
−1.00	156 275	73 210	1	83 063	0	1	44 976	38 089	49.85	54.15
−0.50	156 275	77 158	0	79 116	0	1	43 606	35 511	49.37	55.12
0.00	156 275	48 095	1	108 178	0	1	62 715	45 465	30.78	57.97
0.25	156 275	37 852	1	118 423	0	1	69 684	48 739	24.22	58.84
0.50	156 275	33 626	0	122 649	0	0	71 822	50 827	21.52	58.56
0.75	156 275	31 255	1	125 019	0	0	72 127	52 893	20.00	57.69
1.00	156 275	29 777	2	126 496	0	0	71 563	54 935	19.05	56.57
2.00	156 275	27 561	1	128 713	0	0	65 923	62 791	17.64	51.22
3.00	156 275	28 715	1	127 559	0	0	59 628	67 932	18.37	46.75

Table E.3: A summary of the frequency of different number of solutions (raw number of events) for a system of two invariant mass equations, for the tZq samples including all campaigns. This is done without the anti-top removal. The uncertainties are not shown, which are the square root of the number of events assuming a Poisson distribution.

α	total	no sol	1 sol	2 sol	3 sol	4 sol	real sol	imag sol	no/total (%)	real/imag (%)
−1.00	366	180	0	186	0	0	95	91	46.85	54.15
−0.50	366	148	0	218	0	0	126	92	40.44	57.80
0.00	154	25	0	175	0	0	100	75	17.45	57.14
0.25	366	46	0	320	0	0	215	105	12.57	67.19
0.50	366	39	0	327	0	0	219	108	10.66	66.97
0.75	366	38	0	328	0	0	223	105	10.38	67.99
1.00	366	39	0	327	0	0	230	97	10.66	70.34
2.00	366	37	0	329	0	0	200	129	10.11	60.79
3.00	366	49	0	317	0	0	186	131	13.39	58.68

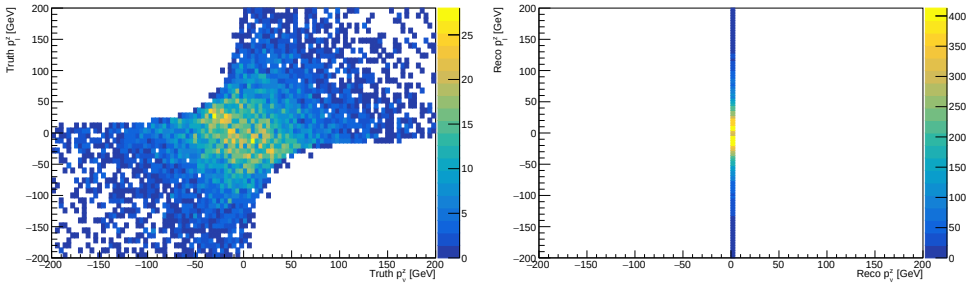
Table E.4: A summary of the frequency of different number of solutions (raw number of events) for a system of two invariant mass equations, for the tHq samples including all campaigns. This is done without the anti-top removal. The uncertainties are not shown, which are the square root of the number of events assuming a Poisson distribution.

α	total	no sol	1 sol	2 sol	3 sol	4 sol	real sol	imag sol	no/total (%)	real/imag (%)
−1.00	86 047	17 881	0	68 166	0	0	46 920	21 246	20.78	68.83
−0.50	86 047	37 142	0	48 905	0	0	28 214	20 691	43.16	57.69
0.00	86 047	14 009	1	72 037	0	0	46 209	25 829	16.28	64.15
0.25	86 047	26 833	1	59 214	0	0	38 247	20 967	31.18	64.59
0.50	86 047	8 331	1	77 715	0	0	55 612	22 104	9.68	66.97
0.75	86 047	7 610	0	78 437	0	0	56 444	21 993	8.84	67.99
1.00	86 047	7 533	3	78 511	0	0	56 068	22 446	8.75	71.41
2.00	86 047	6 929	0	79 118	0	0	51 879	27 239	8.05	65.57
3.00	86 047	8 349	4	77 694	0	0	46 316	31 382	9.70	59.61

Table E.5: A summary of the frequency of different number of solutions (raw number of events) for a system of two invariant mass equations, for the $t\bar{t}$ samples including all campaigns. This is done without the anti-top removal. The uncertainties are not shown, which are the square root of the number of events assuming a Poisson distribution.

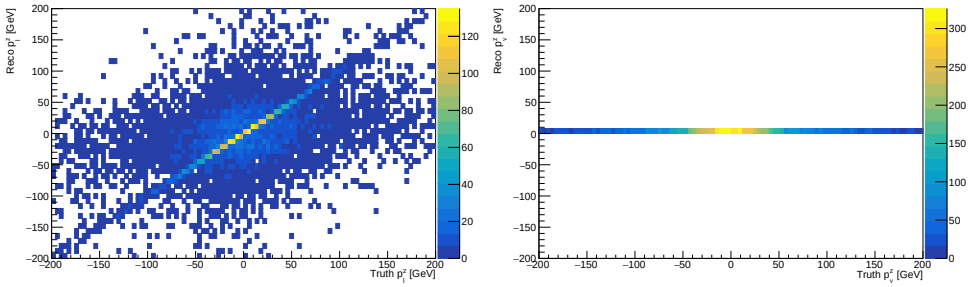
E.2 Additional plots for the truth-matching studies

In this section, additional plots for the truth-matching studies that have been performed with different α values are presented. The α values with $0, -0.5, -1.0$ are shown. As you can see from the plots, the number of events decreases with a decreasing α value (see Section 7.2.4), and there are some cutoffs in the plots which involve the reconstructed quantities. The reason why the distributions of the reco-level variables are cut away in this way, unlike the distributions of the truth-level variables which are faded away (no cutoff but become less dense) with the decreasing α is still unsolved. One possible explanation could be that the mass constraint on the W boson from which the light lepton and the neutrino decay is constraining its decay products into a specific topology. Additionally, it can be seen that the overall tendency of some quantities shown in the plots does not vary much depending on the α value, such as the η-related distributions.



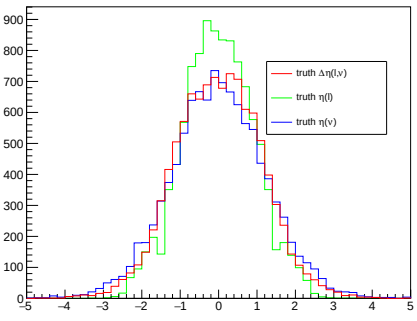
(a) Truth p_l^z vs. Truth p_ν^z

(b) Reco. p_l^z vs. Reco. p_ν^z

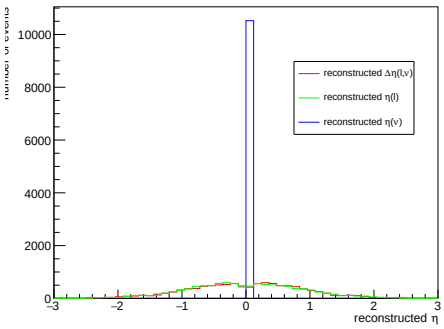


(c) Reco. p_l^z vs. Truth p_l^z

(d) Reco. p_ν^z vs Truth p_ν^z

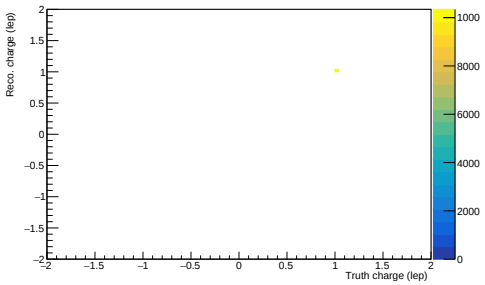


(e) Truth $\eta_l, \eta_\nu, \Delta\eta\ (l, \nu)$

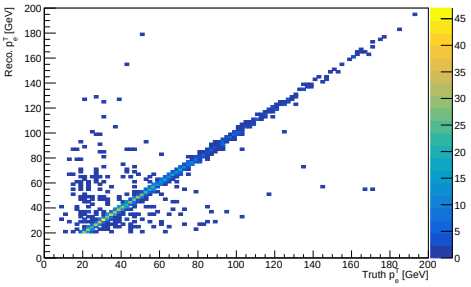


(f) Reco. $\eta_l, \eta_\nu, \Delta\eta\ (l, \nu)$

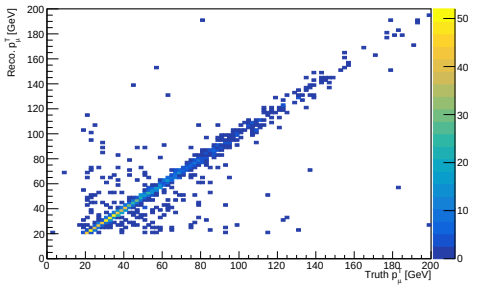
Figure E.1: The distributions showing the correlations of important truth-matching quantities between the truth-level and the reco-level, as well as both of the levels themselves, with $\alpha = 0$. The mc16a tZq sample at $\alpha = 0$ with the anti-top removal is used. The removal of the anti-top can be proven by Fig E.2(a) which shows only the values corresponding to the charge of the light lepton from the top quark.



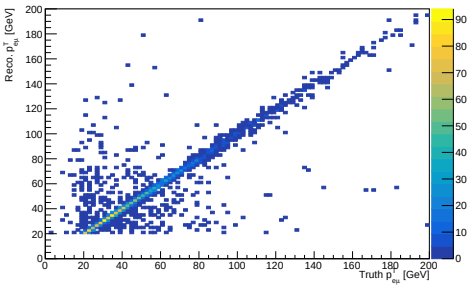
(a) Reco. l charge vs. Truth l charge



(b) Reco. p_e^{T} vs. Truth p_e^{T}

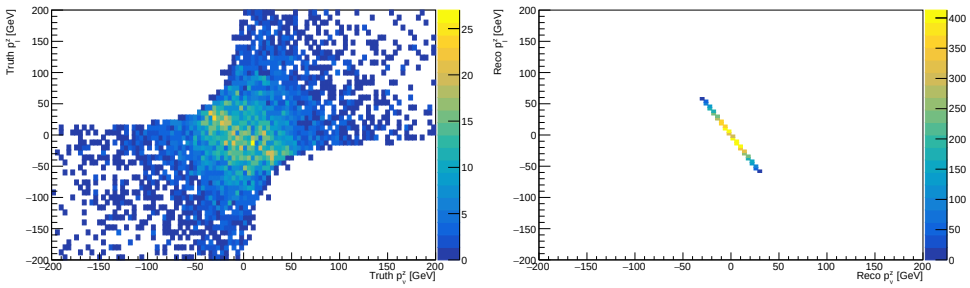


(c) Reco. p_μ^{T} vs. Truth p_μ^{T}



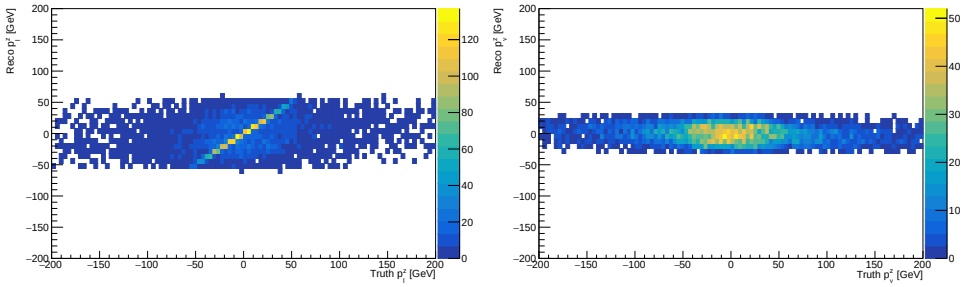
(d) Reco. $p_{(e,\mu)}^{\mathrm{T}}$ vs. Truth $p_{(e,\mu)}^{\mathrm{T}}$

Figure E.2: The distributions showing the correlations of important truth-matching quantities between the truth-level and the reco-level, as well as both of the levels themselves, with $\alpha = 0$. The mc16a tZq sample at $\alpha = 0$ with the anti-top removal is used. The removal of the anti-top can be proven by Fig E.2(a) which shows only the values corresponding to the charge of the light lepton from the top quark.



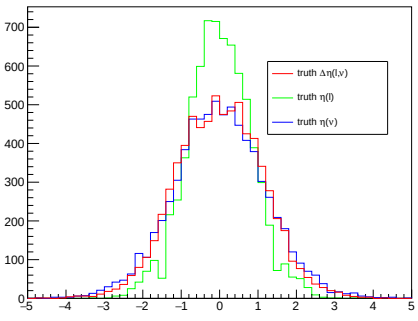
(a) Truth p_l^z vs. Truth p_ν^z

(b) Reco. p_l^z vs. Reco. p_ν^z

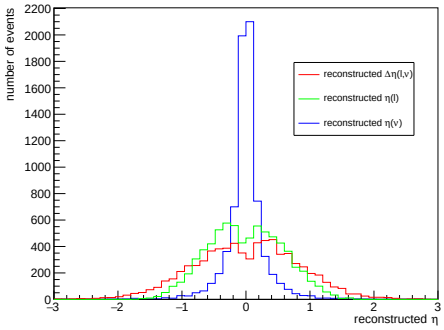


(c) Reco. p_l^z vs. Truth p_l^z

(d) Reco. p_ν^z vs Truth p_ν^z

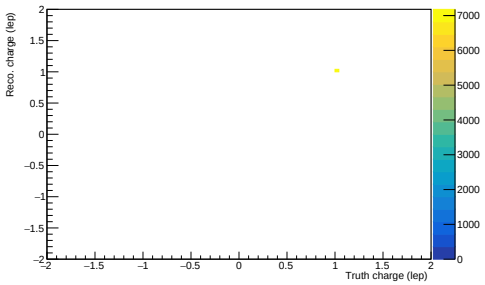


(e) Truth $\eta_l, \eta_\nu, \Delta\eta\ (l, \nu)$

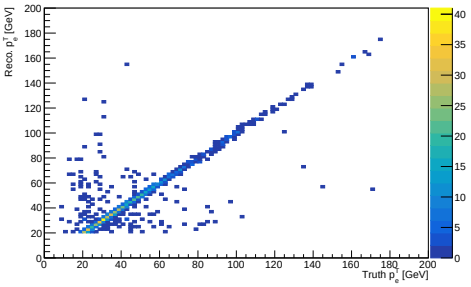


(f) Reco. $\eta_l, \eta_\nu, \Delta\eta\ (l, \nu)$

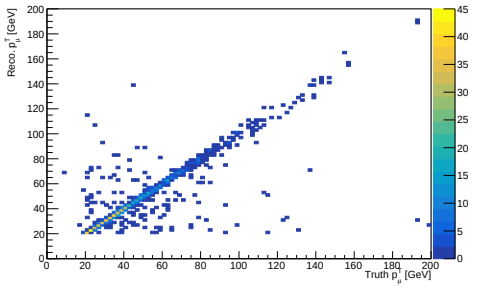
Figure E.3: The distributions showing the correlations of important truth-matching quantities between the truth-level and the reco-level, as well as both of the levels themselves, with $\alpha = -0.5$. The mc16a tZq sample at $\alpha = -0.5$ with the anti-top removal is used. The removal of the anti-top can be proven by Fig E.4(a) which shows only the values corresponding to the charge of the light lepton from the top quark.



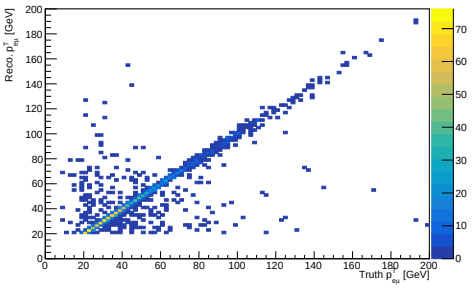
(a) Reco. l charge vs. Truth l charge



(b) Reco. p_e^{T} vs. Truth p_e^{T}

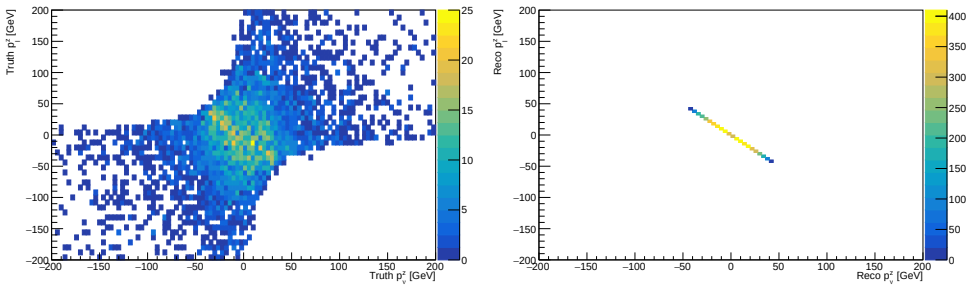


(c) Reco. p_μ^{T} vs. Truth p_μ^{T}



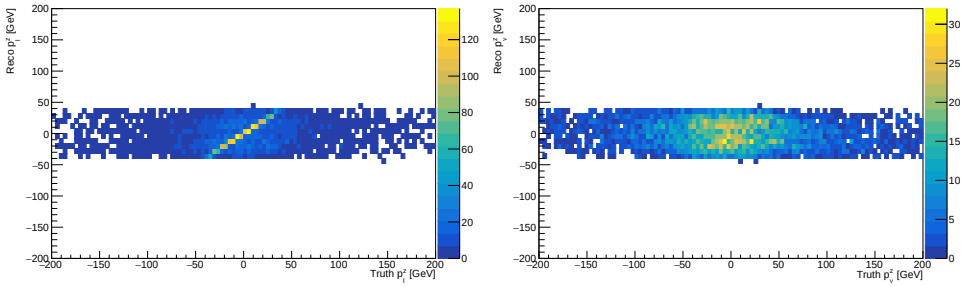
(d) Reco. $p_{(e,\mu)}^{\mathrm{T}}$ vs. Truth $p_{(e,\mu)}^{\mathrm{T}}$

Figure E.4: The distributions showing the correlations of important truth-matching quantities between the truth-level and the reco-level, as well as both of the levels themselves, with $\alpha = -0.5$. The mc16a tZq sample at $\alpha = -0.5$ with the anti-top removal is used. The removal of the anti-top can be proven by Fig E.4(a) which shows only the values corresponding to the charge of the light lepton from the top quark.



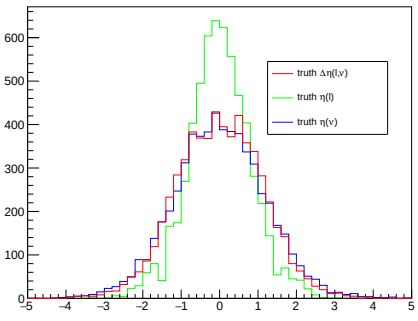
(a) Truth p_l^z vs. Truth p_ν^z

(b) Reco. p_l^z vs. Reco. p_ν^z

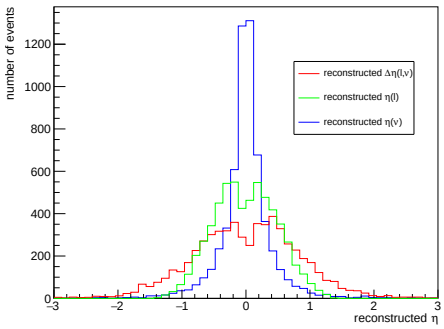


(c) Reco. p_l^z vs. Truth p_l^z

(d) Reco. p_ν^z vs Truth p_ν^z

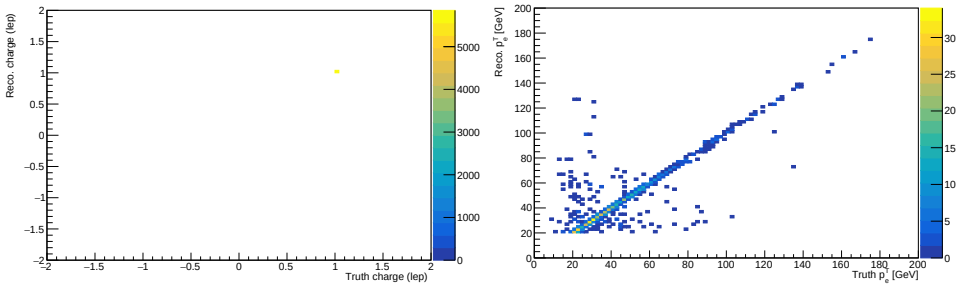


(e) Truth $\eta_l, \eta_\nu, \Delta\eta\ (l, \nu)$



(f) Reco. $\eta_l, \eta_\nu, \Delta\eta\ (l, \nu)$

Figure E.5: The distributions showing the correlations of important truth-matching quantities between the truth-level and the reco-level, as well as both of the levels themselves, with $\alpha = -1.0$. The mc16a tZq sample at $\alpha = -1.0$ with the anti-top removal is used. The removal of the anti-top can be proven by Fig E.6(a) which shows only the values corresponding to the charge of the light lepton from the top quark.



(a) Reco. l charge vs. Truth l charge

(b) Reco. p_e^{T} vs. Truth p_e^{T}

(c) Reco. p_μ^{T} vs. Truth p_μ^{T}

(d) Reco. $p_{(e,\mu)}^{\mathrm{T}}$ vs. Truth $p_{(e,\mu)}^{\mathrm{T}}$

Figure E.6: The distributions showing the correlations of important truth-matching quantities between the truth-level and the reco-level, as well as both of the levels themselves, with $\alpha = -1.0$. The mc16a tZq sample at $\alpha = -1.0$ with the anti-top removal is used. The removal of the anti-top can be proven by Fig E.6(a) which shows only the values corresponding to the charge of the light lepton from the top quark.

List of Figures

List of Tables