Investigating the Use of the Matrix-Element-Method in Reducing the Top-Antitop Quark Background to Higgs Pair Production Processes

Master's Thesis at the Faculty of Physics
of the
Ludwig-Maximilians-University of Munich

submitted by

Patrick Rudolph Schumacher

born in Buffalo, NY, USA

Munich, on 13.07.2026

Untersuchung zur Anwendung der Matrixelementmethode zur Reduzierung der Top-Antitop-Quark Hintergrund von Higgs-Paar-Produktionsprozesse

Masterarbeit der Fakultät für Physik
der
Ludwig-Maximilians-Universität München

vorgelegt von

Patrick Rudolph Schumacher

geboren in Buffalo, NY, USA

München, den 13.07.2026

Abstract

The discovery of the Higgs boson meant that all particles predicted by the Standard Model of Particle Physics had been found. Still, some further predictions continue to evade examination, such as Higgs self-coupling, which is both rare and difficult to detect. In particular, the di-Higgs decay process $HH \to b\bar{b}W^+W^-$ must be distinguished from a large $t\bar{t}$ background. This thesis examines whether the matrix element method as implemented via program MadWeight can be used for this purpose. MadWeight proved to be highly sensitive in its execution which, alongside difficulties with particle identification and stronger-than-expected statistical effects, threatened to limit its utility. However, the introduction of unfolding methods to correct for these statistical effects was extremely successful, leaving the door open for future successful implementation after all.

Contents

Chapter 1

Introduction

The discovery of the Higgs boson by the ATLAS and CMS collaborations at the Large Hadron Collider (LHC) in 2012 was the capstone of decades of research in particle physics. It validated the proposed Higgs mechanism as an explanation for the mass of the W- and Z-bosons in particular, as well as the predictive ability of the Standard Model of particle physics in general [1], as, with the discovery of the Higgs boson, all of the particles predicted by the Standard Model have been found [2].

Nevertheless, many further predictions of the Standard Model remain untested. For example, the Standard Model also predicts the self-coupling of the Higgs boson as well as the value of the coupling constants λ for both trilinear and quartic Higgs self-interaction [2]. An experimental confirmation of these values would further strengthen the current Standard Model, while a deviation from the theoretical values would indicate the need for new physics [2,3]. Either carries wide-ranging implications for various phenomena at both electroweak as well as higher energy scales [2].

As a result, Higgs self-coupling remains a subject of great interest and ongoing investigations at the LHC [3,4]. Yet it is difficult to examine experimentally; the quartic coupling is, for one, largely inaccesible at the LHC due to the small cross-section of the relevant events, and even the trilinear coupling — as observed in Higgs pair production and subsequent decays — is rather rare and must simultaneously compete with large, more common backgrounds in order to be detected in the first place [2]. Thus, while research can produce a general range for the ratio of the experimental trilinear coupling constant to the theoretical value, κ_λ, a well-defined value for the coupling itself remains out of reach [2–4]. In order to get closer to this value, better measurements of Higgs pairs are required, and this necessitates a better identification of them.

Different Higgs pair decay modes have different signatures, and therefore require different methods to attempt to isolate them from their respective backgrounds. This is a particular challenge for $HH \to b\bar{b}W^+W^-$ decays; although it has the second-highest branching ratio of all Higgs pair decay modes and therefore is an attractive target for further study, it has generally been neglected compared to other decay modes due to the large $t\bar{t}$ background from which it needs to be isolated [2].

This thesis explores whether the matrix element method, and in particular its algorithmic implementation via MadWeight, can be used to do so. The matrix element method generates a weight (proportional to the likelihood) that relates experimental results to the governing theoretical framework. In hadron colliders, this weight can be calculated using the squared matrix element [5], giving the method its name. The hope is that potential differences in the weights for a given event can be used to distinguish between these two decay processes, whose final states have identical particle content.

Chapter 2 gives an overview of the theoretical background of Higgs self-coupling, the difficulty in experimentally observing it, and the matrix element method as a possible solution. Chapter 3 outlines the data used and the treatment it undergoes in order to be properly used to calculate the weights via MadWeight. Chapter 4 then examines the results of these calculations and offers analysis attempting to explain them. Chapter 5 describes a change in approach based on these findings, with the preliminary results of this change being covered in Chapter 6. Chapter 7 summarizes the overall work of this thesis and lays out possible next steps.

Chapter 2

Background

2.1 The Standard Model of Particle Physics

The Standard Model provides the now well-established fundamental framework for modern particle physics. Using the methodology of quantum field theory as a unifying medium, it describes the strong, weak, and electromagnetic forces in terms of particle fields, noting that the Standard Model manages to further unify the latter two into a single electroweak theory as well. These different components also correspond to different gauge symmetry groups, altogether comprising the $SU(3) \times SU(2) \times U(1)$ gauge group of the Standard Model. In principle, it establishes that spin $S = \frac{1}{2}$ fermions (and corresponding anti-particles) experience these forces as mediated by spin $S = 1$ bosons, though some bosons are capable of self-interaction and/or interacting with other bosons [6].

The strong component of the Standard Model is described by Quantum Chromodynamics (QCD) and adheres to $SU(3)$ symmetry. The strong force acts through gluon fields and governs the interactions of quarks with gluons as well as gluon self-interaction. Quarks come in six different flavors — up (u), down (d), strange (s), charm (c), bottom (b) and top (t) — and carry one of three color charges. Correspondingly, there are $3^2 - 1 = 8$ types of colored gluons [1], [6].

The electroweak component, on the other hand, is based on the $SU(2) \times U(1)$ group. The gauge bosons of the theory are $W^i_\mu, i = 1, 2, 3$ and B_μ, which correspond to the $SU(2)$ and $U(1)$ groups respectively; however, unlike the gluons in QCD, these do not correspond to physical particles, which are instead combinations thereof:

$$A_\mu = B_\mu \cos\theta_W + W^3_\mu \sin\theta_W \tag{2.1a}$$

$$Z_\mu = -B_\mu \sin\theta_W + W^3_\mu \cos\theta_W \tag{2.1b}$$

$$W^\pm_\mu = \frac{W^1_\mu \mp iW^2_\mu}{\sqrt{2}} \tag{2.1c}$$

which are the photon and the Z- and two W-bosons, respectively. $\theta_W = \tan^{-1}(\frac{g'}{g})$ is the weak mixing angle, where g' and g are the gauge coupling constants associated with $U(1)$ and $SU(2)$ [1].

The photon field mediates electromagnetic interactions, while the weak force acts via the W- and Z-bosons. All charged fermions — both quarks and charged leptons, the latter of which consists of electrons, muons, and τ-leptons [6] — experience electromagnetic interactions as well as weak interactions mediated by the neutral Z-boson. Weak interactions involving either of the charged W-bosons, on the other hand, only occur with left-handed fermion doublets, which for each of the charged leptons includes its corresponding neutrino, which is electrically neutral [1], [6].

2.2 The Higgs Mechanism and the Higgs Boson

The electroweak sector of the Standard Model also includes the Higgs boson, which arises from the implementation of the Higgs mechanism within the Standard Model as a means of explaining the spontaneous breaking of the electroweak gauge symmetry required to account for the masses of the W- and Z-gauge bosons [1]. By introducing the complex scalar field doublet

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2}\phi^+ \\ \phi^0 \end{pmatrix} \tag{2.2}$$

and the associated Lagrangian

$$\mathcal{L}_H = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \tag{2.3}$$

where (noting that σ^α are the Pauli matrices and Y is the hypercharge quantum number)

$$D_\mu \phi = (\partial_\mu + \frac{ig}{2}\sigma^\alpha W_\mu^\alpha + \frac{ig'}{2} Y B_\mu)\phi \tag{2.4a}$$

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{2.4b}$$

to the electroweak sector of the Standard Model and minimizing the potential energy term V for $\mu^2 < 0$ results in a non-zero vacuum expectation value (VEV) for any ϕ satisfying $\langle \phi^\dagger \phi \rangle = \frac{-\mu^2}{2\lambda}$, or $\langle \phi^\dagger \phi \rangle = \frac{v^2}{2}$ when $v = \sqrt{\frac{-\mu^2}{\lambda}}$. If the expectation value of the neutral component of the doublet, $\langle \phi^0 \rangle$, is selected to be equal to this value of v, then the global minimum is simply:

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.5}$$

This sets a non-zero ground state, which breaks the symmetry of the electroweak Lagrangian as needed. Three of the generators of the electroweak gauge group $SU(2) \times U(1)$ are broken, corresponding to three hypothetical Goldstone bosons and three degrees of freedom of $SU(2)$. By adopting the unitary gauge, these can be absorbed by the W- and Z-bosons to give them masses, while leaving one remaining degree of freedom associated with the field. This can be parameterized as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ H + v \end{pmatrix} \tag{2.6}$$

where H is the Higgs field itself, which indicates the existence of the Higgs boson. The discovery of the Higgs boson in 2012 was a landmark in validating the consistency of the Standard Model [1].

The kinematic terms of the Higgs Lagrangian establish the coupling of the Higgs boson to the W- and Z-bosons [1], which has been measured at the LHC [2]. Turning to the potential terms, however, it can be seen that upon expanding with respect to (2.6)

$$\mathcal{L}_H^{potential} = -\lambda v^2 H^2 - \lambda v H^3 - \frac{\lambda}{4} H^4 \tag{2.7}$$

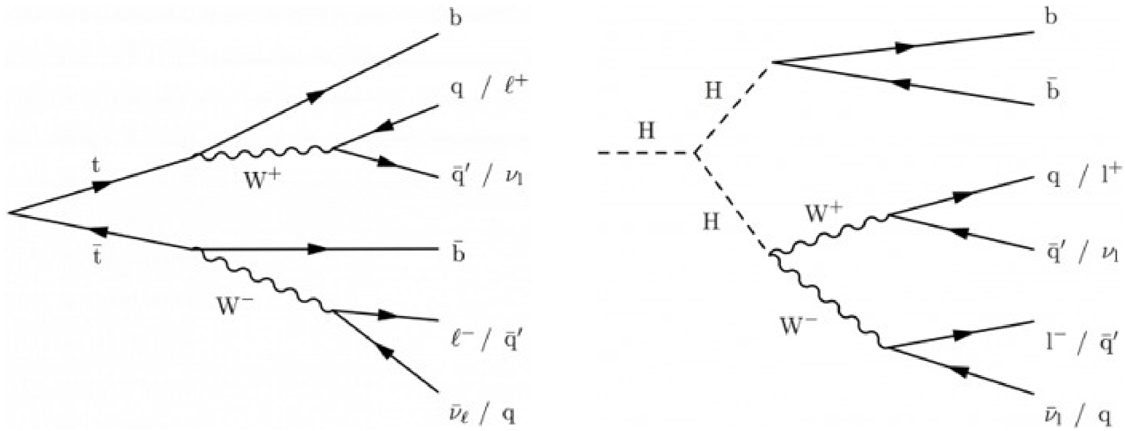
that the Higgs boson also couples to itself, both tri- and quadrilinearly [1]. Recognizing the first term as the Higgs mass term with $m_H^2 = 2\lambda v^2$, the values of these coupling constants λ can then be parametrized in terms of the Higgs mass and the VEV (which is further related to the measured Fermi coupling constant G_F by the relation $v = (\sqrt{2}G_F)^{-1/2}$). However, the true physical values have yet to be measured [1, 2]. While experimental examination of the quartic Higgs coupling λ_{HHHH} at the LHC is precluded due to the small cross-section associated with triple Higgs production, Higgs pair production does occur, and permits investigations into the trilinear Higgs coupling constant λ_{HHH} [2]. Recent experiments have been able to confine the coupling modifier $\kappa_\lambda = \lambda_{HHH}/\lambda_{HHH}^{SM}$, which quantifies the deviation of the observed coupling constant to that predicted by the Standard Model, to $-1.35 \leq \kappa_\lambda \leq 6.37$ (CMS, 2025) [3] and $-1.2 < \kappa_\lambda < 7.2$ (ATLAS, 2024) [4] at 95% confidence level.

Nevertheless, even this is still a difficult process. On the one hand, it is hampered by the rarity of Higgs-pair production. There are multiple Higgs-pair production modes, the largest of which are gluon fusion and vector-boson fusion, with cross-sections of about $\sigma_{ggF}^{HH} = 31.05^{+2.2\%}_{-5.0\%} \pm 3.0\%$ fb and $\sigma_{VBF}^{HH} = 1.73^{+0.03\%}_{-0.04\%} \pm 2.1\%$ fb at $\sqrt{s} = 13$ TeV, respectively [1] [2]. These values are still about three orders of magnitude smaller than for single Higgs production [7].

On the other hand, even when produced, Higgs-pair production events still need to be identified in order to be probed. This presents challenges of its own. Due to the short lifetime of the Higgs boson — with a decay width of $\Gamma_H = 4.1^{+0.7}_{-0.8}$ MeV [1], its mean lifetime is about $\tau = \frac{\hbar}{\Gamma} = 1.6 \times 10^{-22}$ s — it can only be identified by its decay products. However, these signals often have to compete with a strong background. $HH \to b\bar{b}b\bar{b}$ decays, which have the largest branching ratio of all di-Higgs decays at 33.9%, must be identified from a large multi-jet background; $HH \to b\bar{b}W^+W^-$ decays, despite having the second largest branching ratio of 24.9%, have, in particular, been underutilized in spite of this due to the large $t\bar{t}$ background from which these decays must be distinguished [2]. Attempting to do so is the focus of this thesis; any further unclarified references to both top/anti-top decays and di-Higgs decay refer exclusively to this process.

In this process, the two W-bosons further decay into $l^-\bar{\nu}_l$ and $q\bar{q}'$ where for the lepton either electron or muon are considered [2]. Unfortunately, there is a top/anti-top decay process with an identical end state (Figure 2.1). This $t\bar{t}$ background is thus indistinguishable from the targeted di-Higgs process on the grounds of the identities of their end-state products. As these are the only detectable traces of these processes, any attempt at identifying the di-Higgs process must then take into account the properties thereof, and how these might differ between the two.

[1] The uncertainties given in the super-/subscripts are the scale uncertainties, related to the renormalization and factorization scales. The additional uncertainties given are the $\alpha_s + PDF$ uncertainties based on PDF4LHC15 parton distributions [2].



(a) $t\bar{t} \to (bW^+)(\bar{b}W^-) \to [b(q\bar{q}')][\bar{b}(l^-\bar{\nu}_l)]$ (b) $HH \to (b\bar{b})(W^+W^-) \to (b\bar{b})[(q\bar{q}')(l^-\bar{\nu}_l)]$

Figure 2.1: The $HH \to (b\bar{b})(W^+W^-) \to (b\bar{b})[(q\bar{q}')(l^-\bar{\nu}_l)]$ decay and the relevant $t\bar{t}$ background [8]. As the two processes have identical particles in the end state, they cannot be clearly distinguished based on the identity of these particles alone.

2.3 The Matrix Element Method

This thesis examines whether the matrix element method can be used to that end. The matrix element method takes in a set of experimental events x and generates a weight $P(x|\alpha)$ that compares how well the experimental values align with the expected values that follow from the theoretical framework α. This weight is defined as

$$P(x|\alpha) = \int dy P_\alpha(y) W(x, y) \tag{2.8}$$

where $P_\alpha(y)$ is the parton-level probability of the construction of a particular parton configuration y and $W(x, y)$ is a transfer function that encapsulates the various experimental and detector effects that lead to the reconstruction of a parton-level event y into the experimental event x such as parton shower effects and realistic limits on the detector resolution [5].

Specifically within the context of hadron colliders, the parton-level probability can be represented of a product of the squared matrix element $|M_\alpha|^2$ (where M is the amplitude for the given process to occur [9]) as well as the parton distribution functions $f_1(q_1)$ and $f_2(q_2)$ and the phase-space measure $d\phi(y)$ such that the weight takes on the form

$$P(x|\alpha) = \int d\phi(y) dq_1 dq_2 f_1(q_1) f_2(q_2) |M_\alpha|^2(y) W(x, y) \tag{2.9}$$

This value can be normalized into a probability distribution using the total cross section σ_α:

$$P(x|\alpha) = \frac{1}{\sigma_\alpha} \int d\phi(y) dq_1 dq_2 f_1(q_1) f_2(q_2) |M_\alpha|^2(y) W(x, y) \tag{2.10}$$

A fundamental assumption in the use of the matrix element method is that the transfer function can be broken down into the product of single-particle resolution functions

$$W(x, y) = \prod_{i=1}^{n} W_i(x^i, y^i) \tag{2.11}$$

which, in turn, is typically further simplified by writing the single-object resolution functions as the product of the various resolutions of the quantities measured in the detector, namely the energy, rapidity, and azimuthal angle [5].

$$W_i(x^i, y^i) = W_i^E(x^i, y,^i)W_i^\eta(x^i, y^i)W_i^\phi(x^i, y^i) \tag{2.12}$$

Chapter 3

Methodology

MadWeight is an algorithm aimed at calculating the matrix element method weights. Given the decay chain and the appropriate transfer function, it attempts to then optimize the phase-space integration given in Equation 2.8 using adaptive Monte-Carlo integration, the efficiency of which depends on how the phase-space measure is parameterized. This is also optimized by MadWeight based on the considerations of the system [5]. These processes are performed internally; the user input consists of the chosen decay chain, the kinematic information associated with a particular iteration of this event process, and a selection of the transfer function. Of these, the decay chain is simply read into MadWeight at the beginning of execution, leaving the generation and selection of the event data and the choice of the transfer function as the more involved tasks.

3.1 Event Generation

The event data of the targeted processes are produced by version 6.424 of the PYTHIA event generator, which simulates predominantly hard interactions occurring e^+e^-, ep, and pp colliders [10]. In particular, the following data sets are the main one used:

- Proton-proton collision into a top/anti-top decay at $\sqrt{s} = 13$ TeV, 1 million events
- Proton-proton collision into a di-Higgs decay at $\sqrt{s} = 13$ TeV, 1 million events

In addition, to provide a further look into MadWeight's functionality, four data sets focused on the flanks of top quark masses following the Breit-Wigner distribution are also investigated:

- Proton-proton collision into top/anti-top decay with top and anti-top masses less than 150 GeV at $\sqrt{s} = 14$ TeV, 1 million events
- Proton-proton collision into top/anti-top decay with top and anti-top masses greater than 195 GeV at $\sqrt{s} = 14$ TeV, 1 million events
- Proton-proton collision into top/anti-top decay with top masses less than 150 GeV and anti-top masses greater than 195 GeV at $\sqrt{s} = 14$ TeV, 1 million events
- Proton-proton collision into top/anti-top decay with top masses greater than 195 GeV and anti-top masses less than 150 GeV at $\sqrt{s} = 14$ TeV, 1 million events

In the case of the last two, the program cannot functionally distinguish between the top and anti-top quarks such that, though in all cases one of these has a mass less than 150 GeV and the other has a mass exceeding 195 GeV in every event, it is not split necessarily according to the intended division scheme. There is a slight preference toward it: in the top-150GeV/anti-top-195GeV split, there are more events of this type than there are top-195GeV/anti-top-150GeV events and vice versa.

	top-150GeV /anti-top-195GeV Data Set	top-195GeV /anti-top-150GeV Data Set
Number of top-150GeV /anti-top-195GeV events (% of total events)	520720 (52.07%)	479476 (47.95%)
Number of top-195GeV /anti-top-150GeV events (% of total events)	479280 (47.93%)	520524 (52.05%)

Table 3.1: Comparison of the actual division of top and anti-top masses of each split Breit-Wigner data set with the intended division.

This mixing is addressed by treating these two data sets as a single, mixed data set after the conclusion of the data processing due to the similarity of the data they contain.

As the data from the simulation is read in, the PYTHIA's PYEDIT event study routine, whose purpose is to remove undetectable or unstable objects from the event record, is called with MEDIT=2. This removes empty and documentation lines as well as particles that are either undetectable (neutrinos, unknown particles) or that have fragmented or decayed from being passed on to the jet reconstruction algorithm [10], leaving only detectable, end-state particles to be taken into consideration as in an experimental detector.

3.2 Jet Reconstruction

Following the generation of each event, it becomes necessary to combine the end-state hadrons into jets in an attempt to reconstruct an observable record of the original partons. There are many various jet algorithms that sort and combine these hadrons in different ways with different parameters. The FastJet package contains processes which can carry out a wide selection of these. Defining the jets involves selecting one such algorithm as well as clarifying the value of the distance parameter R, the recombination scheme to be used for combining the momenta, and the particular algorithmic strategy used overall [11].

In order to remain in line with the standard practice of ATLAS, the anti-k_t algorithm of the FastJet library is used as the basis for jet reconstruction. Similarly, the R-parameter is set to 0.4 to align with ATLAS' standard calorimeter jet reconstruction practices [12]. Meanwhile, the E_t recombination scheme is selected; it rescales the initial 3-momenta of the hadrons such that they equal the initial energy, rendering the resulting reconstructed jet as massless when delivered by the algorithm [11]; this is an attempt to mirror the processes of the calculations of the weights in which the particles are also treated as though massless [5]. The actual 3-momenta must be manually reconstructed using the 3-momenta of the individual hadrons contained within the jet separately

$$p_{x,y,z} = \sum_{i=1}^{n} p^i_{x,y,z} \tag{3.1}$$

when necessary.

The clustering of the particles according to the jet definition then follows, and finally the inclusive jets are accessed for analysis, as these are the objects generally referred to as "jets" in hadron-collider algorithms and are in any case recommended for use with the anti-k_t algorithm [11].

3.3 Data Preselection and Processing

3.3.1 Preselection

At this point it becomes necessary to filter out unusable events. The factors deciding whether an event is useful can be grouped into two broad categories: limits imposed by the technical requirements of MadWeights, and those imposed in anticipation of expected real-world physical limitations related to conditions within the ATLAS detector.

In order to ensure that MadWeights can properly process provided inputs, it naturally requires that all identifying particles are present - namely, the anti-bottom and bottom quarks, the two lighter quarks, and the lepton. While providing the neutrino information is necessary to uphold the correct formatting of the input, it has no bearing on the ultimate results of the weight calculation, as MadWeights reconstructs the neutrino internally for its own use [5]. Furthermore, given that ATLAS has very-well developed electron detection and reconstruction systems [13] as well as muon identification [14], it is additionally assumed that the leptons emitted directly from the original decay process can always be properly isolated and identified.

The event selection criteria thus falls to whether or not the reconstructed jets can be successfully identified with the quarks of interest. Thus, an event must have a minimum of four jets with a transverse momentum p_t greater than or equal to 25 GeV - though ATLAS has a technical minimum of 20 GeV [12], for the purposes of data processing the higher value selected here proved more useful (see Chapter 3.4). Additionally, the jets reconstructed at ATLAS are required to have pseudorapidity $\eta < 2.5$ [12], so only jets that met this requirement are taken into consideration for the tagging processes to reflect this. Once a jet is considered identified, it is no longer taken into consideration for further tagging processes.

Similarly to the leptons, ATLAS has done much work in b-tagging, aided by the particular qualities of b-jets that make them distinguishable from other particles [15]. Taking this into consideration, it is assumed that the b-jets can be clearly determined. Thus, the b-tagging procedure used here performs an angular separation (ΔR) comparison between the bottom as well as anti-bottom quarks and the various reconstructed jets for each event; the limit was set to $\Delta R = 0.5$ to keep the difference on a similar scale to size of the jets themselves.

The ΔR-values between each individual jet and both the bottom and anti-bottom quarks were calculated, with only the smaller value of the two being tested against this limit. This selects strictly for the particle closest to a given jet, which should improve the accuracy of the procedure, while also precluding the same jet from being potentially tagged as two different objects. A jet was identified as a b-/anti-b-jet (depending on which quark the tested value corresponded to) if the angular separation was lower than the cutoff; if multiple jets fell within the limit, the jet with the minimum ΔR-value to the b-/anti-b-jet was selected as the identified particle. Though a physical detector would naturally not permit access to the directional knowledge of the original partons, they are used here to enable a method that reflects the accuracy of detector-based b-tagging while remaining simple enough to not distract from the actual goals of this thesis.

The additional condition of an energy difference restriction was considered, but while between 71.15% and 75.76% of events that had jets that passed the ΔR filter for either the bottom or anti-bottom quark also had those jets pass an energy filter of $\Delta E \leq 10\%$ between the actual and reconstructed energy, only between 55.06% and 57.31% of events where both the bottom and anti-bottom quark were identified had both jets then pass through the additional energy filter, as recorded in Appendix A. While loosening the restriction gives more events, it comes at the cost of diminishing the additional accuracy that is the goal of implementing the condition in the first place; this result, and the fact that it is more desirable to have a larger pool of events to work with for the purposes of this research, lead to any such additional checks being dropped entirely.

Light quarks, on the other hand, lack the distinguishing characteristics of bottom quarks and are subsequently difficult to identify concretely. However, it is known that they have originated from a W-boson. Using conservation of energy and momentum, it is possible to construct the mass of the preceding particles in the decay chain:

$$m = \sqrt{(\sum_i E_i) - (\sum_i p_{xi}) - (\sum_i p_{yi}) - (\sum_i p_{zi})} \tag{3.2}$$

If this mass is already known, then the probable products can be deduced by selecting jet combinations which best approximate the mass value. For the top/anti-top decay, this is simply the invariant W-boson mass $m_W = 80.3692$ GeV [1]

$$m_W = \sqrt{(E_1 + E_2) - (p_{x1} + p_{x2}) - (p_{y1} + p_{y2}) - (p_{z1} + p_{z2})} \tag{3.3}$$

with the subscripts "1" and "2" denoting properties of the two jets being used. For the di-Higgs process, however, one of the W-bosons is off-shell, and which W-boson it will be is unknown [2]. Only the invariant Higgs boson mass $m_H = 125.11$ GeV [16] is consistent over all events and can therefore be used for particle identification

$$m_H = \sqrt{(E_1 + E_2 + E_l + E_\nu) - (p_{x1} + p_{x2} + p_{xl} + p_{x\nu}) - (p_{y1} + p_{y2} + p_{yl} + p_{y\nu}) - (p_{z1} + p_{z2} + p_{zl} + p_{z\nu})} \tag{3.4}$$

with the addition of the properties of the lepton (l) and the neutrino (ν) that comprise the second W-boson produced in the decay process.

In both cases, every combination of jets not tagged as a bottom or anti-bottom quark is used to calculate the target mass. For each jet combination, the momentum values are reconstructed according to Equation 3.1; while not strictly necessary for the energy, this is calculated from the particles along the same lines for consistency. The difference between the calculated mass and the expected mass is then taken for each jet pair and the minimum such difference is saved. The jets used in calculating the mass with the minimum difference are identified as with the light quarks.

Since the two events cannot be told apart at the detector level as mentioned in Chapter 2.2, it is also not possible to know which tagging method to employ. The necessity of two separate tagging methodologies for different events thus imposes the need to analyze each data set with both. This doubles the amount of data that ultimately needs to be processed with a hypothesis-matched group, where the correct mass is reconstructed, and a wrong hypothesis group, where the event does not match the mass reconstruction procedure used. In this thesis, every data set is henceforth referred to by both the actual event as well as according to which hypothesis it was treated — for example, a “top-as-top” event refers to a top/anti-top decay event that was analyzed as such, while a “top-as-Higgs” event is a top/anti-top decay event that has been analyzed as though it were a di-Higgs decay event. These are often abbreviated using their starting letter; with respect to the given example, this would be “tt” and “tH,” respectively. The hypothesis selection occurs at the beginning of the execution of the program based on user input.

As a sanity check, for the top-hypothesis process, the mass of the top quark is calculated as well, using the values from the identified light quarks and the bottom quark:

$$m_t = \sqrt{(E_1 + E_2 + E_b) - (p_{x1} + p_{x2} + p_{xb}) - (p_{y1} + p_{y2} + p_{yb}) - (p_{z1} + p_{z2} + p_{zb})} \quad (3.5)$$

The results of the tagging process are given below:

Data Set	Events Remaining After B-Tagging	Events Remaining After Subsequent Light-Quark-Tagging
tt	479409	418354
tH	489804	365610
HH	329777	270754
Ht	318609	265344
tt ($m_t, m_{\bar{t}} < 150$ GeV)	254517	98073
tH ($m_t, m_{\bar{t}} < 150$ GeV)	261862	196393
tt ($m_t, m_{\bar{t}} > 195$ GeV)	594567	396214
tH ($m_t, m_{<}\bar{t} > 195$ GeV)	604585	420987
tt (mixed flank)	419451	220346
tH (mixed flank)	428941	309979

Table 3.2: Results of the b- and light-quark-tagging processes for each data set. Each data set starts out with one million events except for those based on the mixed Breit-Wigner flanks, which are out of two million.

3.3.2 Processing

If an event has all the necessary particles identified, the next step for that event is to extract the energy and the 3-momenta values, which can collectively be used to construct all the information required for MadWeights’ calculations. As it is necessary to calculate the weights associated with the original partons as well as those associated with the reconstructed jets in order to assess how well the weights derived from the reconstructed data reflect the weights of the actual event, this information must accordingly be collected from both sources and then output separately as to permit both calculations.

The required information from the original data is extracted directly from the simulation as the particles in question are read in. PYTHIA’s event record contains not only the particle ID of line I (K(I,2)), but also the status code (K(I,1)) and the parent (K(I,3)) and daughter (K(I,4)-K(I,5)) particles [10]. By cross-referencing the particle ID of the correct particle type with the status of the particle as well as with the IDs of the parent and daughter particles and the status thereof, it is possible to clearly identify particles of interest. The initial processes of the event are recorded in the lines with the status code of K(I,1) =21 [10].

It can be seen in Figure 3.1 that this includes the immediate decay products of the top/anti-top and the di-Higgs decay processes that are the subject of investigation. These particles can be isolated by filtering for the correct status code as well as the corresponding particle IDs; the process used here begins to check for them only after the last listed top quark/Higgs boson in order to omit any quarks which might originate from the original proton collision.

There is an additional check for the number of light quarks, as while which light quarks may appear in a particular event is unknown in general, the order in which the particles appear remains consistent such that the first two light quarks listed after the decaying particles are the quarks of interest. Therefore, limiting the number of quarks read into the particle data listing to two by checking whether one or no quarks have already been read prevents any potential extra light quarks that might arise during the decay process in some events from being accidentally included.

The process for the lepton/neutrino pair is also more involved, as, based on the assumption lined out in Chapter 3.3.1, the same information is used for both the particle data and the jet data listings. As a result, the lepton and neutrino listed in the information about the original event must be used to select the correct end-state particle, whose kinematic information is passed into both lists. By searching the list for the end-state lepton (K(I,1)=1) whose parent is the lepton contained in the initial event listing, both it and the neutrino — which is always the next entry in the list — can be located.

Event listing (standard)

I	particle/jet	K(I,1)	K(I,2)	K(I,3)	K(I,4)	K(I,5)	P(I,1)	P(I,2)	P(I,3)	P(I,4)	P(I,5)
1	!p+!	21	2212	0	3	27	0.00000	0.00000	6499.99993	6500.00000	0.93827
2	!p+!	21	2212	0	4	56	0.00000	0.00000	-6499.99993	6500.00000	0.93827
3	!g!	21	21	1	5	134	0.79074	0.55107	474.62499	474.62596	0.00000
4	!g!	21	21	2	6	139	1.03403	2.13015	-2685.62347	2685.62451	0.00000
5	!g!	21	21	3	0	0	21.42073	52.31799	165.50309	174.89223	0.00000
6	!g!	21	21	4	0	0	16.25476	115.38756	-1250.33265	1255.75086	0.00000
7	!g!	21	21	5	30	165	8.38838	444.96325	-346.90515	564.27512	0.75000
8	!t!	21	6	5	10	11	-43.44330	-117.78146	-326.47686	387.34984	166.41204
9	!tbar!	21	-6	5	12	13	72.73040	-159.47625	-411.44750	479.01809	171.60098
10	!b!	21	5	8	166	167	-25.11475	-93.76915	-263.46628	280.82187	4.80000
11	!W+!	21	24	8	14	15	-18.32855	-24.01231	-63.01058	106.52797	80.40739
12	!bbar!	21	-5	9	103	113	4.87839	-15.45681	11.07844	20.21097	4.80000
13	!W-!	21	-24	9	16	17	67.85201	-144.01945	-422.52594	458.80712	81.42671
14	!sbar!	21	-3	11	168	178	-4.35954	2.08933	-74.55759	74.71583	0.50000
15	!c!	21	4	11	179	179	-13.96901	-26.10164	11.54701	31.81214	1.50000
16	!e-!	21	11	13	20	20	60.15214	-139.74805	-423.45047	449.95344	0.00051
17	!nu_ebar!	21	-12	13	21	21	7.69987	-4.27140	0.92454	8.85368	0.00000
18	(W+)	11	24	11	22	22	-13.70992	-1.46257	-37.79826	89.91194	80.40739
19	(W-)	11	-24	13	23	23	66.34192	-143.65513	-425.54360	461.25449	81.42671
20	e-	1	11	16	0	0	56.76955	-135.36208	-413.47319	438.75478	0.00051
21	nu_ebar	1	-12	17	0	0	7.63332	-4.18486	1.12201	8.77721	0.00000

(a) Sample initial event record for $t\bar{t}$ decay

Event listing (standard)

I	particle/jet	K(I,1)	K(I,2)	K(I,3)	K(I,4)	K(I,5)	P(I,1)	P(I,2)	P(I,3)	P(I,4)	P(I,5)
1	!p+!	21	2212	0	0	0	0.00000	0.00000	6499.99993	6500.00000	0.93827
2	!p+!	21	2212	0	0	0	0.00000	0.00000	-6499.99993	6500.00000	0.93827
3	!g!	21	21	1	0	0	-0.32826	-0.56048	346.53892	346.53953	0.00000
4	!g!	21	21	2	0	0	-0.47486	0.29064	-782.19845	782.19865	0.00000
5	!g!	21	21	3	0	0	-14.17080	-5.11602	308.60526	308.97280	0.00000
6	!g!	21	21	4	0	0	9.69429	21.72006	-149.54360	151.42334	0.00000
7	!h0!	21	25	5	0	0	-172.92492	-34.25726	77.82941	229.68944	124.99294
8	!h0!	21	25	5	0	0	168.44842	50.86130	81.23225	230.70670	125.16054
9	!b!	21	5	7	0	0	-93.42455	-79.43007	59.22108	136.25898	4.70000
10	!bbar!	21	-5	7	0	0	-79.50038	45.17281	18.60833	93.43047	4.70000
11	!W+!	21	24	8	0	0	23.85732	5.67178	-11.66324	35.99069	23.62113
12	!W-!	21	-24	8	0	0	144.59110	45.18952	92.89549	194.71601	79.59945
13	!e+!	21	-11	11	0	0	16.72809	-8.03056	-5.42618	19.33293	0.00000
14	!nu_e!	21	12	11	0	0	7.12923	13.70234	-6.23707	16.65776	0.00000
15	!d!	21	1	12	0	0	6.73047	26.08813	34.52574	43.79402	0.00000
16	!ubar!	21	-2	12	0	0	137.86063	19.10139	58.36975	150.92198	0.00000
17	(h0)	11	25	7	185	198	-172.92492	-34.25726	77.82941	229.68944	124.99294
18	(h0)	11	25	8	19	20	168.44842	50.86130	81.23225	230.70670	125.16054
19	(W+)	11	24	11	21	22	23.85732	5.67178	-11.66324	35.99069	23.62113
20	(W-)	11	-24	12	199	211	144.59110	45.18952	92.89549	194.71601	79.59945
21	e+	1	-11	13	0	0	16.72809	-8.03056	-5.42618	19.33293	0.00000
22	nu_e	1	12	14	0	0	7.12923	13.70234	-6.23707	16.65776	0.00000

(b) Sample initial event record for di-Higgs decay

Figure 3.1: Sample initial event records for $t\bar{t}$ and di-Higgs decay events.

Having identified all six particles, the 3-momenta and energy can be read from the event record; these are stored under P(I,1), P(I,2), P(I,3) and P(I,4), these being the x-, y-, and z-momenta as well as the energy, respectively [10]. These values are stored in an instance of ROOT's TLorentzVector object [17], one vector for each particle. These vectors allow these values to be readily accessible throughout the program where necessary, including outputting the data.

The information sourced from the jets, on the other hand, is read in following the successful identification of all particles. In practice, this is just for the quarks, as the lepton and neutrino have already been sourced as outlined above. The energy and 3-momenta of the identified jets are stored in new Lorentz vector objects in the same manner as for the original particles.

In both cases, this information is collected in a string listing out the particles with their Pythia particle identification, their x-, y-, and z-momenta, and their total energy. These strings are output as a single line per event into two separate data files, one for the original information and another for the reconstructed data. The outputs for data derived from the two split data sets, top-150GeV/anti-top-195GeV and top-195GeV/anti-top-150GeV, are merged into single data files so that they can be analyzed as a single data set as outlined in Chapter 3.1. These files are formatted into the LHCO format required by MadWeights as input using a GAWK program.

3.4 MadWeight and Weight Calculation

With the data gathered, all that remains before the execution of MadWeight is the selection of the transfer function. MadWeight breaks these up along the lines of Equations 2.11 and 2.12; furthermore, the default setting for both W_i^η and W_i^ϕ is simply a delta function. As a transfer function, this represents an ideal detector that measures these values exactly as they are [5]. For the weights of the original partons, a delta function for W_i^E is sufficient as well, as the energy values are the actual values and do not represent observables. To compare the weights calculated from the reconstructed jets to the weights of the particles directly, they are also originally treated with a delta function for W_i^E; this also gives a sense of how much of a correction an appropriate transfer function needs to provide.

As for a more realistic approach to the transfer function for the weights derived from the jets, MadWeight comes with several prepared options, as well as the ability to construct one from scratch if need be. For a first glance at MadWeights' results, the simplest of the already accompanying transfer functions, a single Gaussian smearing, was selected. It is defined as

$$W = \frac{1}{\sqrt{2\pi} * loc_2} e^{\frac{-(p_0 - p_{exp0} - loc_1)^2}{2 * loc_2^2}} \tag{3.6}$$

where p_0 is the partonic event energy, p_{exp0} is the experimental event energy, and loc_1 and loc_2 are local variables that need to be input. Based on the structure of the transfer function, they are analogous to the mean and standard deviation of a Gaussian distribution, where the value under consideration is the difference between the original and reconstructed energies. Thus, the definition of these variables, which are by default

$$loc_1 = var_1 + var_2\sqrt{p_0} + var_3 * p_0 \tag{3.7}$$

$$loc_2 = var_4 + var_5\sqrt{p_0} + var_6 * p_0 \tag{3.8}$$

is achieved through parameterizing the mean and standard error of these energy differences in terms of the square root of the parton energy. In order to do so, these two values were plotted against each other. It should be noted that this transfer function is also applied to the leptons, but as the values for these are treated as definitely known, the variables for these were left such that the mean was simply set at 0 with a constant width of 1.

An issue became immediately apparent when looking at these distributions of energy differences, visible on the left in Figures 3.2 and 3.3 as well as Appendix B. The expected even distribution about a difference of 0 was not present; indeed, there appeared a hard cutoff for the positive differences, where the parton energy exceeds the jet energy. This occurs as any given parton energy value (x^2) represents a limit on the maximum positive value the energy difference can assume which is then reduced by the minimum value of the jet energy, which is equivalent to the minimum p_t value of 25 GeV. Thus, the positive energy differences are limited to $\Delta E < x^2 - 25$. The negative values do not have such a boundary; for a given parton energy, the corresponding jet value can be arbitrarily larger.

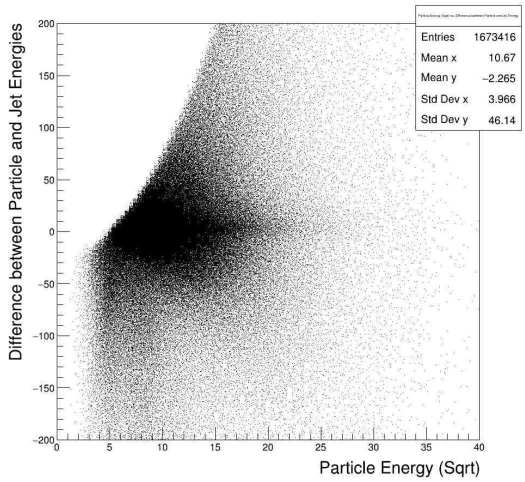
In order to deal with this issue, it was decided to clean up this boundary by imposing a matching manual limit on the maximum of plotted energy differences and mirroring this to the negative differences, $y \geq -(x^2 - 25) = 25 - x^2$, which then results in a distribution which can be modeled with the single-Gaussian approach. After $x = 10$, this limit was fixed to the corresponding value +/- 75, as after this point the natural limit widens as the parton energy values increase, permitting increasing high differences in both the positive and negative directions. This constant limit instead serves to limit any distorting impacts of particles with extremely high energy differences to the tagged jets on the fitting process. The important resulting distributions can be seen in the right column of Figures 3.2 and 3.3, with noteworthy further examples in Appendix B.

The imposition of these limits generally removes about 20% of all particle-jet pairs in a data set, with notable exceptions being the tt and tH Breit-Wigner mixed-flank data sets, which see slightly over 30% of the pairs removed. While these removed results are unlikely to be treated properly by the transfer function, the vast majority of data points do fall within this range such that the end results of the weights calculation are expected to be largely representative of the overall behavior.

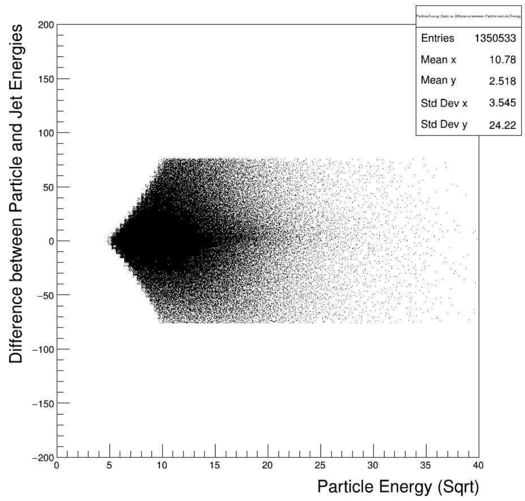
This resolves the most glaring problem; with the imposition of these limits, the distribution appears roughly symmetrical about 0 across all data sets. The structure is, in fact, otherwise almost identical between different data sets and hypotheses, irrespective of the particle type or the mass distribution of the top quark. One slight difference that appears is that this clustering about 0 remains more prominent at higher particle energies in the data sets based on the right- and mixed flanks of the Breit-Wigner-distributed top mass (Figures B.1 and B.2), which follows naturally from the higher energies associated with the larger top quark masses.

What remains is to extract the necessary information by passing the remaining data into profile histograms with a bin width of 1 $\sqrt{GeV}$. This provides the mean of each bin as well as the error, which was set to use the standard deviation to align with the values needed for the transfer function. The error values of each bin can then also be extracted and plotted onto a separate profile histogram.

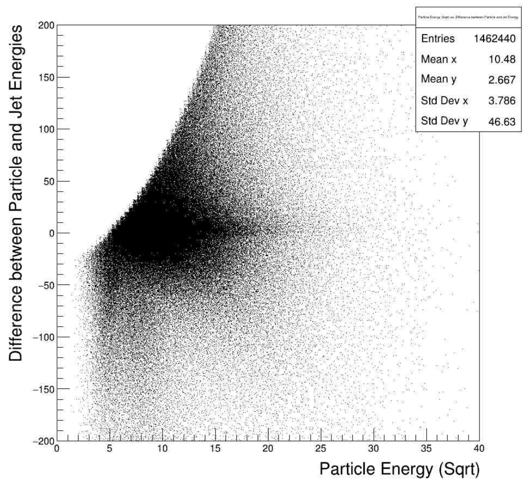
The parametrization of Equation 3.7 then follows through using ROOT's native Fit Panel [18] function on the profile histogram of the energy differences, which uses the recorded means for each bin. The predefined second-degree polynomial — which, implemented on the square root of the particle energy, recreates the exact structure — is fitted over a range of 5 $\sqrt{GeV}$ to 20 $\sqrt{GeV}$. This cutoff reflects what can be seen in the energy difference distributions: the clear clustering of differences about 0 disappears nearly entirely such that the distribution becomes highly scattered. This most likely follows from the scarcity of relevant data beyond this point, which is clearly visible in the original energy difference distributions. For all data



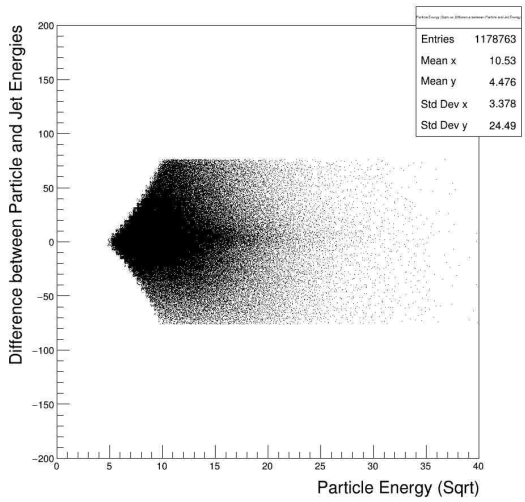
(a) tt, Untrimmed



(b) tt, Trimmed



(c) tH, Untrimmed

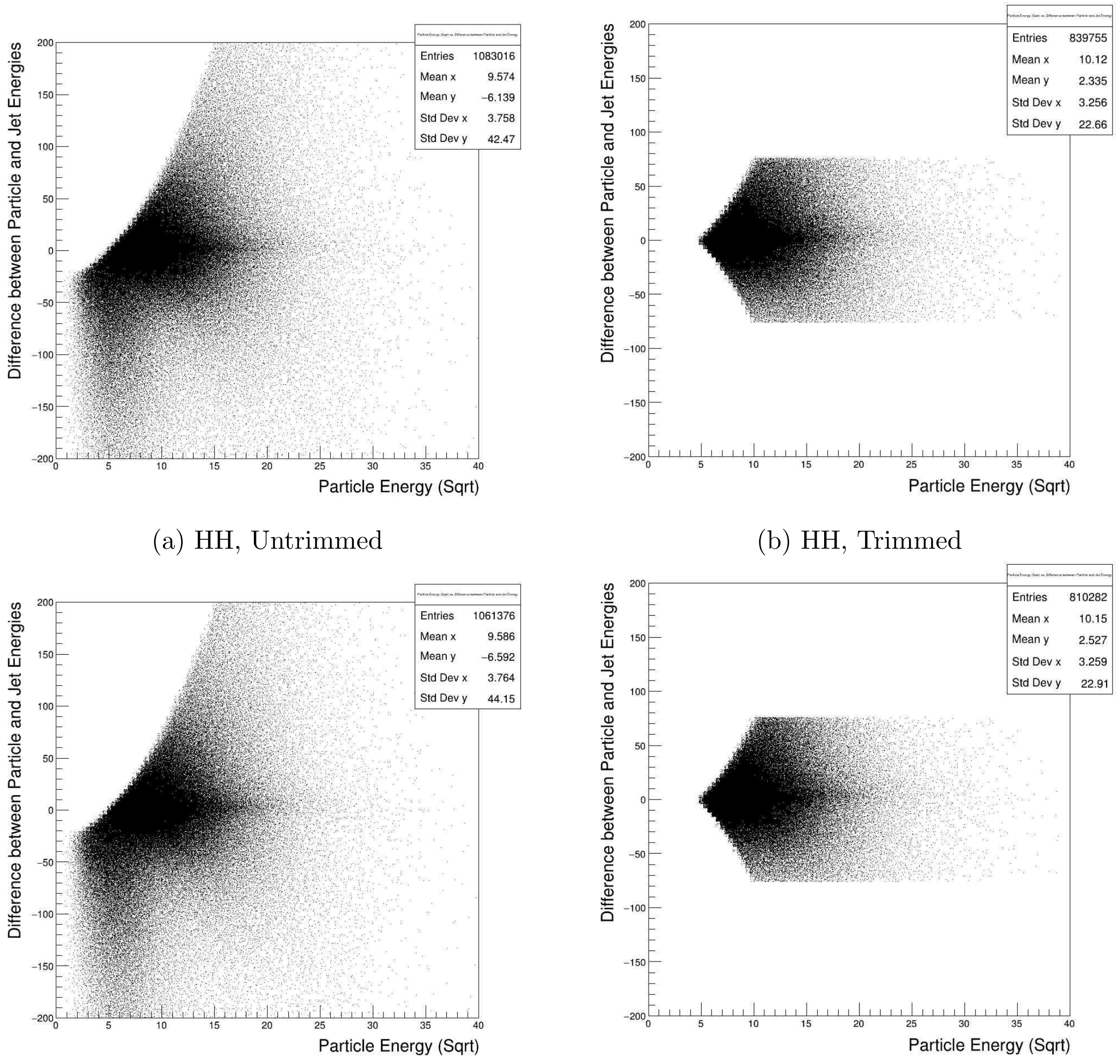


(d) tH, Trimmed

Figure 3.2: Energy difference between the original particles and their associated jet compared to the square root of the particle energy. The left column shows the full distribution, while the distributions shown on the right have had a limit $|\Delta E| \leq x^2 - 25$ (fixed at $|\Delta E| \leq 75$ at and after $x = 10$) imposed on both positive and negative differences that reflects the organic upper limit and mirrors it to the negative differences. The data set used for each plot is given underneath.

sets, less than 10% of partons have energies whose root exceeds 20 $\sqrt{GeV}$; for all data sets except for the tt and tH right flank of the Breit-Wigner top quark mass ($m_t, m_{\bar{t}} > 195$ GeV) and the tt mixed flanks, this value was lower than 5%. As a result, information in this range cannot be considered to be as reliable or meaningful as the more well-represented lower energies, justifying its exclusion from the parameterization. This parameterization procedure also motivates the elevation of the minimum p_t from the ATLAS standard value of 20 GeV to 25 GeV, as this makes the minimum possible energy a perfect square, ensuring that the data for the jet energies begins exactly at 5 $\sqrt{GeV}$ and not between two integers.

This same process is then repeated for the profile histogram of the errors for the parameterization of Equation 3.8 across the same range such that only the errors from the relevant bins are utilized. The main results for both parametrizations are shown in Figures 3.4 and 3.5, with additional examples in Appendix C.



(a) HH, Untrimmed

(b) HH, Trimmed

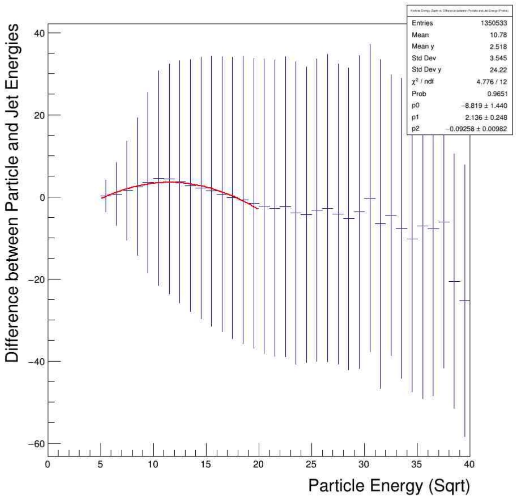
(c) Ht, Untrimmed

(d) Ht, Trimmed

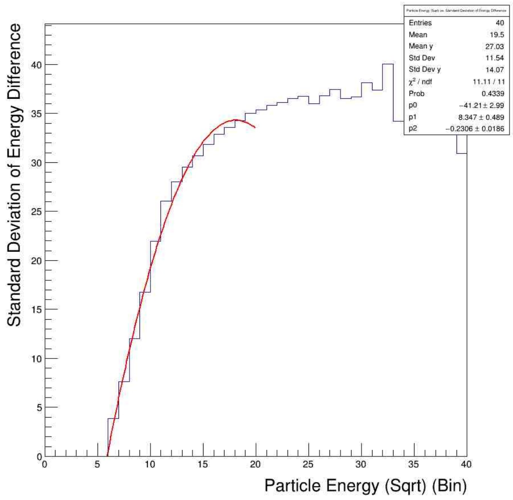
Figure 3.3: Energy difference between the original particles and their associated jet compared to the square root of the particle energy, continued. The left column shows the full distribution, while the distributions shown on the right have had a limit $|\Delta E| \leq x^2 - 25$ (fixed at $|\Delta E| \leq 75$ at and after $x = 10$) imposed on both positive and negative differences that reflects the organic upper limit and mirrors it to the negative differences. The data set used for each plot is given underneath.

The profile histograms and their standard deviation distributions confirm the main observation made of the original energy difference distributions, namely that there are no major discernible differences in the energy difference distributions between different data sets. The increased rate of high-energy particles ($\sqrt{E} \geq 20\sqrt{GeV}$) noted above in data sets involving high-mass top quarks is mirrored by a greater stability of the means given by the profile histogram in this energy range (Figures C.1 and C.2). Due to the exclusion of these data points, however, it has no bearing on the fit, which sees only small numerical shifts across all data sets. All parameterizations of the means have the same shape consisting of a fairly flat downward-opening parabola with a maximum roughly between 10 and 15 $\sqrt{GeV}$, while those of the standard deviations increase rapidly until reaching a maximum slightly before the cutoff at 20 $\sqrt{GeV}$.

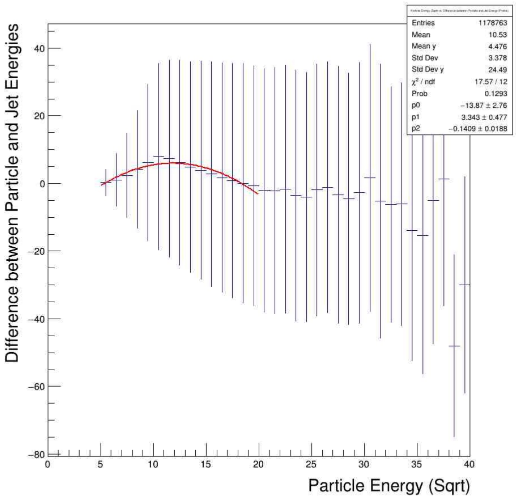
With both parameterizations, and thus all necessary variables of the transfer function on



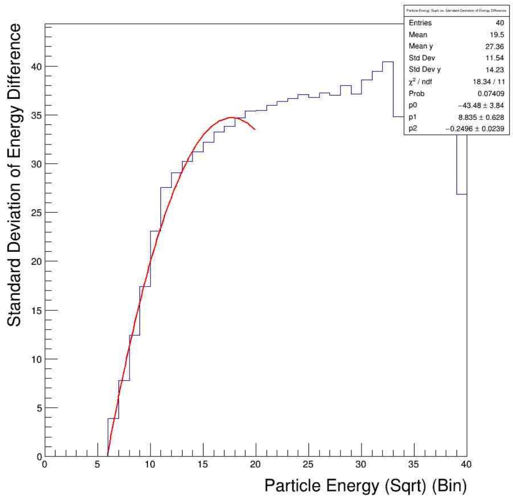
(a) tt



(b) tt



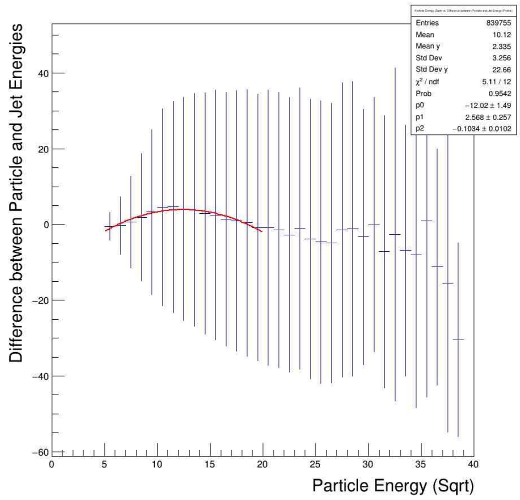
(c) tH



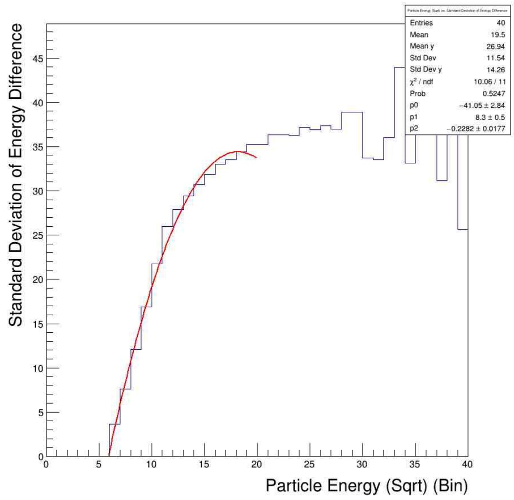
(d) tH

Figure 3.4: Profile histograms of the (trimmed) energy differences between original particles and the associated jet (left) and of the standard deviation of each bin of the former (right) compared to the square root of the particle energy. The second-degree-polynomial parametrization over a range of $\sqrt{E} = 5$ to $\sqrt{E} = 20$ for each plot is also shown. The data set used for each plot is given underneath.

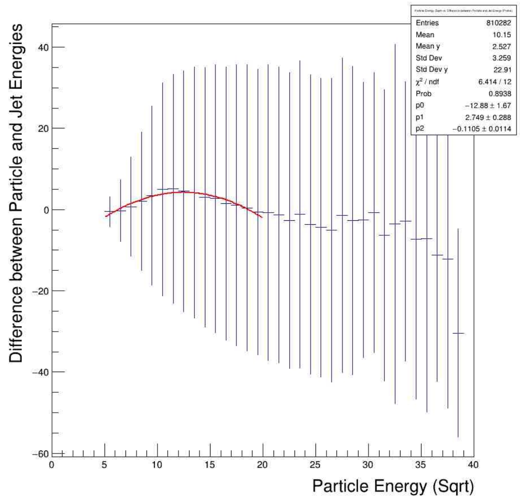
hand, these can be passed onto MadWeight during execution and all of the desired weights can be calculated.



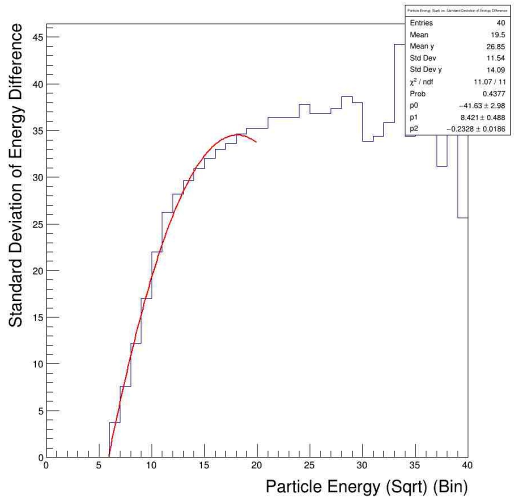
(a) HH



(b) HH



(c) Ht



(d) Ht

Figure 3.5: Profile histograms of the (trimmed) energy differences between original particles and the associated jet (left) and of the standard deviation of each bin of the former (right) compared to the square root of the particle energy, continued. The second-degree-polynomial parametrization over a range of $\sqrt{E} = 5$ to $\sqrt{E} = 20$ for each plot is also shown. The data set used for each plot is given underneath.

Chapter 4

Results and Preliminary Conclusions

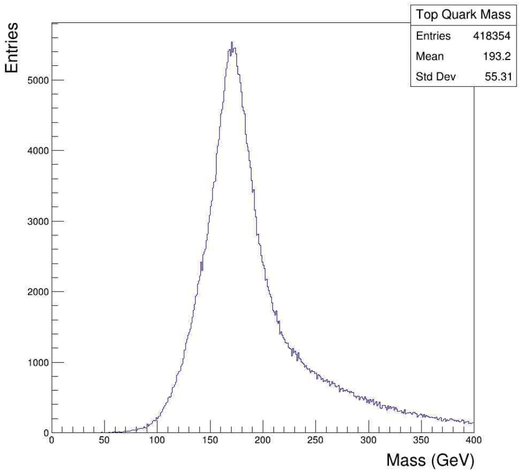
4.1 Event Reconstruction and Subsequent Weights

The tagging procedure for the light quarks was, at first glance, a success. For the hypothesis-matching procedure, the distribution of the reconstructed masses of the original particles demonstrates reasonable behavior, with a peak near the expected mass of each particle and a wider distribution for the mass of the reconstructed top quark than for the Higgs boson, as is visible in Figure 4.1.

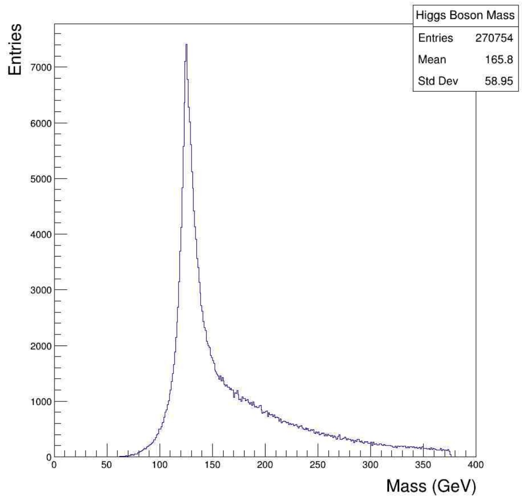
For the wrong-hypothesis procedure, on the other hand, the mass distributions are very broad and not centered on the otherwise expected mass of the particle, which can be seen in Figure 4.2. These broad distributions indicate very clearly the mismatch between the data available from the jets and the incorrectly-hypothesized reconstruction.

Since no issues were immediately apparent with respect to the identification of the necessary jets, the values of the calculated weights were originally taken at face-value. In order to get a sense of the reliability of the weights associated with the reconstructed events, the negative base-10 logarithm of those weights calculated using the delta function as the transfer function were plotted against the negative logarithm of the weights calculated using the original events; the use of negative logarithms here was necessitated by the small value of the weights themselves. Due to the time-intensive nature of the weight calculation, only about a third of all valid events were ultimately plotted. With the weights of the original events representing the best estimate of the actual of the weights of each event, a near 1:1 correspondence between the two sets of weights would indicate that the jet reconstructions captured the particles well and that the weights calculated directly from those results are thus just as representative.

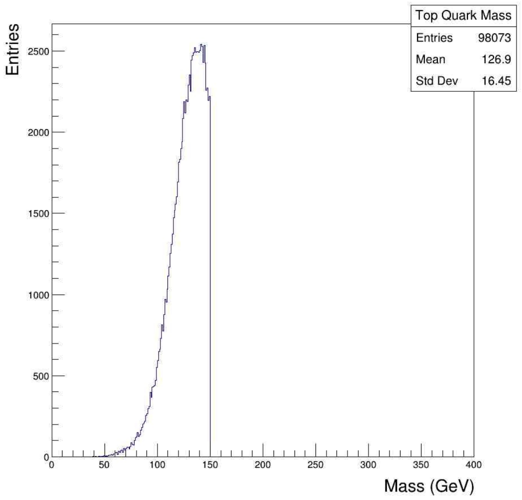
The results, shown in Figures 4.3 and 4.4, with selected quantitative data in Table 4.1, are somewhat surprising. The plotted weights of the original events demonstrate the expected behavior on their own: those associated with the (non-Breit-Wigner) correct hypothesis generally have the largest weights (reflected in the smallest negative logarithm thereof) with a mean weight of $10^{-24.05}$ for tt events and $10^{-27.1}$ for HH events, while the wrong-hypothesis weights are significantly smaller at $10^{-26.72}$ for Ht events and $10^{-35.49}$ for tH events. However, the weights of the reconstructed events are several orders of magnitude smaller for non-Breit-Wigner correct-hypothesis events (nearly 4 for tt events and about 8.5 for HH), while for all other event types the two weights are within a two orders of magnitude of each other — aside from tH events, even within a single order of magnitude. This was admittedly



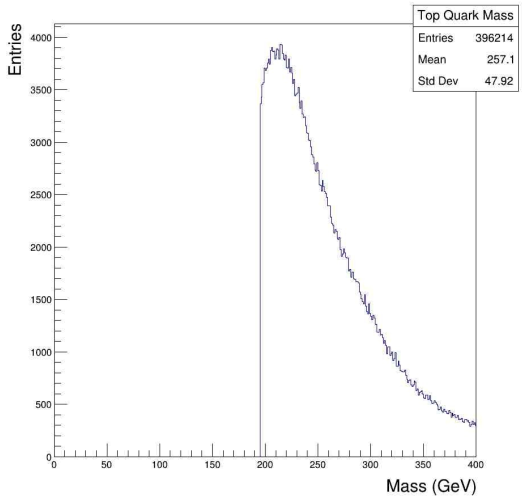
(a) Top Quark Mass, tt



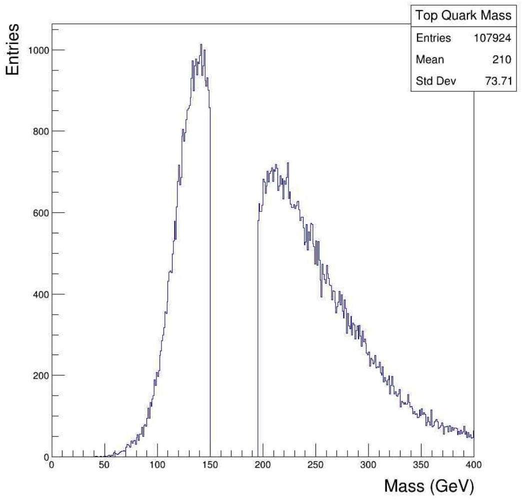
(b) Higgs Boson Mass, HH



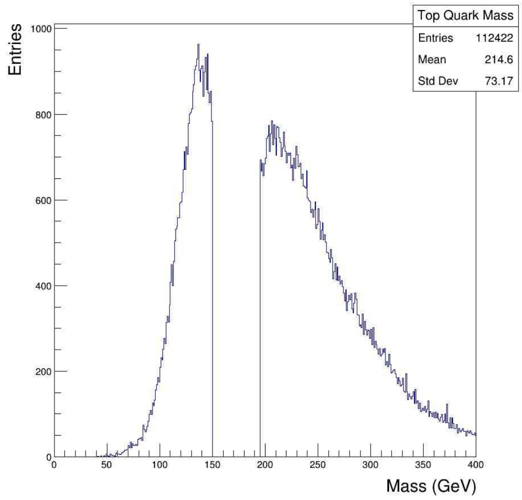
(c) Top Quark Mass, tt ($m_t, m_{\bar{t}} < 150$ GeV)



(d) Top Quark Mass, tt ($m_t, m_{\bar{t}} > 195$ GeV)



(e) Top Quark Mass, tt ($m_t < 150, m_{\bar{t}} > 195$ GeV)



(f) Top Quark Mass, tt ($m_t > 195, m_{\bar{t}} < 150$ GeV)

Figure 4.1: Mass distributions (in GeV) of the reconstructed top quark/Higgs boson using the reconstruction method of the correct hypothesis. The data set used for each plot is given underneath.

the inverse of the expected relationship where the correct-hypothesis reconstructed event weights, being based on the correct physics, would be closer to the weights of the original event while the wrong-hypothesis reconstructions might be expected to correlate less. Most worryingly, this places the weights of the correctly reconstructed events well within range of the weights of the incorrectly reconstructed events; for example, the average weight of the reconstructed tt events is only $10^{-27.98}$, while the average weight of reconstructed Ht events is even slightly larger at $10^{-27.24}$. Similarly, for the di-Higgs hypothesis, the correctly reconstructed HH events had a mean weight of $10^{-35.66}$ while the same for tH events was $10^{-36.9}$. Weights derived from data based on the Breit-Wigner-mass top quarks function similarly to the wrong-hypothesis results, even when the correct hypothesis for the data is used; the comparisons of the weights of the original events with those of the reconstructed events for these data sets are given in Appendix D.

This dissonance, the strength of the discrepancies, and the resulting loss of distiguishability certainly began to raise questions. Nevertheless, it was believed that the single-Gaussian transfer function might yet correct the weights of the reconstructed jets of the correct-hypothesis data to a more reasonable difference from the original particle weights. To compare, the negative logarithms of the reconstructed event weights, now calculated using the single-Gaussian transfer function, were again plotted against the negative logarithm of the original event weights.

The different choice of transfer function gave mixed results at first glance. While the gap between the mean weights of the reconstructed events and those of the original events apparently closes by about 2.14 orders of magnitude for HH events, it increases by 0.27 orders of magnitude for tt events. Indeed, the two hypotheses react differently to the change overall; data sets analyzed using the top-quark hypothesis mostly see the mean weights increase, while the mean weights decrease for data analyzed with the Higgs-hypothesis. Looking at the plots themselves, even this may be misleading: for the top-hypothesis (for example in Figure 4.3), it can be seen how the distribution of the negative logarithm of the weights scatters upwards after introducing the single-Gaussian transfer function — indicating a decrease in the weights — while for events analyzed with the Higg-hypothesis they simply appear to thin out. As the number of plotted events also decreases, it seems more likely that the same upward scattering is occurring with the Higgs-hypothesis, only that it removes them from the bounds of the plot (which were kept fixed to provide a better direct comparison across different data sets), which would imply that the observed decreases in the mean weights are reflecting this omission instead of the actual behavior. The reasonable conclusion is thus that the implementation of the single-Gaussian transfer function led to a further decrease in weights across the board while also leaving the issue of distinguishing between different events under the same hypothesis untouched.

Based on all of the observed behavior of the weights, some of it can be explained. The weights of original events of the two correct-hypothesis, non-Breit-Wigner data sets certainly represent the best estimate of the actual value. Meanwhile, the much lower weights of the original events for the wrong-hypothesis and all of the Breit-Wigner top mass data sets reflects that these do not match onto the theoretical expectations as well. That the weights of the reconstructed events for these match so well when calculated using the delta function could simply be a result of the fact that the original events are already so inaccurate to the hypothesized events that discrepancies between the original and reconstructed events contribute less to the calculation of the weights than the use of the incorrect model. The questions that remained, then, were why the weights of reconstructed events for the non-Breit-Wigner correct hypotheses should vary so strongly from those of the original events that they nearly resemble incorrect weights in spite of the apparent success of reconstruction and tagging and why the chosen transfer function failed to improved the situation.

Data Set	Average -log(Weight) of Original Events (X)	Average - log(Weight) of Reconstructed Event (Y) (Delta)	Average -log(Weight) of Reconstructed Event (Y') (Single-Gaussian)	X-Y	X-Y'	Y-Y'
tt	24.05	27.98	28.25	-3.93	-4.2	-0.27
tt (150)	27.18	27.38	26.98	-0.2	0.2	0.4
tt (195)	28.84	29.32	29.75	-0.48	-0.91	-0.43
tt (Mix)	27.96	28.39	28.55	-0.43	-0.59	-0.16
Ht	26.72	27.24	27.43	-0.52	-0.71	-0.19
HH	27.1	35.66	33.52	-8.56	-6.42	2.14
tH	35.49	36.9	34.76	-1.41	1.09	2.14
tH (150)	36.11	36.72	34.4	-0.61	1.06	2.32
tH (195)	36.83	37.23	35.05	-0.4	1.78	2.18
tH (Mix)	36.47	37	34.96	-0.53	1.51	2.04

Table 4.1: Comparison of the mean negative base-10 logarithms of the plotted weights associated with the original and reconstructed events, the latter calculated with both the delta function and the single-Gaussian function as transfer functions for the examined data sets. For the data sets featuring top masses along the flanks of the Breit-Wigner curve, the mass cutoffs are given in parentheses after.

4.2 Investigating the Energy Distributions

Concurrent investigations [19] proceeded to show that MadWeight demonstrates a very high sensitivity to even slight discrepancies; a minor difference the mass of the W-boson mass used, 80.4 GeV to 79.8 GeV, results in a difference of up to a factor of 10^3. This naturally spurred discussion concerning what model of transfer function should be adopted in order to make the necessary precise corrections. In particular, how the transfer function should correct for particles with large energy differences came to the forefront. Indeed, even though many highly divergent particles were removed from the analysis as described in Chapter 3.4, the new insights into the particularity of MadWeights broadened the scope of what differences needed to be considered as "large" immensely. Ultimately, this line of inquiry raised the question of why such strong corrections should be necessary in the first place and thus of what exactly was causing these energy differences. These had heretofore been assumed to be statistical phenomena, and further that these would be limited in both frequency and impact. In light of the results of the calculation of the weights and MadWeights' sensitivity, it was evident that these differences would need to be investigated to bring some clarity to the situation. The full results of these investigations can be seen in Appendices E and F; the averages given below are taken over the percentages of valid events fulfilling the stated diagnostic criteria within each hypothesis, broken down by the number of jets in the event:

$$\bar{\%} = \frac{\sum_j^{n-jets=6} \sum_i^{hypothesis=10} \%_{ij}}{60} \tag{4.1}$$

as there are 6 blocks of numbers of jets (4, 5, 6, 7, 8 or 9+ jets in an event) and 10 hypotheses, 5 matching the data and 5 incorrect.

Two problems that had been disguised by the apparent success of the tagging procedure came to light. The first is that, contrary to initial assumptions, energy-distorting effects were clearly occurring fairly frequently. This is exemplified by the behavior of the energies of the bottom/anti-bottom quarks, which are in general more likely to be correctly identified, as the only way the wrong jet would be selected is if there are multiple jets in close vicinity to each other in roughly the same direction as the quark in question. In spite of this, the average percentage of valid events featuring a bottom quark whose energy differed from the identified jet by more than 10% rests at about 55.78%, while the same figure for the anti-bottom quarks is at 56.48%. In an average of 31.79% of selected events, both the bottom and the anti-bottom quarks have energies that differ more than 10% from the tagged jets.

These values vary only slightly between events with different jet amounts; while, for example, the average percentage of events with four jets with both bottom and anti-bottom quark energies diverging more than 10% from the jet energies is only 29.06%, this figure rises to 35.23% for events with nine or more jets. The analogous results for the bottom and anti-bottom quarks separately see a rise in the average from 53.34% to 58.37% and from 54.00% to 59.69%, respectively. This is best explained by the increased likelihood of false positives as the number of jets increases. The differences between hypotheses, for events with the same number of jets, are generally even smaller.

This problem is, unsurprisingly, even worse with respect to the light quarks. On average, 40.14% of valid events have just one light quark whose energy differs in excess of 10% from the tagged jet, and for another 45.15% this was strictly the case for both. As these values were recorded disjointly — for the light quarks, events where both quarks matched this criteria were removed from the single-quark count — it comes out to a stunning average of 85.31% of events where at least one light quark's energy differs from that of its supposed jet by more than 10%. Even higher energy differences are no rarity, either; whereas only an average of 7.618% and 8.276% of events feature a bottom quark and anti-bottom quark, respectively, whose energies differ more than 50% from the corresponding tagged jets, an average of 26.87% of events feature exactly one light quark where this was the case. Here, too, a dependence on the number of jets in the event can be seen, and it is much stronger. The average percentage of events with at least one light quark with a difference of 10% to the tagged jet ranges from 80.50% for events with four jets to 90.01% for events with nine or more jets. The equivalent figures are 24.40% and 45.17% at a 50% difference.

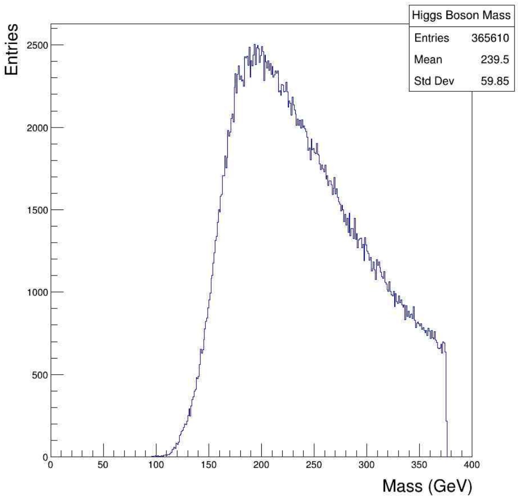
Given the shifts caused by a 0.6 GeV change in the target W-boson mass, differences at the scale and frequency seen here are almost certainly playing a role in skewing the weights of the jets away from the weights of the particles. There are various probable causes, all of which are simultaneously playing a role in establishing these energy differences between particles and jets. Most of these are indeed statistical in their nature, reflected in the stability of the percentages associated with the more-accurately tagged b-quarks. Among the likely contributors is gluon radiation, which PYTHIA models and would result in changes to energy as well as momentum during the decay process [10]. This was expected on some level, but it was assumed that these would largely be filtered out at the tagging level due to the resulting directional changes, but the large number of divergent-energy events indicates that this assumption may have been misplaced and that there are enough instances where the emitted gluon may not be radically interfering with the momentum of the daughter particles while still changing the energy.

Another statistical effect that is likely warping the energies of the jets concerns inevitable shortcomings with respect to how the particles are sorted into jets. The jet algorithm cannot discern origin, after all; if two particles have similar momenta or if their decay products otherwise overlap, some of these will be sorted into jets whose contents are otherwise largely the daughter particles of a different parton. Some jets will thus lose energy relative to the particle they are supposed to represent while others gain energy. It also occurs that products of the original proton collision are thrown away from the beam into the range of the calorimeter; these particles will also be sorted into jets and will account for another net gain of energy.

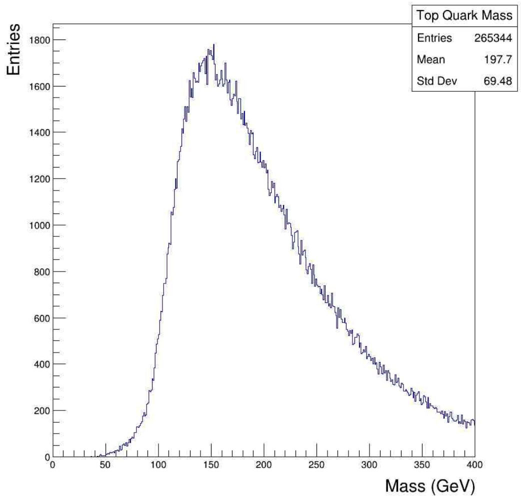
These effects are undoubtedly also present with the light quarks. However, the rate of failure is much higher than for the bottom quarks, which indicates that in addition to the statistical issues, there is a second, methodological problem. Since the only difference between the two quark categories is in the tagging procedure, it was worth investigating whether it held up as well as it appeared to at first glance. Knowing that the b-tagging procedure was selected to represent a higher degree of accuracy, it was decided to take the deciding criterion — the angular separation between the jet and the particle — and use it as a benchmark for the performance of the light quark tagging.

Whereas all bottom/anti-bottom quarks are required to fall within the $\Delta R = 0.5$ range, an average percentage of 42.05% of events feature one light quark with a greater angular separation to the tagged jet associated with it, and for an average percentage of 24.11% of events this is strictly the case for both light quarks. As, again, these values were counted separately, a total average percentage of 66.15% of events have at least one light quark with an angular separation in excess of R=0.5 with respect to its tagged jet. This phenomenon, as might be expected, was heavily dependent on the number of jets in an event, with a rise from an average percentage of 48.15% of events with four jets with at least one such light quark to 80.55% for events with nine or more jets. Much of this discrepancy occurs in events with strictly two light quarks with a high angular separation from their respective jets — the corresponding values are 5.089% for four-jet events and 35.57% for events with nine jets or more.

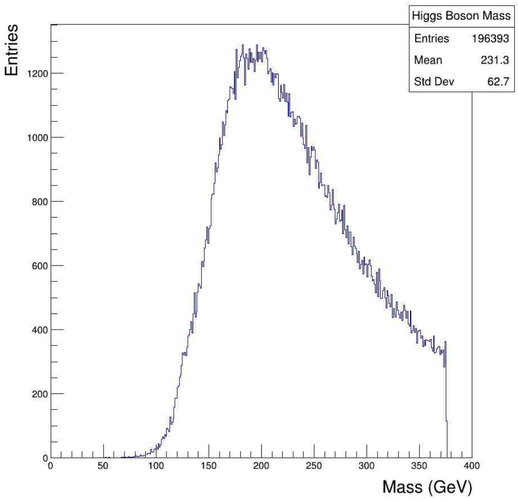
While it cannot be ruled out that, due to radiative effects, the light quark decay products have a noticeably different momentum and thus direction from the original parton and that the associated jet is actually successfully identified, in light of the information regarding the energy discrepancies of the light quarks and their jets as well as the fact that the fraction of events with high angular separations between the two increases with the number of jets present, it seems more reasonable to assume that if this does happen, it is largely accidental. Instead, the results indicate that the light quark jets are likely being misidentified, and of course the chance that this occurs rises with the number of jets which are present. Not only is this another source of error for the energies which explains the heightened rate at which the light quarks have diverging energies from the tagged jets, but it causes problems for MadWeight in and of itself. The difference of the jet/experimental momenta for the light quarks from the parton/theoretical values can also undoubtedly contribute to the lowering the weights for the former and makes treating them without correction nonviable.



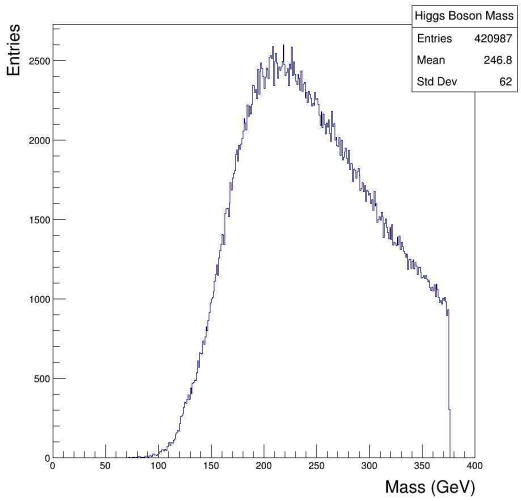
(a) Higgs Boson Mass, tH



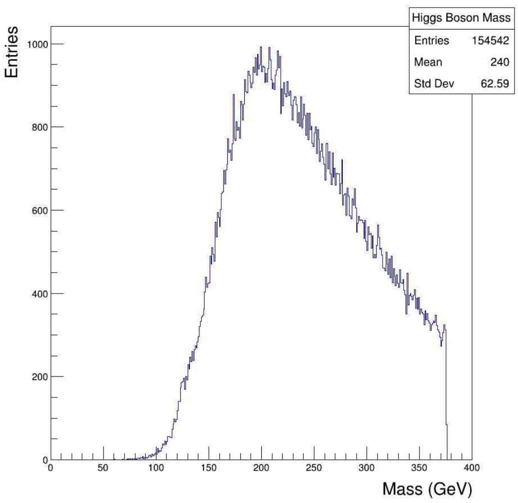
(b) Top Quark Mass, Ht



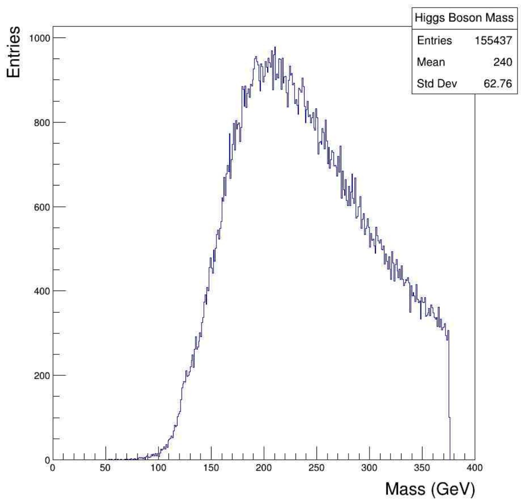
(c) Higgs Boson Mass, tH ($m_t, m_{\bar{t}} < 150$ GeV)



(d) Higgs Boson Mass, tH ($m_t, m_{\bar{t}} > 195$ GeV)

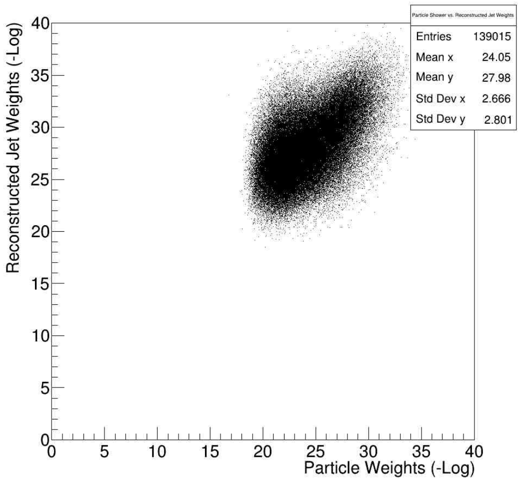


(e) Higgs Boson Mass, tH ($m_t < 150, m_{\bar{t}} > 195$ GeV)

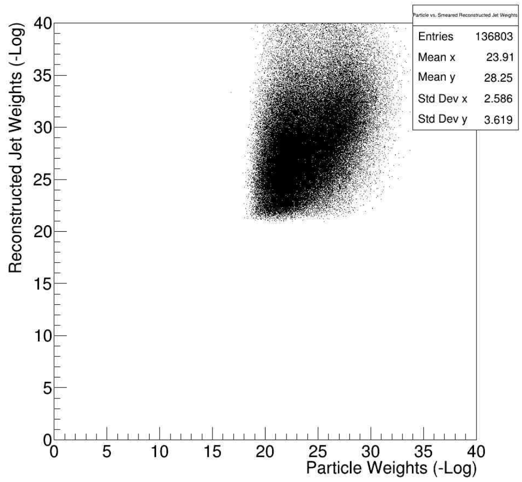


(f) Higgs Boson Mass, tt ($m_t > 195, m_{\bar{t}} < 150$ GeV)

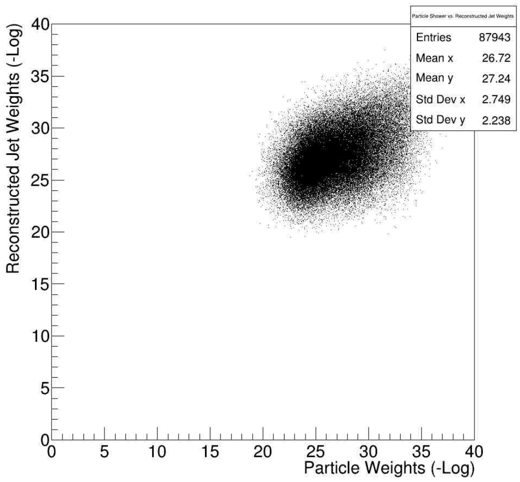
Figure 4.2: Mass distributions (in GeV) of the reconstructed top quark/Higgs boson using the reconstruction method of the wrong hypothesis. The data set used for each plot is given underneath.



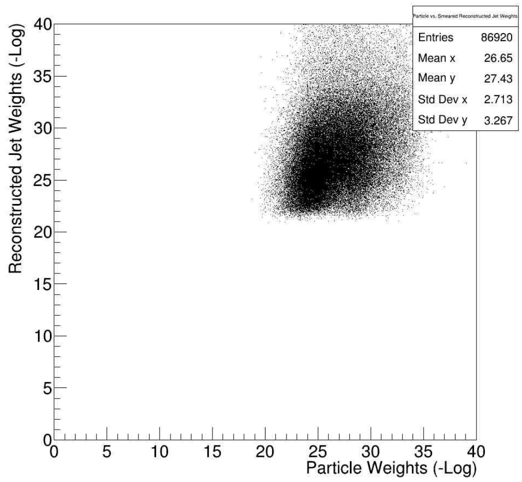
(a) tt, Original vs Reconstructed (Delta)



(b) tt, Original vs Reconstructed (Single-Gaussian)

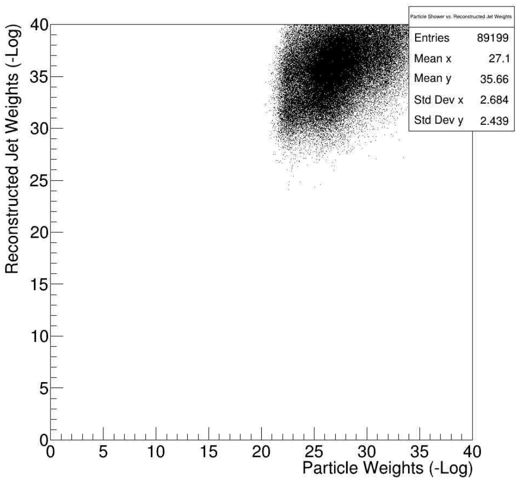


(c) Ht, Original vs Reconstructed (Delta)

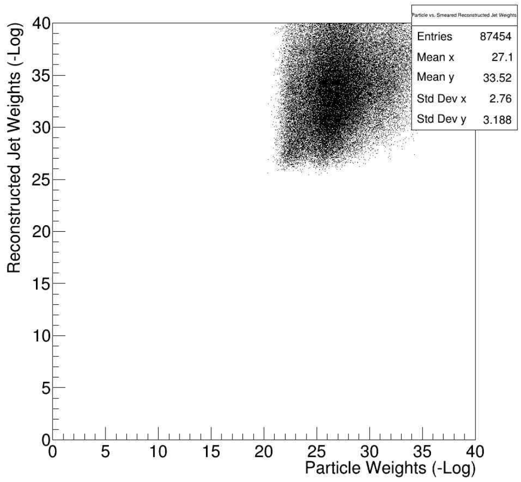


(d) Ht, Original vs Reconstructed (Single-Gaussian)

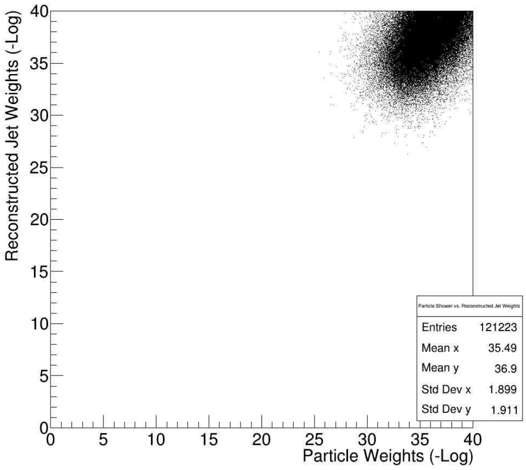
Figure 4.3: Plot of the negative base-10 logarithms of the weights of the reconstructed events against those of the original events. The left column uses the weights of the reconstructed events where the delta function was used as the transfer function, while the right column uses a single-Gaussian distribution as the transfer function. The data set used for each plot is given underneath.



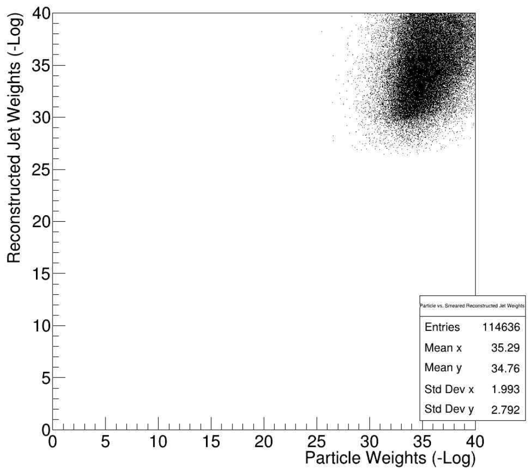
(a) HH, Original vs Reconstructed (Delta)



(b) HH, Original vs Reconstructed (Single-Gaussian)



(c) tH, Original vs Reconstructed (Delta)



(d) tH, Original vs Reconstructed (Single-Gaussian)

Figure 4.4: Plot of the negative base-10 logarithms of the weights of the reconstructed events against those of the original events, continued. The left column uses the weights of the reconstructed events where the delta function was used as the transfer function, while the right column uses a single-Gaussian distribution as the transfer function. The data set used for each plot is given underneath.

Chapter 5

Methodology II

5.1 Justifying an Alternate Approach

The previous results, taken together, indicate that a significantly more involved approach with respect to the chosen transfer function is needed if the matrix element method is to deliver more workable results via MadWeight. Noting that there is precedent for successfully selecting transfer functions that correct to the parton shower level [20], it is reasonable to suspect that most of the strongly-distorting effects discussed in Chapter 4 occur during the parton shower. It is therefore worth considering whether a two-step approach for the selection for the transfer function offers a better opportunity for success. In this case, the established methods for correcting from the hadronic end state to the parton shower can be utilised; these prescribe a double Gaussian approach to the transfer function, as well as for different transfer function parameters depending on the value of the pseudorapidity of the jet in question [20]. Since this covers the correction between the hadron jets and the parton shower, what remains to be done in the second step is to correct between the parton shower and the original partons themselves.

A promising option for this correction is unfolding. Unfolding aims to map experimental quantities back to their original values in light of distorting statistical effects, such as detector imperfections or uncontrollable physical processes such as radiation [21], which map well onto most of the issues presented in Chapter 4. Based on the premise that measured ($g(y)$) and actual distribution ($f(x)$) can be mapped by a resolution function $A(y, x)$:

$$g(y) = \int A(y, x) f(x) dx \tag{5.1}$$

If the distributions $g(y)$ and $f(x)$ are represented as histograms, then they can be written as vectors y and x with respective dimensions n and m representing the number of bins contained within each histogram. (5.1) can then be represented as the matrix equation

$$y = Ax \tag{5.2}$$

where A is an $n \times m$ matrix that transforms the vector x containing the original data into the the vector y containing the observed data, with the elements a_{ij} of A reflecting the probability of an observed event in the ith bin of y originating from an event in the jth bin of x. A, however, is generally not known analytically; it needs to be constructed from sample data such as from a Monte Carlo simulation of the process in question. Since y is known, once A is also known, it becomes possible to approximate the original distribution by solving for

x [21].

Through unfolding, then, it should be possible to estimate the original parton data from the parton shower results and thereby give insight as to how this correction should be factored into the overall transfer function. In order to carry out the unfolding procedure, efforts turned to the various algorithmic implementations for unfolding provided by the software package RooFitUnfold, an extension of the ROOT-based RooFit statistical software [22]. The introduction of the use of this software and unfolding in general motivates several changes of approach with respect to the treatment of the particle data.

5.2 Jet Reconstruction

In order to investigate the parton shower as an event of interest, it is necessary to collect the particles associated with it into separate jets that can be compared to both the original partons as well as the end-state hadron jets. In PYTHIA, the particle codes 91 and 92 denote the cluster and string fragmentations of the parton system, respectively [10] (See Figure 5.1); the particles of the parton shower can thus be identified as all of those (non-gluon) particles whose daughters are either one of these objects, as shown in Figure 5.2.

212	(cluster)		11	91	70	213	213	-0.10519	0.11450	1522.44715	1522.44759	1.14958
213	(Delta+)		11	2214	212	553	554	-0.08410	0.14304	1646.25486	1646.25534	1.25360
214	(dbar)	V	11	-1	68	555	555	0.34712	0.49280	2291.07236	2291.07244	0.00000
215	(s)	A	11	3	69	555	555	1.45748	1.20506	78.26033	78.28477	0.49919
216	(string)		11	92	23	217	248	-8.48511	-0.14134	24.71842	228.73820	227.24028
217	(rho-)		11	-213	216	563	564	0.17646	0.75080	42.72711	42.74136	0.78933
218	(omega)		11	223	216	565	567	0.42222	0.57090	46.60360	46.61557	0.78222
219	(omega)		11	223	216	568	570	-0.04952	0.83522	13.33435	13.38349	0.78276
220	(rho+)		11	213	216	571	572	-0.63867	0.25232	10.30399	10.33762	0.47193
221	(rho0)		11	113	216	573	574	-0.18284	-0.15637	1.23750	1.46142	0.73923
222	pbar-		1	-2212	216	0	0	-1.21934	-0.41196	1.43441	2.14345	0.93827
223	Lambda0		1	3122	216	0	0	-0.03823	-0.36425	-0.52307	1.28549	1.11568
224	(K*0)		11	313	216	575	576	-0.69204	-0.08862	-0.19779	1.20401	0.96111
225	(Deltabar-)		11	-2214	216	577	578	-0.65008	-0.60862	-0.52368	1.57947	1.19476
226	(rho+)		11	213	216	579	580	0.54606	-0.59454	-4.17037	4.28535	0.56620
227	p+		1	2212	216	0	0	-0.87881	-0.69705	-1.48918	2.08715	0.93827
228	(eta)		11	221	216	581	583	0.35595	-0.29469	-1.93429	2.06270	0.54745
229	pbar-		1	-2212	216	0	0	-0.63136	-0.47651	-27.46303	27.49043	0.93827
230	n0		1	2112	216	0	0	0.19378	-0.51476	-9.12500	9.18972	0.93957
231	(rho+)		11	213	216	584	585	-0.69282	-0.66865	-17.65720	17.71334	1.02895
232	(omega)		11	223	216	586	588	-0.33554	0.74509	-5.17380	5.29347	0.76473
233	(pi0)		11	111	216	589	590	0.18540	-0.24830	-0.81512	0.88242	0.13498
234	(pi0)		11	111	216	591	592	-0.36664	0.88163	-0.57281	1.12162	0.13498
235	(omega)		11	223	216	593	595	-0.54208	-0.15371	-0.20789	0.95316	0.74014
236	(omega)		11	223	216	596	598	-0.66632	0.56755	1.32241	1.76716	0.77973
237	(rho-)		11	-213	216	599	600	0.68665	0.16435	0.93027	1.41569	0.80017
238	(Deltabar+)		11	-1114	216	601	602	-0.34578	0.12548	2.64384	2.93552	1.22151
239	(pi0)		11	111	216	603	604	-0.89793	0.61807	-0.49948	1.20664	0.13498
240	n0		1	2112	216	0	0	0.15700	0.40506	0.24425	1.06357	0.93957
241	(rho0)		11	113	216	605	606	-0.28417	-0.43313	0.21652	1.01390	0.84425
242	(pi0)		11	111	216	607	608	0.01245	-0.11700	-0.89228	0.91007	0.13498
243	(eta)		11	221	216	609	611	-0.49684	-0.11617	0.42336	0.85982	0.54745
244	(omega)		11	223	216	612	614	-0.30225	0.06556	-1.51900	1.72445	0.75546
245	pi-		1	-211	216	0	0	-0.71809	-0.43502	-9.45496	9.49319	0.13957
246	(pi0)		11	111	216	615	616	-0.04956	0.25952	-4.39454	4.40455	0.13498
247	(pi0)		11	111	216	617	618	-0.14745	-0.07529	-5.16933	5.17374	0.13498
248	pi+		1	211	216	0	0	-0.39476	0.07175	-4.92036	4.93866	0.13957

Figure 5.1: Sample cluster (lines 212-215) and string (lines 216-248) with content.

With the parton shower elements identified from the event list, these can be fed into another instance of FastJet. The parameters used were identical to those used in the generation of the hadron jets described in Chapter 3 in order to permit consistent comparisons and in the absence of any apparent reasons to make any particular alterations to them. The jet reconstruction of the hadron jets themselves remains unchanged.

23	(ubar)	A	11	-2	1	216	216	1.04244	0.92967	58.56530	58.58289	0.33000
24	(g)	V	11	21	1	216	216	-0.38499	1.46766	34.32307	34.35659	0.00000
25	(g)	A	11	21	1	216	216	-0.00460	-0.42283	10.34325	10.35189	0.00000
26	(g)	V	11	21	1	216	216	-0.94129	0.48667	11.87060	11.91781	0.00000
27	(g)	A	11	21	2	216	216	-0.05275	-0.49529	0.01686	0.49838	0.00000
28	(g)	V	11	21	2	216	216	-2.09091	-0.12888	1.34700	2.49057	0.00000
29	(g)	A	11	21	2	216	216	0.34083	-0.50315	0.09414	0.61498	0.00000
30	(g)	V	11	21	2	216	216	-1.14030	-1.10775	-3.46912	3.81604	0.00000
31	(g)	A	11	21	2	216	216	0.29346	-1.15898	-2.53326	2.80120	0.00000
32	(g)	V	11	21	2	216	216	0.32530	-0.60311	-4.47341	4.52559	0.00000
33	(g)	A	11	21	2	216	216	-1.19581	-1.19171	-47.52019	47.55017	0.00000
34	(g)	V	11	21	2	216	216	-0.10680	0.19531	-1.52085	1.53705	0.00000
35	(g)	A	11	21	2	216	216	-0.24677	0.18852	-7.50479	7.51122	0.00000
36	(g)	V	11	21	1	216	216	-0.82079	0.39761	-4.22439	4.32172	0.00000
37	(g)	A	11	21	1	216	216	0.39582	0.04978	-0.20587	0.44892	0.00000
38	(g)	V	11	21	1	216	216	0.29430	1.11511	1.08787	1.58541	0.00000
39	(g)	A	11	21	1	216	216	-1.11497	0.58974	2.91312	3.17446	0.00000
40	(g)	V	11	21	1	216	216	-0.48303	1.24948	1.29874	1.86581	0.00000
41	(g)	A	11	21	1	216	216	-0.05556	-0.05582	0.43605	0.44311	0.00000
42	(g)	V	11	21	1	216	216	0.21653	-0.21502	0.72239	0.78420	0.00000
43	(g)	A	11	21	1	216	216	-0.92104	-0.67331	1.07871	1.57012	0.00000
44	(u)	V	11	2	2	216	216	-1.83417	-0.25503	-27.92681	27.99008	0.33000
45	(u)	A	11	2	1	249	249	1.59205	0.10457	-16.08317	16.16548	0.33000
46	(g)	V	11	21	1	249	249	2.60136	2.02759	-5.05248	6.03372	0.00000
47	(g)	A	11	21	1	249	249	0.88600	1.47666	-4.31624	4.64709	0.00000
48	(g)	V	11	21	1	249	249	1.87554	-0.10173	-14.49946	14.62062	0.00000
49	(g)	A	11	21	1	249	249	-0.55546	-0.44301	-7.14125	7.17651	0.00000
50	(g)	V	11	21	1	249	249	-0.50378	0.22365	-7.36662	7.38721	0.00000
51	(g)	A	11	21	1	249	249	-0.26200	-0.89620	-34.37442	34.38710	0.00000
52	(g)	V	11	21	2	249	249	-1.11402	0.16565	-103.41246	103.41859	0.00000
53	(g)	A	11	21	2	249	249	0.28857	0.20044	-109.95278	109.95334	0.00000
54	(g)	V	11	21	1	249	249	-0.44156	-0.08148	0.71180	0.84159	0.00000
55	(g)	A	11	21	1	249	249	-0.21094	0.35686	0.16468	0.44606	0.00000
56	(g)	V	11	21	1	249	249	0.67295	-0.50504	1.36873	1.60665	0.00000
57	(g)	A	11	21	1	249	249	0.63105	-0.77716	0.56382	1.14895	0.00000
58	(g)	V	11	21	1	249	249	-0.25447	-1.22068	2.12728	2.46579	0.00000
59	(g)	A	11	21	1	249	249	-0.77895	0.58829	42.33418	42.34543	0.00000
60	(g)	V	11	21	1	249	249	0.14493	0.45185	12.41955	12.42861	0.00000
61	(g)	A	11	21	1	249	249	4.95995	-0.20794	155.71332	155.79243	0.00000
62	(g)	V	11	21	1	249	249	-0.30644	-0.10250	13.89023	13.89399	0.00000
63	(g)	A	11	21	4	249	249	0.33837	0.37319	0.22123	0.55019	0.00000
64	(g)	V	11	21	4	249	249	-1.45938	0.16258	-2.01544	2.49363	0.00000
65	(g)	A	11	21	4	249	249	-1.89379	-1.07854	-1.63132	2.72229	0.00000
66	(g)	V	11	21	4	249	249	-0.30424	-0.04429	-1.94174	1.96593	0.00000
67	(ubar)	A	11	-2	4	249	249	0.20977	-0.54805	-9.78195	9.80509	0.33000

Figure 5.2: Sample portion of the parton shower; the string shown in Figure 5.1 originates from the antiup quark in line 23, indicating that it is a part of the parton shower.

5.3 Data Preselection and Processing

5.3.1 Preselection

With both the parton shower jets and the hadron jets on hand, the events must once again be evaluated for their utility. As before, this involves ensuring that all of the signature particles can be identified, and many of the original assumptions regarding this process still hold — the leptons can still be treated as being clearly identifiable, and the rigorousness of experimental b-tagging continues to endorse the more accurate approach of using the angular separation. However, the results presented in Chapter 4 revealed the shortcomings of attempting to identify the light particles by reconstructing the mass of their parent particles and the difficulties this ultimately presented to MadWeight. This informed a move away from the original methodology as described in Chapter 3.3.1. Instead, the ΔR approach is extended to the light quarks as well.

This is admittedly difficult to justify experimentally, for the same reasons that motivated only using this method for the bottom and anti-bottom quarks and not for the light quarks outlined in Chapter 3.3.1 — while bottom/anti-bottom quarks are readily identified and thus permit a tagging procedure that is more invasive than physical constraints would otherwise allow in order to reflect that, no such justification follows for the light quarks. The need to now compare the hadron jets to the parton shower jets instead of the original partons — and thus to an object whose quantitative properties are not necessarily known — alongside the very low tolerance to discrepancies demonstrated by MadWeight nevertheless demands a more precise method. In order to continue the investigation into the merits of the matrix element method in general and its implementation through MadWeight in particular, this particular weakness was accepted, as the results can in any case act as a proof-of-concept, even if not all of the methods employed have a strict experimental counterpart at this point.

Additionally, as the hadron jets are now to be corrected only to the parton shower level, the tagging process must attempt identify them with respect to the parton shower jets. This is, of course, only meaningful if the parton shower jets have already successfully tagged, which must then occur with reference to the original partons for both methodological (the correction of the parton shower jets to the parton level) as well as practical (there is no other reference where the identity of the particles is already known) reasons.

Thus, jet tagging has also been elevated into a two-step process; both, however, correspond closely to the b-tagging process described in Chapter 3.3.1, including that only jets with pseudorapidity $\eta < 2.5$ are taken into consideration. The minimum transverse momentum is relaxed for the parton shower jets, down to 5 GeV for events analyzed under the top/anti-top hypothesis and even to 0 GeV for events analyzed with the di-Higgs hypothesis. This was done to ensure that a workable number of usable events were available. As the justification of this limit, being a stricter version of that used by ATLAS, hinged on its experimental applicability, abandoning it for the non-detectable and therefore non-experimental parton shower jets in favor of a larger, workable sample of events seems acceptable.

The main change is that instead of a binary check between the angular separations with the bottom and anti-bottom quarks, the ΔR-values between the jets and all four quarks are evaluated. Again, only the smallest value among these is tested, and the jet is identified as either a bottom, anti-bottom, or one of the two light quarks according to the particle associated with that value, provided it lies under the limit of $\Delta R = 0.5$. If multiple jets match these criteria for any particle, the jet with the minimum ΔR-value is selected. This process is carried out once with the original partons as the reference for the parton shower jets, and then repeated with identified parton shower jets as the reference for the hadron jets for all events where all four quarks have been matched to a parton shower jet.

As before, all particles of interest are needed for MadWeight's calculations. Thus, only events where all four quarks can be matched to jets on both the parton shower and the hadron level are submitted to processing; any events with less than four total jets on either the parton shower level or the hadron level are excluded from tagging and processing entirely. The number of events with sufficient parton shower and hadron jets per data set are given in Table 5.1.

Data Set	Events Remaining After Parton-Shower Jet-Matching	Events Remaining After Subsequent Hadron Jet-Matching
tt	353525	203098
tH	406608	214733
HH	269661	105750
Ht	219084	98191

Table 5.1: Results of the parton-to-parton-shower and parton-shower-to-hadron jet-matching processes for each data set. Each data set starts out with one million events.

5.3.2 Processing

The processing of the valid events remains largely unchanged from how it is described in Chapter 3.3.2. The process for the original partons and the hadron jets is identical, and it is simply now repeated for the parton shower jets. The necessary values are stored in TLorentzVector objects during the tagging process and are passed on in a string into a separate file for further analysis, where they can be treated in the same manner as the other data as desired.

However, additional data is needed for the unfolding process, namely as training data (see Chapter 5.4). For this, the particle data for every event as well as the particle shower data for every event for which all four quarks could be identified were written to two additional files in the same format as the main data.

5.4 Unfolding

RooFitUnfold uses a common interfaces across all of its unfolding methods [22], so it can be executed consistently irrespective of the unfolding algorithm that is ultimately chosen.

The first step is to build and train the resolution matrix A described in Chapter 5.1, also referred to as the response matrix R in RooFitUnfold documentation [22]. To mirror as closely as possible the methodology used in Chapter 3.4 for the investigation and correction of the energy differences between the original partons and the hadron jets, the response matrix is initialized with 15 bins over a range from 5 $\sqrt{GeV}$ to 20 $\sqrt{GeV}$, representing once again the square roots of the energies. This approach aims to allow a more direct comparison of the resulting distributions.

The training data, drawn from the two files mentioned at the end of Chapter 5.3, consists of the energies of the particles of the first 25% of events along with the valid parton shower jet energies where applicable. For each event used, there is a check to see if the corresponding parton shower data is available. If it is, then the squares of both energies are submitted as a pair to fill the response matrix. If not, then the square of the particle energies alone are passed along as “missed” events.

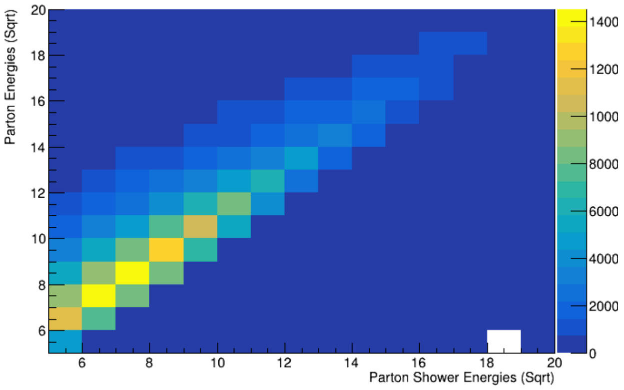
After generating the response matrix, a histogram representation of the square roots of the measured particle shower jet energies is constructed with the same parameters: 15 bins spanning 5 to 20 $\sqrt{GeV}$. With the response matrix and the histogram containing the measured data at hand, the unfolding procedure proper can be carried out. For the purposes of this investigation, RooFitUnfold’s implementation of D’Agostini’s iterative Bayesian unfolding methodology, also known as Richardson-Lucy deconvolution, was used [22]. For a first look, the default value of four interactions was kept.

Chapter 6

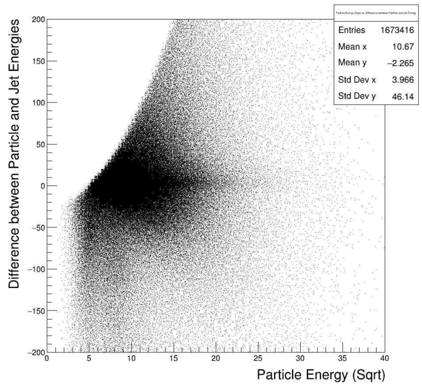
Results II

The initial results of the unfolding procedure were promising. Firstly, the structure of the resolution matrix resembled closely the observed behavior of the parton/(hadron) jet energy differences (Figure 6.1). The higher concentration near the diagonal of the matrix, representing partons and parton jets that have the same binning or are at least only a couple of bins removed from each other and thus have similar energies, aligns with the clustering of parton/hadron jet energy differences about 0. Meanwhile, the larger number of bins where the energy of the parton exceeds that of the parton shower jet than vice versa — some of the bins with a larger parton shower energy are even empty — mirrors the other clustering of positive energy differences along the hard upper limit that defines the shape of the energy difference distribution. This simultaneously seems to confirm the suspicion that the causes for the observed energy differences originate during the parton shower, since these trends are already observable at that level, as well gives confidence that the unfolding process will be able to address these issues.

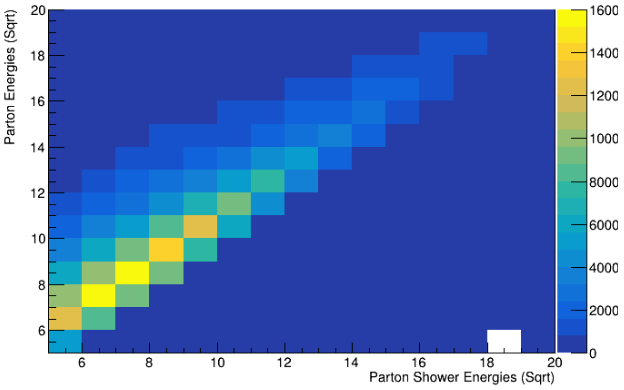
Indeed, the results of the unfolding itself matched or even exceeded original expectations. Figure 6.2 demonstrates the impact; the distribution of the (square roots) of the energies of the remaining original parton data is given in green, while the same distribution for the parton shower jets is given in solid blue. The unfolded distribution is represented as individual line segments with an error bar, also in blue. In all cases, the unfolding process restored the parton shower jet energy distribution to almost exactly the same shape as that of the original particle energy distribution as desired.



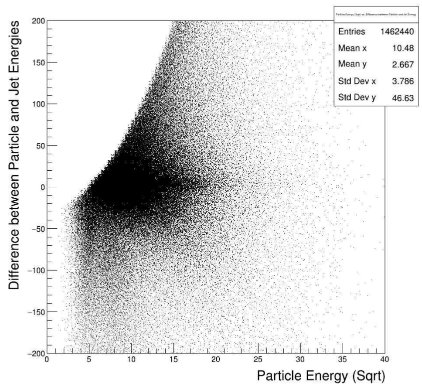
(a) tt



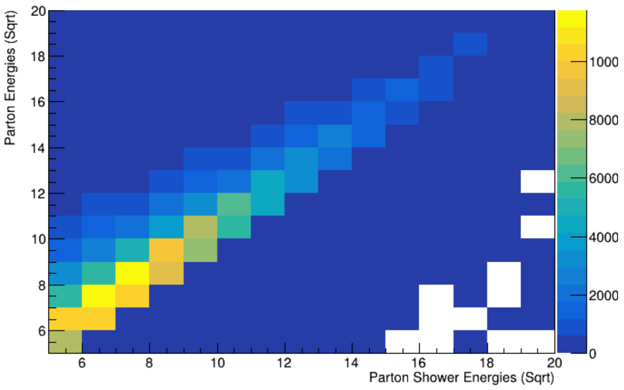
(b) tt



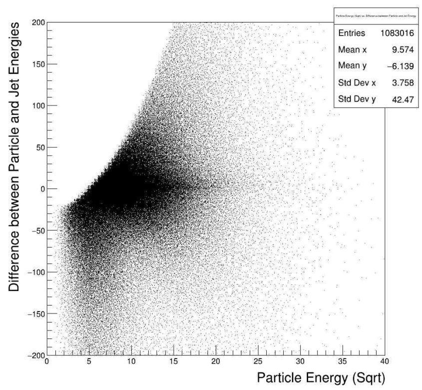
(c) tH



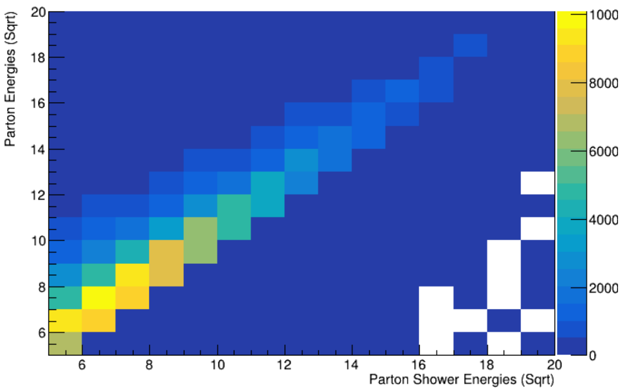
(d) tH



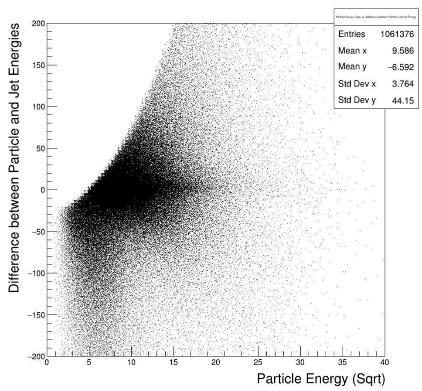
(e) HH



(f) HH

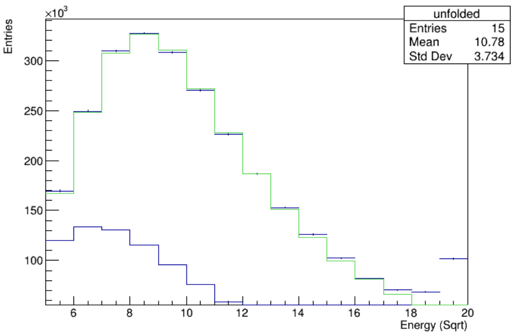


(g) Ht

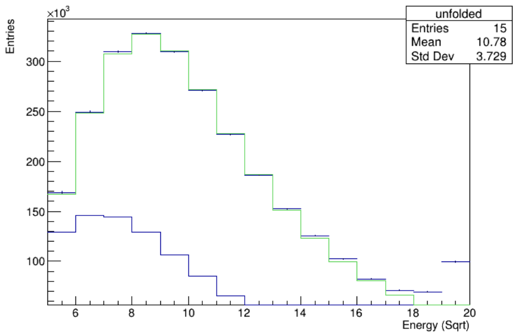


(h) Ht

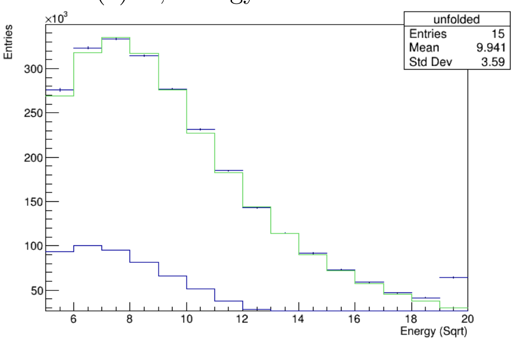
Figure 6.1: Comparison of the response matrix between the parton and parton shower energies (left) and the distribution of energy differences between the partons and reconstructed jets and the parton energies (right).



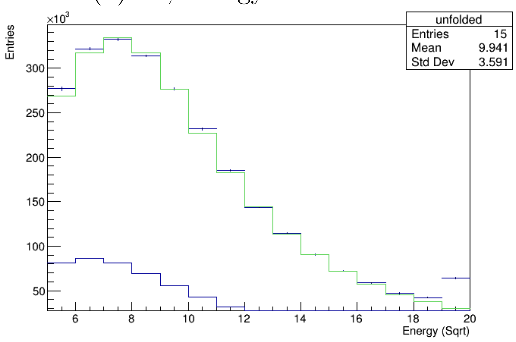
(a) tt, Energy Distributions



(b) tH, Energy Distributions



(c) HH, Energy Distributions



(d) Ht, Energy Distributions

Figure 6.2: Distribution of the energies (in $\sqrt{GeV}$) of the original particles (green), parton shower jets (continuous blue), and unfolded parton shower jets (individual blue bars).

Chapter 7

Conclusions

As an analysis of the utility of MadWeight as a tool for separating $HH \rightarrow b\bar{b}W^+W^-$ decays from their top/anti-top background, the results of this research have been insightful, if not initially as straightforward as might have been hoped. While the weights calculated from the original events indicated noticeable differences between decay events where the correct matching/reconstruction method is chosen and those where the matching hypothesis was incorrect, this did not carry over to the weights calculated from the reconstructed events, which more closely represent observable data. The initial approach, which adopted a single-Gaussian transfer function, ultimately produced weights for the correctly reconstructed events that were only marginally distinguishable from events incorrectly evaluated with the same initial hypothesis.

However, the fact that a distinction was visible on the level of the original events demonstrates that the issue is not strictly with MadWeight, but systematic errors with the particular approach, which means that MadWeight's utility has not been precluded either — giving hope that changes in this approach might yet offer better results. After MadWeight's high sensitivty was uncovered, it became clear that the possible energy differences between the original particles and their matched jets were a liability. As the causes for these were predominantly statistical, either due to uncontrollable physical process or uncertainties associated with detector conditions, the changes that could be made to improve things from the data processing perspective were limited. Instead, making changes to the selection of the transfer function — whose job it is to deal with precisely these issues — was the most workable approach.

The chosen solution of splitting the tagging process into two stages to permit an additional referencing of the parton shower to be compared to the original particles via unfolding proved to be very successful. Not only did the structure of the resolution matrix verify the suspicion that the majority of the distorting statistical effects were occurring during the parton shower, but the subsequent unfolding of the data managed to almost perfectly restore the energies of the original particles from the parton shower jets, clearing the way for the implementation of a further transfer function that corrects between the parton-shower and the hadronic end-state jets.

More work remains to be done to determine a way to integrate the results of the unfolding into a transfer function that is compatible with MadWeight's methodology as well as with the transfer function needed to correct for differences between the parton shower and the end-state of the event, which should ideally permit the calculation of weights from the reconstructed event that match the actual values well. If this can be achieved, then there is potential to be found in using neural networks to predict whether an event is the result of top/anti-top or di-Higgs decay based on training data that is the result of the successful calculation of meaningful weights; this would also limit the amount of further such weight calculations that need to be run, which would be a serious improvement due to the time-intensive nature of said calculations. Additionally, if any improvements could be made to particle/jet identification, this would almost certainly further improve the results of any weight calculations made with MadWeight.

Appendix A

Impact of a 10% Energy Difference Filter on B-Tagging Results

	Tagged w/out Enegry Filter(E_{NF})	Tagged w/ Filter($\Delta E \leq 10\%$)(E_F)	Impact of Energy Energy Filter(E_F/E_{NF})
Detected b	578724	438412	0.757549
Detected $\bar{b}$	586052	439267	0.749536
Both Detected	479409	274338	0.572242

Table A.1: Count of tt events where the bottom and anti-bottom quarks were tagged with and without an additional filter that limited the energy difference to 10%, as well as the same count where both b-quarks were tagged. The last column gives the fraction of events where the respective particles remained tagged after the imposition of the filter.

	Tagged w/out Enegry Filter(E_{NF})	Tagged w/ Filter($\Delta E \leq 10\%$)(E_F)	Impact of Energy Energy Filter(E_F/E_{NF})
Detected b	590879	447658	0.757614
Detected $\bar{b}$	598268	448156	0.749089
Both Detected	489804	280202	0.572070

Table A.2: Count of tH events where the bottom and anti-bottom quarks were tagged with and without an additional filter that limited the energy difference to 10%, as well as the same count where both b-quarks were tagged. The last column gives the fraction of events where the respective particles remained tagged after the imposition of the filter.

	Tagged w/out Enegry Filter(E_{NF})	Tagged w/ Filter($\Delta E \leq 10\%$)(E_F)	Impact of Energy Energy Filter(E_F/E_{NF})
Detected b	397252	282655	0.711526
Detected $\bar{b}$	397343	283010	0.712256
Both Detected	329777	181569	0.550581

Table A.3: Count of HH events where the bottom and anti-bottom quarks were tagged with and without an additional filter that limited the energy difference to 10%, as well as the same count where both b-quarks were tagged. The last column gives the fraction of events where the respective particles remained tagged after the imposition of the filter.

	Tagged w/out Enegry Filter(E_{NF})	Tagged w/ Filter($\Delta E \leq 10\%$)(E_F)	Impact of Energy Energy Filter(E_F/E_{NF})
Detected b	384286	273551	0.711842
Detected $\bar{b}$	384399	273845	0.712398
Both Detected	318609	175602	0.551152

Table A.4: Count of Ht events where the bottom and anti-bottom quarks were tagged with and without an additional filter that limited the energy difference to 10%, as well as the same count where both b-quarks were tagged. The last column gives the fraction of events where the respective particles remained tagged after the imposition of the filter.

	Tagged w/out Enegry Filter(E_{NF})	Tagged w/ Filter($\Delta E \leq 10\%$)(E_F)	Impact of Energy Energy Filter(E_F/E_{NF})
Detected b	350044	258958	0.739787
Detected $\bar{b}$	359397	272209	0.757405
Both Detected	254517	143413	0.563471

Table A.5: Count of Breit-Wigner tt ($m_t, m_{\bar{t}} < 150$ GeV) events where the bottom and anti-bottom quarks were tagged with and without an additional filter that limited the energy difference to 10%, as well as the same count where both b-quarks were tagged. The last column gives the fraction of events where the respective particles remained tagged after the imposition of the filter.

	Tagged w/out Enegry Filter(E_{NF})	Tagged w/ Filter($\Delta E \leq 10\%$)(E_F)	Impact of Energy Energy Filter(E_F/E_{NF})
Detected b	359478	265543	0.738691
Detected $\bar{b}$	368844	279285	0.757190
Both Detected	261862	147331	0.562628

Table A.6: Count of Breit-Wigner tH ($m_t, m_{\bar{t}} < 150$ GeV) events where the bottom and anti-bottom quarks were tagged with and without an additional filter that limited the energy difference to 10%, as well as the same count where both b-quarks were tagged. The last column gives the fraction of events where the respective particles remained tagged after the imposition of the filter.

	Tagged w/out Enegry Filter(E_{NF})	Tagged w/ Filter($\Delta E \leq 10\%$)(E_F)	Impact of Energy Energy Filter(E_F/E_{NF})
Detected b	697278	528244	0.757580
Detected $\bar{b}$	700054	525229	0.750269
Both Detected	594567	340768	0.573136

Table A.7: Count of Breit-Wigner tt ($m_t, m_{\bar{t}} > 195$ GeV) events where the bottom and anti-bottom quarks were tagged with and without an additional filter that limited the energy difference to 10%, as well as the same count where both b-quarks were tagged. The last column gives the fraction of events where the respective particles remained tagged after the imposition of the filter.

	Tagged w/out Enegry Filter(E_{NF})	Tagged w/ Filter($\Delta E \leq 10\%$)(E_F)	Impact of Energy Energy Filter(E_F/E_{NF})
Detected b	708864	537043	0.757611
Detected $\bar{b}$	711643	533673	0.749917
Both Detected	604585	346489	0.573102

Table A.8: Count of Breit-Wigner tH ($m_t, m_{\bar{t}} > 195$ GeV) events where the bottom and anti-bottom quarks were tagged with and without an additional filter that limited the energy difference to 10%, as well as the same count where both b-quarks were tagged. The last column gives the fraction of events where the respective particles remained tagged after the imposition of the filter.

	Tagged w/out Enegry Filter(E_{NF})	Tagged w/ Filter($\Delta E \leq 10\%$)(E_F)	Impact of Energy Energy Filter(E_F/E_{NF})
Detected b	539210	397594	0.737364
Detected $\bar{b}$	538916	404413	0.750419
Both Detected	419451	234618	0.559345

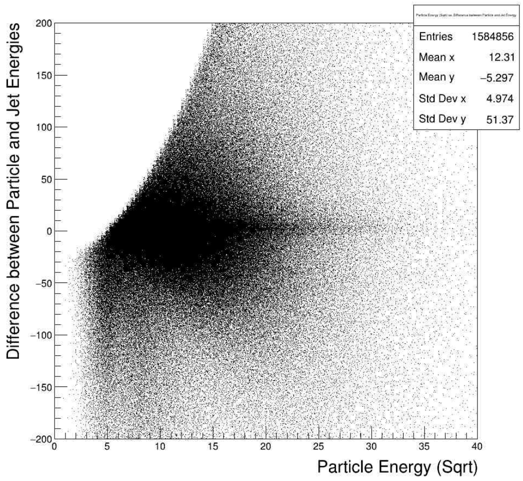
Table A.9: Count of Breit-Wigner tt (mixed flank) events where the bottom and anti-bottom quarks were tagged with and without an additional filter that limited the energy difference to 10%, as well as the same count where both b-quarks were tagged. The last column gives the fraction of events where the respective particles remained tagged after the imposition of the filter.

	Tagged w/out Enegry Filter(E_{NF})	Tagged w/ Filter($\Delta E \leq 10\%$)(E_F)	Impact of Energy Energy Filter(E_F/E_{NF})
Detected b	550842	405818	0.736723
Detected $\bar{b}$	550478	412839	0.749965
Both Detected	428941	239549	0.558466

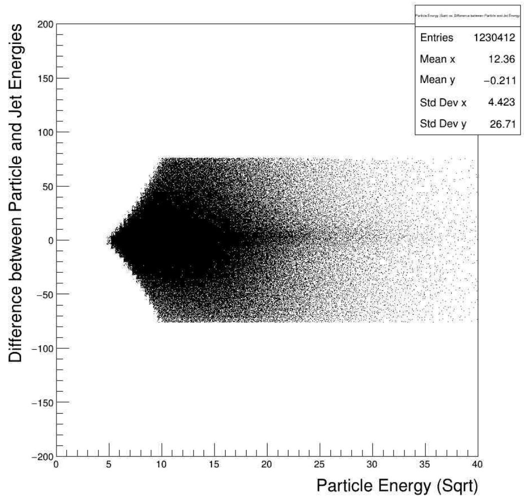
Table A.10: Count of Breit-Wigner tH (mixed flank) events where the bottom and anti-bottom quarks were tagged with and without an additional filter that limited the energy difference to 10%, as well as the same count where both b-quarks were tagged. The last column gives the fraction of events where the respective particles remained tagged after the imposition of the filter.

Appendix B

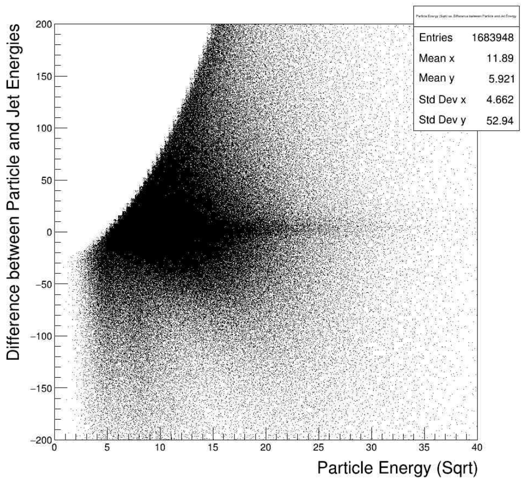
Noteworthy Energy Difference Distributions for $t\bar{t}$ Decay Events for Breit-Wigner-Distributed Top Quark Masses



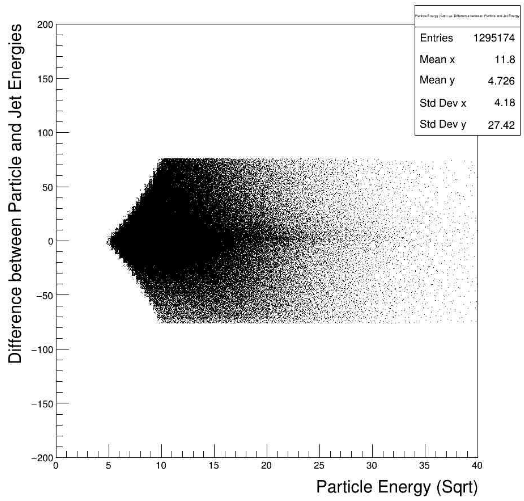
(a) tt ($m_t, m_{\bar{t}>195}$ GeV), Untrimmed



(b) tt ($m_t, m_{\bar{t}>195}$ GeV), Trimmed

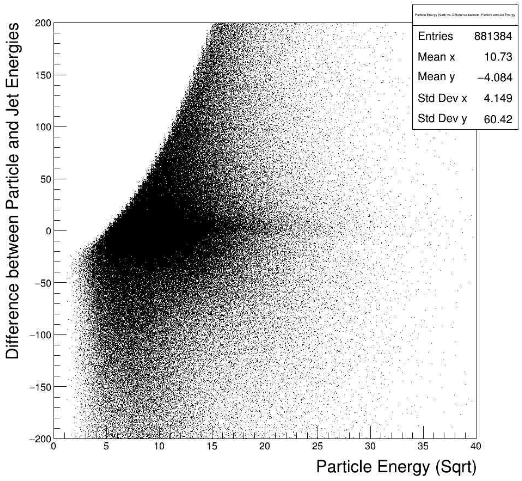


(c) tH ($m_t, m_{\bar{t}>195}$ GeV), Untrimmed

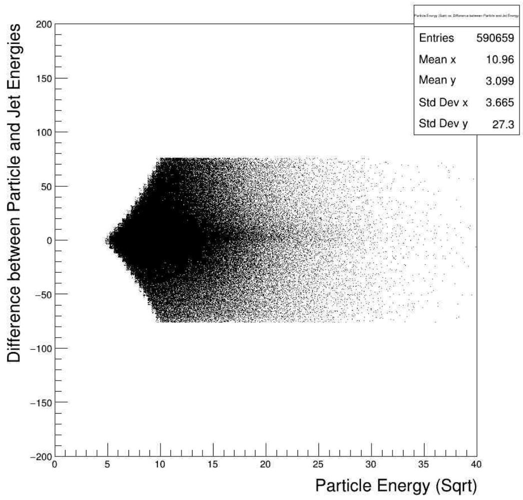


(d) tH ($m_t, m_{\bar{t}>195}$ GeV), Trimmed

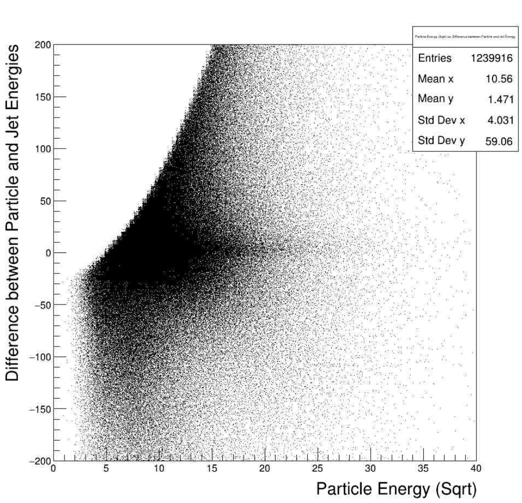
Figure B.1: Energy difference between the original particles and their associated jet compared to the square root of the particle energy, continued. The left column shows the full distribution, while the distributions shown on the right have had a limit $|\Delta E| \leq x^2 - 25$ (fixed at $|\Delta E| \leq 75$ at and after $x = 10$) imposed on both positive and negative differences that reflects the organic upper limit and mirrors it to the negative differences. The data set used for each plot is given underneath.



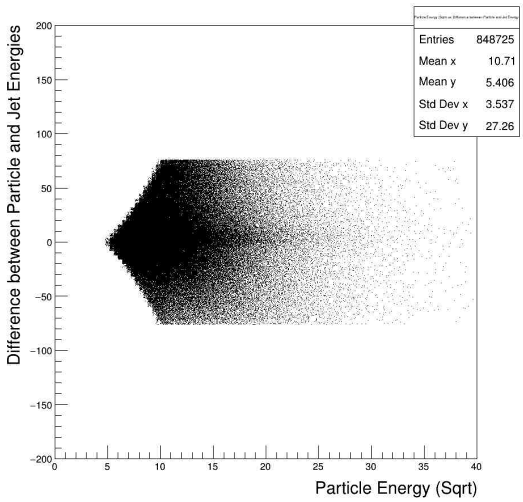
(a) tt (mixed flank), Untrimmed



(b) tt (mixed flank), Trimmed



(c) tH (mixed flank), Untrimmed

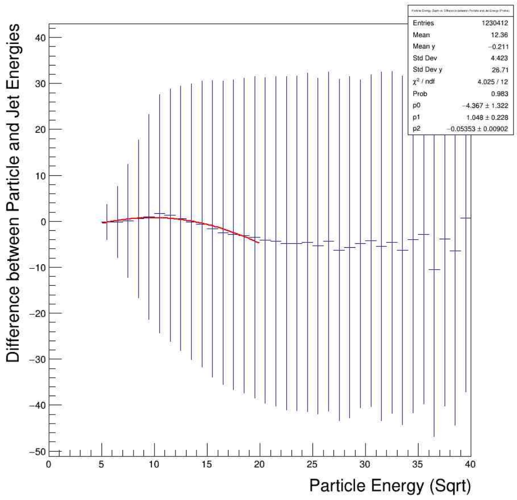


(d) tH (mixed flank), Trimmed

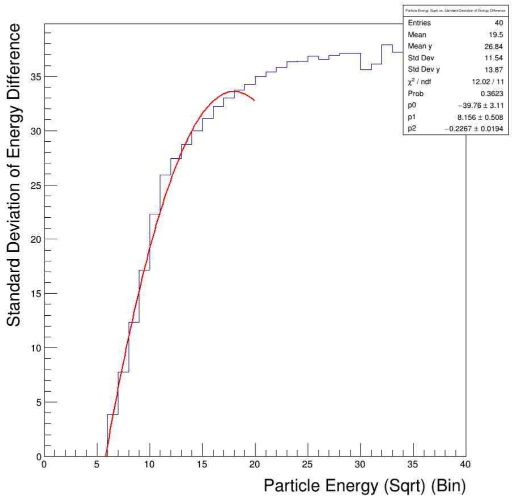
Figure B.2: Energy difference between the original particles and their associated jet compared to the square root of the particle energy, continued. The left column shows the full distribution, while the distributions shown on the right have had a limit $|\Delta E| \leq x^2 - 25$ (fixed at $|\Delta E| \leq 75$ at and after $x = 10$) imposed on both positive and negative differences that reflects the organic upper limit and mirrors it to the negative differences. The data set used for each plot is given underneath.

Appendix C

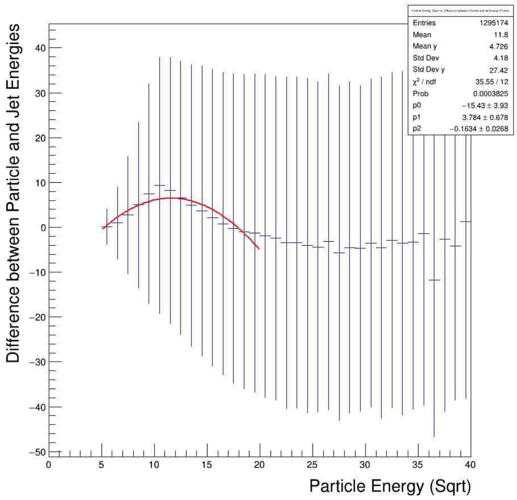
Noteworthy Parametrizations of the Profile Histograms of the Energy Differences and the Standard Deviation Thereof for $t\bar{t}$ Decay Events for Breit-Wigner-Distributed Top Quark Masses



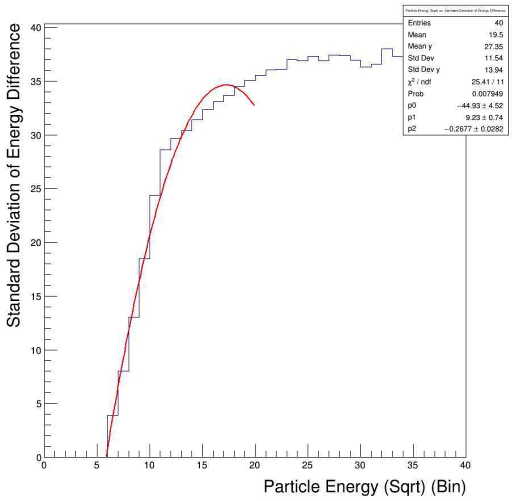
(a) tt ($m_t, m_{\bar{t}} > 195$ GeV)



(b) tt ($m_t, m_{\bar{t}} > 195$ GeV)

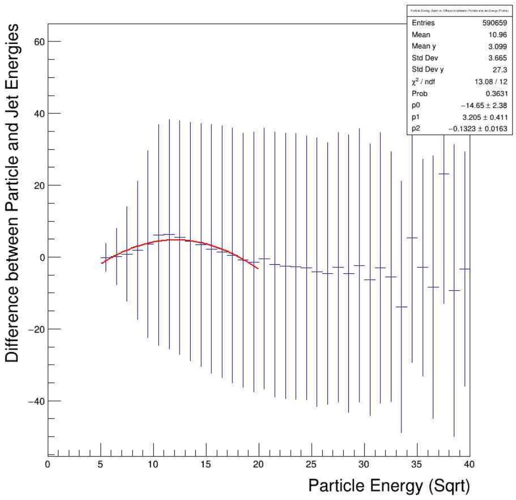


(c) tH ($m_t, m_<\bar{t} > 195$ GeV)

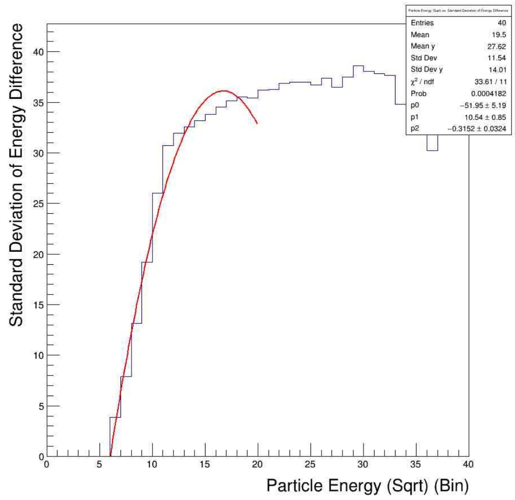


(d) tH ($m_t, m_{\bar{t}} > 195$ GeV)

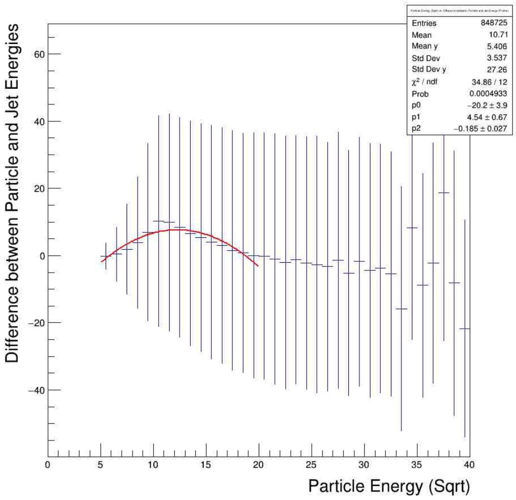
Figure C.1: Profile histograms of the (trimmed) energy differences between original particles and the associated jet (left) and of the standard deviation of each bin of the former (right) compared to the square root of the particle energy, continued. The second-degree-polynomial parametrization over a range of $\sqrt{E} = 5$ to $\sqrt{E} = 20$ for each plot is also shown. The data set used for each plot is given underneath.



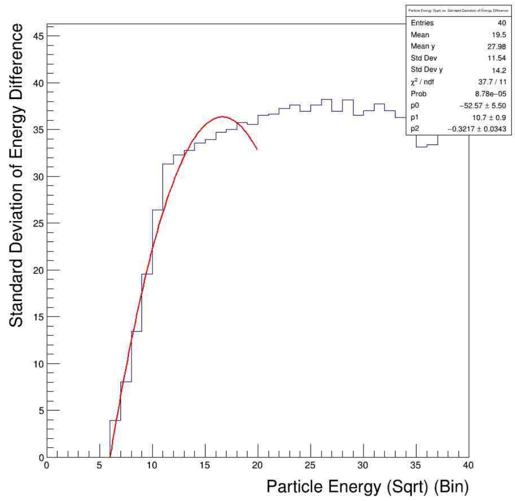
(a) tt (mixed flank)



(b) tt (mixed flank)



(c) tH (mixed flank)

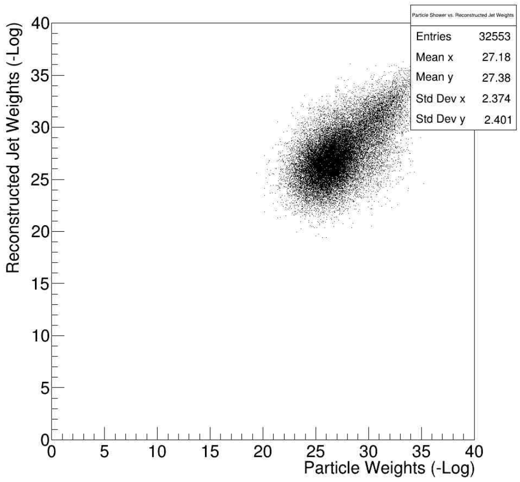


(d) tH (mixed flank)

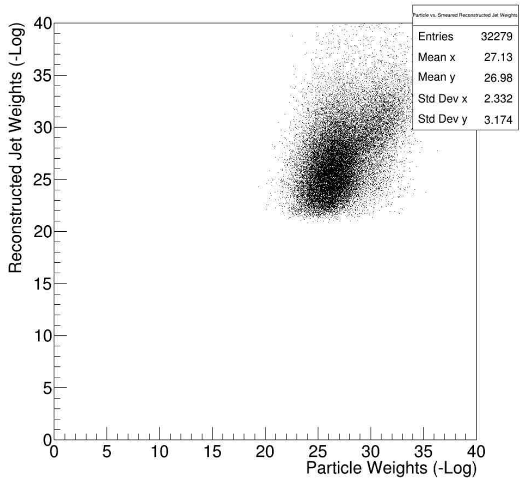
Figure C.2: Profile histograms of the (trimmed) energy differences between original particles and the associated jet (left) and of the standard deviation of each bin of the former (right) compared to the square root of the particle energy, continued. The second-degree-polynomial parametrization over a range of $\sqrt{E} = 5$ to $\sqrt{E} = 20$ for each plot is also shown. The data set used for each plot is given underneath.

Appendix D

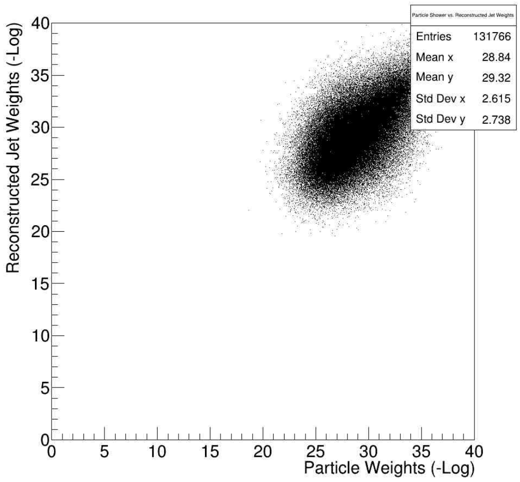
Plot of the Negative Base-10 Logarithms of the Weights of the Reconstructed Events Against Those of the Original Events for $t\bar{t}$ Decay Events for Breit-Wigner-Distributed Top Quark Masses



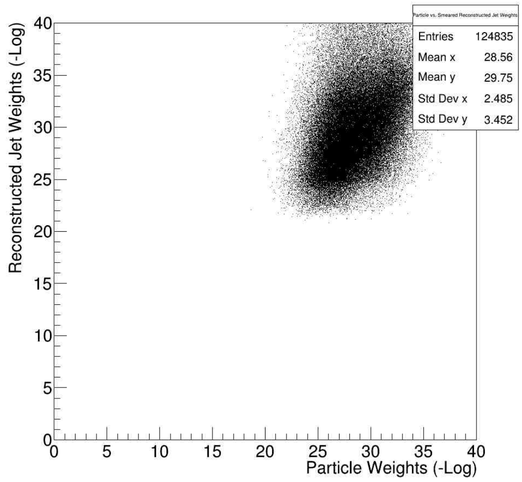
(a) tt ($m_t, m_{\bar{t}} < 150$ GeV), Original vs Reconstructed (Delta)



(b) tt ($m_t, m_{\bar{t}} < 150$ GeV), Original vs Reconstructed (Single-Gaussian)

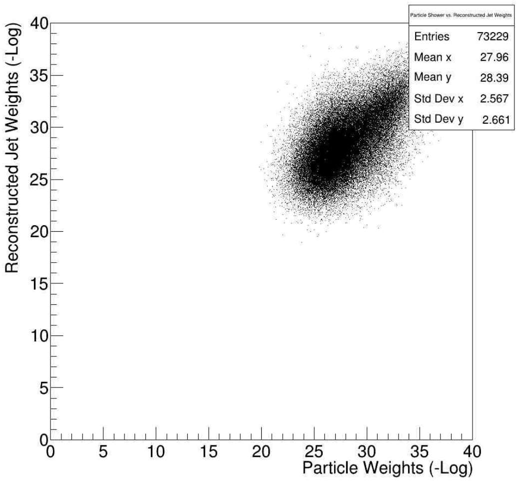


(c) tt ($m_t, m_{\bar{t}} > 195$ GeV), Original vs Reconstructed (Delta)

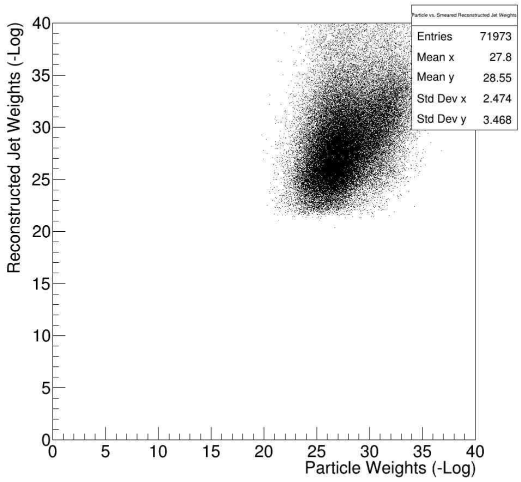


(d) tt ($m_t, m_{\bar{t}} > 195$ GeV), Original vs Reconstructed (Single-Gaussian)

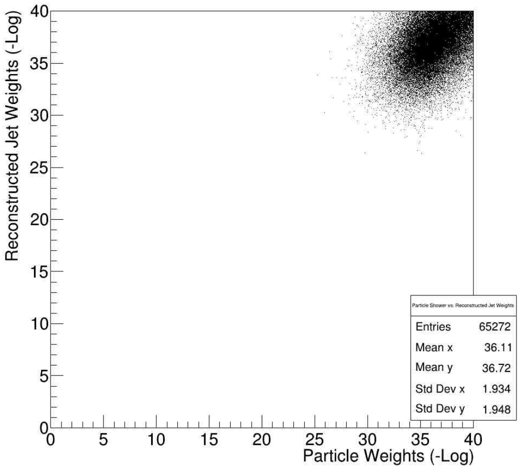
Figure D.1: Plot of the negative base-10 logarithms of the weights of the reconstructed events against those of the original events. The left column uses the weights of the reconstructed events where the delta function was used as the transfer function, while the right column uses a single-Gaussian distribution as the transfer function. The data set used for each plot is given underneath.



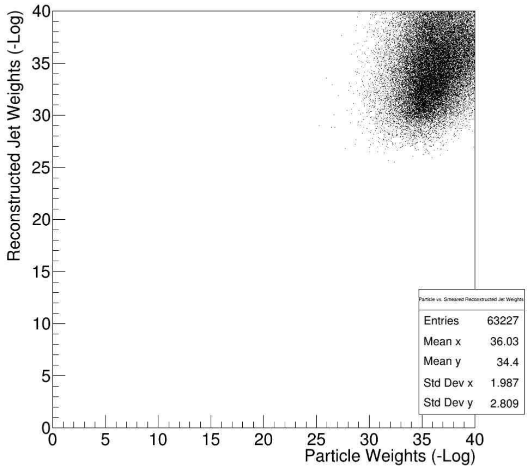
(a) tt (mixed flank), Original vs Reconstructed (Delta)



(b) tt (mixed flank), Original vs Reconstructed (Single-Gaussian)

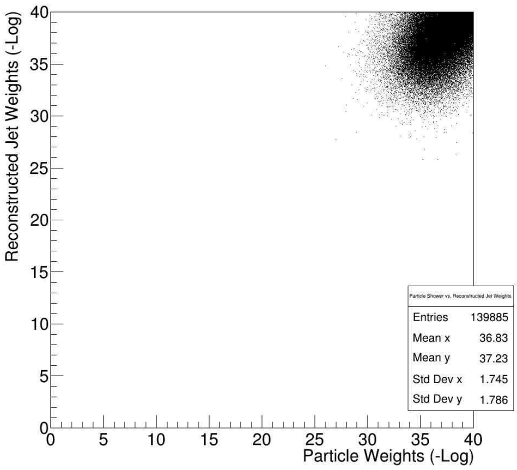


(c) tH ($m_t, m_{\bar{t}} < 150$ GeV), Original vs Reconstructed (Delta)

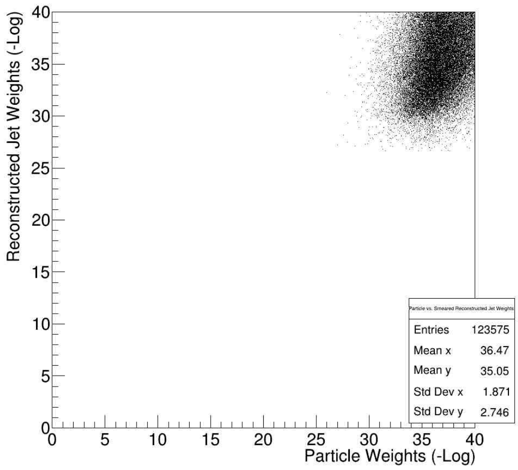


(d) tH ($m_t, m_{\bar{t}} < 150$ GeV), Original vs Reconstructed (Single-Gaussian)

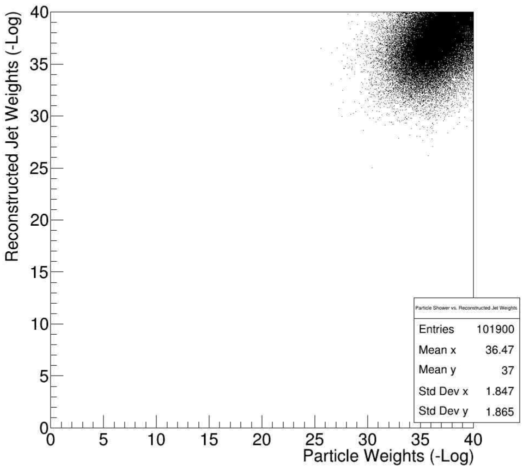
Figure D.2: Plot of the negative base-10 logarithms of the weights of the reconstructed events against those of the original events, continued. The left column uses the weights of the reconstructed events where the delta function was used as the transfer function, while the right column uses a single-Gaussian distribution as the transfer function. The data set used for each plot is given underneath.



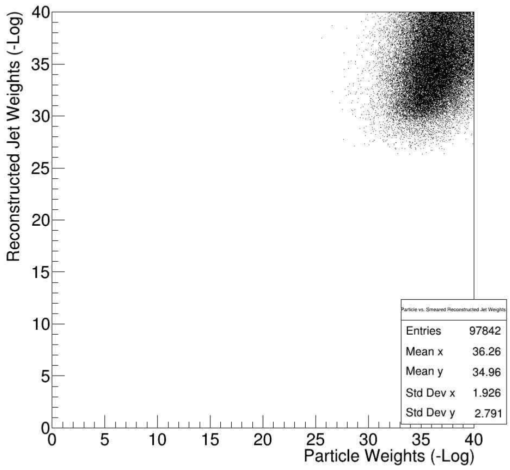
(a) tH ($m_t, m_{\bar{t}} > 195$ GeV), Original vs Reconstructed (Delta)



(b) tH ($m_t, m_{\bar{t}} > 195$ GeV), Original vs Reconstructed (Single-Gaussian)



(c) tH (mixed flank), Original vs Reconstructed (Delta)



(d) tH (mixed flank), Original vs Reconstructed (Single-Gaussian)

Figure D.3: Plot of the negative base-10 logarithms of the weights of the reconstructed events against those of the original events, continued. The left column uses the weights of the reconstructed events where the delta function was used as the transfer function, while the right column uses a single-Gaussian distribution as the transfer function. The data set used for each plot is given underneath.

Appendix E

Overview of Events with Large Energy Differences for Bottom/Anti-Bottom Quarks (10%, 50%) between the Original Particle and the Reconstructed Jet

		4-Jet Events: 132133	5-Jet Events: 137995	6-Jet Events: 84495	7-Jet Events: 39536	8-Jet Events: 15515	9+ Jet Events: 8680
b	$\Delta E > 10\%$	70616(53.443%)	74917(54.290%)	46621(55.176%)	22137(55.992%)	8889(57.29%)	4961(57.15%)
	$\Delta E > 50\%$	7065(5.347%)	8318(6.028%)	5692(6.737%)	3020(7.639%)	1325(8.540%)	806(9.29%)
$\bar{b}$	$\Delta E > 10\%$	70933(53.683%)	75180(54.480%)	46701(55.271%)	22051(55.774%)	8895(57.33%)	5107(58.84%)
	$\Delta E > 50\%$	8079(6.114%)	9615(6.968%)	6365(7.533%)	3322(8.403%)	1370(8.830%	871(10.0%)
$Both$	$\Delta E > 10\%$	37994(28.754%)	40954(29.678%)	25695(30.410%)	12326(31.177%)	5102(32.88%)	2879(33.17%)
	$\Delta E > 50\%$	473(0.358%)	639(0.463%)	436(0.516%)	271(0.685%)	117(0.754%)	101(1.16%)

Table E.1: Count of tt events where the bottom, anti-bottom, and both b-quarks have percent differences in energy between the particle and the assigned jet exceeding 10% as well as 50%. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events: 112042	5-Jet Events: 119263	6-Jet Events: 75392	7-Jet Events: 36155	8-Jet Events: 14515	9+ Jet Events: 8243
b	$\Delta E > 10\%$	60078(53.621%)	64713(54.261%)	41725(55.344%)	20266(56.053%)	8336(57.43%)	4693(56.93%)
	$\Delta E > 50\%$	6013(5.367%)	7146(5.992%)	5043(6.689%)	2752(7.612%)	1250(8.612%)	762(9.24%)
$\bar{b}$	$\Delta E > 10\%$	60512(54.008%)	65326(54.775%)	41962(55.658%)	20271(56.067%)	8359(57.59%)	4850(58.84%)
	$\Delta E > 50\%$	6952(6.205%)	8497(7.125%)	5791(7.681%)	3100(8.574%)	1288(8.874%)	835(10.1%)
$Both$	$\Delta E > 10\%$	32502(29.009%)	35554(29.811%)	23156(30.714%)	11369(31.445%)	4796(33.04%)	2721(33.01%)
	$\Delta E > 50\%$	399(0.356%)	552(0.463%)	400(0.531%)	254(0.703%)	106(0.730%)	97(1.2%)

Table E.2: Count of tH events where the bottom, anti-bottom, and both b-quarks have percent differences in energy between the particle and the assigned jet exceeding 10% as well as 50%. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events: 123852	5-Jet Events: 91165	6-Jet Events: 38640	7-Jet Events: 12555	8-Jet Events: 3427	9+ Jet Events: 1115
b	$\Delta E > 10\%$	69927(56.460%)	52063(57.109%)	22637(58.584%)	7478(59.56%)	2013(58.74%)	686(61.5%)
	$\Delta E > 50\%$	8068(6.514%)	6523(7.155%)	3160(8.178%)	1110(8.841%)	323(9.43%)	122(10.9%)
$\bar{b}$	$\Delta E > 10\%$	69993(56.513%)	52282(57.349%)	22400(57.971%)	7363(58.65%)	2026(59.12%)	685(61.4%)
	$\Delta E > 50\%$	8120(6.556%)	6671(7.318%)	2999(7.761%)	1066(8.491%)	317(9.25%)	101(9.06%)
Both	$\Delta E > 10\%$	40336(32.568%)	30593(33.558%)	13452(34.814%)	4551(36.25%)	1236(36.07%)	440(39.5%)
	$\Delta E > 50\%$	685(0.553%)	675(0.740%)	354(0.916%)	142(1.13%)	42(1.2%)	18(1.6%)

Table E.3: Count of HH events where the bottom, anti-bottom, and both b-quarks have percent differences in energy between the particle and the assigned jet exceeding 10% as well as 50%. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events: 119055	5-Jet Events: 90394	6-Jet Events: 38698	7-Jet Events: 12632	8-Jet Events: 3449	9+ Jet Events: 1116
b	$\Delta E > 10\%$	67071(56.336%)	51585(57.067%)	22650(58.530%)	7513(59.48%)	2023(58.65%)	687(61.6%)
	$\Delta E > 50\%$	7649(6.425%)	6421(7.103%)	3158(8.161%)	1117(8.843%)	324(9.39%)	121(10.8%)
$\bar{b}$	$\Delta E > 10\%$	67173(56.422%)	51777(57.279%)	22400(57.884%)	7396(58.55%)	2040(59.15%)	686(61.5%)
	$\Delta E > 50\%$	7748(6.508%)	6558(7.255%)	2997(7.745%)	1071(8.478%)	318(9.22%)	101(9.05%)
Both	$\Delta E > 10\%$	38641(32.456%)	30311(33.532%)	13445(34.743%)	4565(36.14%)	1243(36.04%)	441(39.5%)
	$\Delta E > 50\%$	652(0.548%)	662(0.732%)	350(0.904%)	144(1.14%)	42(1.2%)	18(1.6%)

Table E.4: Count of Ht events where the bottom, anti-bottom, and both b-quarks have percent differences in energy between the particle and the assigned jet exceeding 10% as well as 50%. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events: 34662	5-Jet Events: 33405	6-Jet Events: 18495	7-Jet Events: 7585	8-Jet Events: 2734	9+ Jet Events: 1192
b	$\Delta E > 10\%$	18444(53.211%)	18089(54.151%)	10331(55.858%)	4236(55.85%)	1601(58.56%)	705(59.1%)
	$\Delta E > 50\%$	1817(5.242%)	1987(5.948%)	1305(7.056%)	525(6.92%)	236(8.63%)	140(11.7%)
$\bar{b}$	$\Delta E > 10\%$	18369(52.995%)	17951(53.737%)	10097(54.593%)	4240(55.90%)	1555(56.88%)	732(61.4%)
	$\Delta E > 50\%$	2044(5.897%)	2194(6.568%)	1372(7.418%)	610(8.04%)	246(9.00%)	130(10.9%)
Both	$\Delta E > 10\%$	9852(28.42%)	9796(29.32%)	5659(30.60%)	2359(31.10%)	906(33.1%)	438(36.7%)
	$\Delta E > 50\%$	124(0.358%)	154(0.461%)	121(0.654%)	42(0.55%)	16(0.59%)	19(1.6%)

Table E.5: Count of Breit-Wigner tt ($m_t, m_{\bar{t}} < 150$ GeV) events where the bottom, anti-bottom, and both b-quarks have percent differences in energy between the particle and the assigned jet exceeding 10% as well as 50%. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events: 68386	5-Jet Events: 64609	6-Jet Events: 37104	7-Jet Events: 16480	8-Jet Events: 6506	9+ Jet Events: 3308
b	$\Delta E > 10\%$	38072(55.672%)	36370(56.292%)	21457(57.829%)	9608(58.30%)	3878(59.61%)	1985(60.01%)
	$\Delta E > 50\%$	5477(8.009%)	5684(8.798%)	3668(9.886%)	1699(10.31%)	729(11.2%)	444(13.4%)
$\bar{b}$	$\Delta E > 10\%$	37141(54.311%)	35470(54.899%)	20740(55.897%)	9354(56.76%)	3739(57.47%)	2004(60.58%)
	$\Delta E > 50\%$	4799(7.018%)	4837(7.487%)	3107(8.374%)	1480(8.981%)	631(9.70%)	386(11.7%)
Both	$\Delta E > 10\%$	20737(30.323%)	20106(31.120%)	11998(32.336%)	5429(32.94%)	2221(34.14%)	1203(36.37%)
	$\Delta E > 50\%$	389(0.569%)	472(0.731%)	343(0.924%)	181(1.10%)	75(1.2%)	61(1.8%)

Table E.6: Count of Breit-Wigner tH ($m_t, m_{\bar{t}} < 150$ GeV) events where the bottom, anti-bottom, and both b-quarks have percent differences in energy between the particle and the assigned jet exceeding 10% as well as 50%. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events: 105961	5-Jet Events: 125630	6-Jet Events: 86676	7-Jet Events: 45002	8-Jet Events: 20157	9+ Jet Events: 12788
b	$\Delta E > 10\%$	51209(48.328%)	61605(49.037%)	43035(49.650%)	22865(50.809%)	10528(52.230%)	6815(53.29%)
	$\Delta E > 50\%$	3382(3.192%)	4295(3.419%)	3287(3.792%)	2021(4.491%)	1077(5.343%)	762(5.96%)
$\bar{b}$	$\Delta E > 10\%$	55578(52.451%)	67315(53.582%)	47298(54.569%)	25184(55.962%)	11544(57.270%)	7427(58.08%)
	$\Delta E > 50\%$	6124(5.779%)	8464(6.737%)	6529(7.533%)	3839(8.531%)	1899(9.421%)	1379(10.78%)
$Both$	$\Delta E > 10\%$	26884(25.372%)	33089(26.338%)	23549(27.169%)	12737(28.303%)	5991(29.72%)	3947(30.86%)
	$\Delta E > 50\%$	268(0.253%)	377(0.300%)	286(0.330%)	204(0.453%)	112(0.556%)	94(0.74%)

Table E.7: Count of Breit-Wigner tt ($m_t, m_{\bar{t}} > 195$ GeV) events where the bottom, anti-bottom, and both b-quarks have percent differences in energy between the particle and the assigned jet exceeding 10% as well as 50%. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events: 95728	5-Jet Events: 129730	6-Jet Events: 98554	7-Jet Events: 54687	8-Jet Events: 25745	9+ Jet Events: 16543
b	$\Delta E > 10\%$	49540(51.751%)	69546(53.608%)	53754(54.543%)	30386(55.563%)	14737(57.242%)	9553(57.75%)
	$\Delta E > 50\%$	5006(5.229%)	8321(6.414%)	6977(7.079%)	4333(7.923%)	2299(8.930%)	1598(9.660%)
$\bar{b}$	$\Delta E > 10\%$	50405(52.654%)	70479(54.327%)	54457(55.256%)	30825(56.366%)	14892(57.844%)	9684(58.54%)
	$\Delta E > 50\%$	5599(5.849%)	9119(7.029%)	7754(7.868%)	4899 (8.958%)	2508(9.742%)	1814(10.97%)
$Both$	$\Delta E > 10\%$	26106(27.271%)	37821(29.154%)	29723(30.159%)	16994(31.075%)	8478(32.93%)	5607(33.89%)
	$\Delta E > 50\%$	377(0.394%)	625(0.482%)	539(0.547%)	371(0.678%)	228(0.886%)	178(1.08%)

Table E.8: Count of Breit-Wigner tH ($m_t, m_{\bar{t}} > 195$ GeV) events where the bottom, anti-bottom, and both b-quarks have percent differences in energy between the particle and the assigned jet exceeding 10% as well as 50%. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events: 66929	5-Jet Events: 72353	6-Jet Events: 45426	7-Jet Events: 21860	8-Jet Events: 8857	9+ Jet Events: 4921
b	$\Delta E > 10\%$	33652(50.280%)	37030(51.180%)	23585(51.920%)	11440(52.333%)	4743(53.56%)	2735(55.58%)
	$\Delta E > 50\%$	2579(3.853%)	3120(4.312%)	2287(5.035%)	1238(5.663%)	532(6.01%)	356(7.23%)
$\bar{b}$	$\Delta E > 10\%$	35450(52.967%)	38751(53.558%)	25073(55.195%)	12282(56.185%)	5090(57.47%)	2876(58.44%)
	$\Delta E > 50\%$	4305(6.432%)	5084(7.027%)	3585(7.892%)	1841(8.422%)	858(9.69%)	533(10.8%)
$Both$	$\Delta E > 10\%$	18019(26.923%)	19702(27.230%)	13045(28.717%)	6407(29.31%)	2742(30.96%)	1627(33.06%)
	$\Delta E > 50\%$	216(0.323%)	241(0.333%)	193(0.425%)	112(0.512%)	55(0.621%)	35(0.71%)

Table E.9: Count of Breit-Wigner tt (mixed flank) events where the bottom, anti-bottom, and both b-quarks have percent differences in energy between the particle and the assigned jet exceeding 10% as well as 50%. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events: 85835	5-Jet Events: 99511	6-Jet Events: 66945	7-Jet Events: 34439	8-Jet Events: 14679	9+ Jet Events: 8570
b	$\Delta E > 10\%$	46602(54.293%)	55352(55.624%)	37882(56.587%)	19721(57.264%)	8572(58.40%)	5210(60.79%)
	$\Delta E > 50\%$	6172(7.191%)	8268(8.309%)	6198(9.258%)	3470(10.08%)	1593(10.85%)	1023(11.94%)
$\bar{b}$	$\Delta E > 10\%$	46326(53.971%)	54336(54.603%)	37335(55.770%)	19599(56.909%)	8484(57.80%)	5077(59.24%)
	$\Delta E > 50\%$	5957(6.940%)	7547(7.584%)	5578(8.332%)	3086(8.961%)	1470(10.01%)	951(11.1%)
$Both$	$\Delta E > 10\%$	25334(29.515%)	30238(30.387%)	21142(31.581%)	11314(32.852%)	4951(33.73%)	3105(36.23%)
	$\Delta E > 50\%$	525(0.612%)	696(0.699%)	564(0.842%)	346(1.00%)	198(1.35%)	112(1.31%)

Table E.10: Count of Breit-Wigner tH (mixed flank) events where the bottom, anti-bottom, and both b-quarks have percent differences in energy between the particle and the assigned jet exceeding 10% as well as 50%. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

Appendix F

Overview of Events with Large Energy Differences for Light Quarks(10%, 50%), Large Angular Separation $(\Delta R > 0.5)$ between the Original Particle and the Reconstructed Jet

Note: The consideration of events with one quark meeting the selection criteria versus events with both quarks meeting them are disjoint — if both quarks are found to meet the selection criteria, the event is removed from the "single quark" category and added only to the "both quarks" category.

		4-Jet Events: 132133	5-Jet Events: 137995	6-Jet Events: 84495	7-Jet Events: 39536	8-Jet Events: 15515	9+ Jet Events: 8680
One q	$\Delta R > 0.5$	66574(53.753%)	45041(49.406%)	18164(47.008%)	5631(44.85%)	1468(42.84%)	488(43.8%)
	$\Delta E > 10\%$	62067(46.973%)	59754(43.302%)	33712(39.898%)	14892(37.667%)	5464(35.22%)	2801(32.27%)
	$\Delta E > 50\%$	30787(23.300%)	32276(23.389%)	21186(25.074%)	10905(27.582%)	4713(30.38%)	2883(33.21%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	28947(21.907%)	26513(19.213%)	14575(17.250%)	6284(15.89%)	2421(15.60%)	1203(13.86%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	24701(18.694%)	21994(15.938%)	12238(14.484%)	5498(13.91%)	2120(13.66%)	1101(12.68%)
Both q	$\Delta R > 0.5$	5407(4.092%)	17870(12.950%)	18018(21.324%)	11449(28.958%)	5659(36.47%)	3985(45.91%
	$\Delta E > 10\%$	45120(34.147%)	53384(38.685%)	36260(42.914%)	18509(46.816%)	7871(50.73%)	4915(56.62%)
	$\Delta E > 50\%$	1805(1.366%)	5632(4.081%)	5437(6.435%)	3340(8.448%)	1553(10.01%)	1126(12.97%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	3865(2.925%)	13508(9.7888%)	13788(16.318%)	8737(22.10%)	4349(28.03%)	3094(35.65%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	834(0.631%)	4275(3.098%)	4557(5.393%)	2856(7.224%)	1368(8.817%)	1011(11.65%)

Table F.1: Count of tt events with either one or both light quarks where the angular separation between the particle and the assigned jet is in excess of 0.5, the percent differences in energy between the two exceed 10% as well as 50%, and where both these angular separations and the energy differences exceed these limits. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events:	5-Jet Events:	6-Jet Events:	7-Jet Events:	8-Jet Events:	9+ Jet Events:
One q	$\Delta R > 0.5$	48633(43.406%)	61954(51.947%)	38663(51.283%)	17120(47.352%)	6166(42.48%)	3146(38.17%)
	$\Delta E > 10\%$	52456(46.818%)	52784(44.258%)	30671(40.682%)	13569(37.530%)	5074(34.96%)	2515(30.51%)
	$\Delta E > 50\%$	23577(21.043%)	29837(25.018%)	21156(28.061%)	10798(29.866%)	4422(30.47%)	2633(31.94%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	24115(21.523%)	30439(25.523%)	18467(24.495%)	8058(22.29%)	2853(19.66%)	1331(16.15%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	18890(16.860%)	21708(18.202%)	13059(17.321%)	5678(15.70%)	2035(14.02%)	1069(12.97%)
Both q	$\Delta R > 0.5$	4555(4.065%)	15388(12.903%)	18472(24.501%)	12543(34.692%)	6252(43.07%)	4235(51.38%)
	$\Delta E > 10\%$	37858(33.789%)	48699(40.833%)	35686(47.334%)	19221(53.163%)	8282(57.06%)	5199(63.07%)
	$\Delta E > 50\%$	1423(1.270%)	4908(4.115%)	5730(7.600%)	4066(11.25%)	2035(14.02%)	1483(17.99%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	3183(2.841%)	11560(9.6929%)	13873(18.401%)	9618(26.60%)	4784(32.96%)	3320(40.28%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	637(0.569%)	3583(3.004%)	4671(6.196%)	3452(9.548%)	1779(12.26%)	1337(16.22%)

Table F.2: Count of tH events with either one or both light quarks where the angular separation between the particle and the assigned jet is in excess of 0.5, the percent differences in energy between the two exceed 10% as well as 50%, and where both these angular separations and the energy differences exceed these limits. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events:	5-Jet Events:	6-Jet Events:	7-Jet Events:	8-Jet Events:	9+ Jet Events:
One q	$\Delta R > 0.5$	66574(53.753%)	45041(49.406%)	18164(47.008%)	5631(44.85%)	1468(42.84%)	488(43.8%)
	$\Delta E > 10\%$	57681(46.573%)	40938(44.905%)	16833(43.564%)	5532(44.06%)	1466(42.78%)	462(41.4%)
	$\Delta E > 50\%$	37307(30.122%)	22960(25.185%)	8984(23.25%)	2801(22.31%)	758(22.1%)	270(24.2%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	34054(27.496%)	22420(24.593%)	8779(22.72%)	2828(22.52%)	712(20.8%)	226(20.3%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	30108(24.310%)	17737(19.456%)	6587(17.05%)	2024(16.12%)	504(14.7%)	179(16.1%)
Both q	$\Delta R > 0.5$	11087(8.9518%)	9227(10.12%)	4559(11.80%)	1618(12.89%)	540(15.8%)	182(16.3%)
	$\Delta E > 10\%$	46884(37.855%)	33971(37.263%)	14562(37.686%)	4602(36.65%)	1300(37.93%)	462(41.4%)
	$\Delta E > 50\%$	5550(4.481%)	3892(4.269%)	1688(4.369%)	487(3.88%)	158(4.61%)	54(4.8%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	8732(7.050%)	7136(7.828%)	3413(8.833%)	1183(9.423%)	364(10.6%)	141(12.6%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	3689(2.979%)	2703(2.965%)	1159(2.999%)	364(2.90%)	113(3.30%)	41(3.7%)

Table F.3: Count of HH events with either one or both light quarks where the angular separation between the particle and the assigned jet is in excess of 0.5, the percent differences in energy between the two exceed 10% as well as 50%, and where both these angular separations and the energy differences exceed these limits. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events:	5-Jet Events:	6-Jet Events:	7-Jet Events:	8-Jet Events:	9+ Jet Events:
One q	$\Delta R > 0.5$	62679(52.647%)	44009(48.686%)	17912(46.287%)	5739(45.432%)	1392(40.36%)	474(42.5%)
	$\Delta E > 10\%$	55079(46.263%)	39225(43.393%)	15909(41.111%)	5097(40.35%)	1292(37.46%)	386(34.6%)
	$\Delta E > 50\%$	35929(30.178%)	25648(28.374%)	11164(28.849%)	3890(30.79%)	1106(32.07%)	385(34.5%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	32117(26.977%)	21928(24.258%)	8736(22.57%)	2866(22.69%)	691(20.0%)	217(19.4%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	28667(24.079%)	18297(20.241%)	7053(18.23%)	2215(17.53%)	538(15.6%)	181(16.2%)
Both q	$\Delta R > 0.5$	11599(9.7426%)	14563(16.111%)	8987(23.22%)	3660(28.97%)	1274(36.94%)	469(42.0%)
	$\Delta E > 10\%$	45366(38.105%)	36682(40.580%)	16837(43.509%)	5760(45.60%)	1677(48.62%)	604(54.1%)
	$\Delta E > 50\%$	5780(4.855%)	5428(6.005%)	2831(7.316%)	993(7.86%)	329(9.54%)	116(10.4%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	9212(7.738%)	11424(12.638%)	6838(17.67%)	2733(21.64%)	952(27.6%)	360(32.3%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	4018(3.375%)	4299(4.756%)	2343(6.055%)	863(6.83%)	295(8.55%)	101(9.05%)

Table F.4: Count of Ht events with either one or both light quarks where the angular separation between the particle and the assigned jet is in excess of 0.5, the percent differences in energy between the two exceed 10% as well as 50%, and where both these angular separations and the energy differences exceed these limits. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events:	5-Jet Events:	6-Jet Events:	7-Jet Events:	8-Jet Events:	9+ Jet Events:
One q	$\Delta R > 0.5$	9779(28.21%)	11872(35.540%)	7289(39.41%)	3058(40.32%)	1170(42.79%)	511(42.9%)
	$\Delta E > 10\%$	15287(44.103%)	14644(43.838%)	7978(43.14%)	3167(41.75%)	1102(40.31%)	485(40.7%)
	$\Delta E > 50\%$	3789(10.93%)	5116(15.32%)	3419(18.49%)	1710(22.54%)	669(24.5%)	306(25.7%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	4486(12.94%)	5561(16.65%)	3425(18.52%)	1464(19.30%)	539(19.7%)	229(19.2%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	2781(8.023%)	3630(10.87%)	2362(12.77%)	1120(14.77%)	435(15.9%)	173(14.5%)
Both q	$\Delta R > 0.5$	1075(3.101%)	2362(7.071%)	2085(11.27%)	1215 (16.02%)	520(19.0%)	320(26.8%)
	$\Delta E > 10\%$	10419(30.059%)	11165(33.423%)	6746(36.47%)	2962(39.05%)	1161(42.47%)	553(46.4%)
	$\Delta E > 50\%$	349(1.01%)	674(2.02%)	561(3.03%)	314(4.14%)	162(5.93%)	86(7.2%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	700(2.02%)	1676(5.017%)	1464(7.916%)	884(11.7%)	385(14.1%)	230(19.3%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	127(0.366%)	429(1.28%)	391(2.11%)	228(3.01%)	119(4.35%)	70(5.9%)

Table F.5: Count of Breit-Wigner tt ($m_t, m_{\bar{t}} < 150$ GeV) events with either one or both light quarks where the angular separation between the particle and the assigned jet is in excess of 0.5, the percent differences in energy between the two exceed 10% as well as 50%, and where both these angular separations and the energy differences exceed these limits. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events:	5-Jet Events:	6-Jet Events:	7-Jet Events:	8-Jet Events:	9+ Jet Events:
One q	$\Delta R > 0.5$	33864(49.519%)	34274(53.048%)	18841(50.779%)	7640(46.36%)	2841(43.67%)	1265(38.24%)
	$\Delta E > 10\%$	31523(46.096%)	28303(43.807%)	15092(40.675%)	6090(36.95%)	2251(34.60%)	1070(32.35%)
	$\Delta E > 50\%$	17347(25.366%)	17353(26.858%)	10438(26.858%)	4825(29.28%)	1944(29.88%)	1014(30.65%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	16541(24.188%)	16635(25.747%)	8976(24.19%)	3588(21.77%)	1299(19.97%)	556(16.8%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	13884(20.302%)	12339(19.098%)	6306(17.00%)	2474(15.01%)	902(13.9%)	405(12.2%)
Both q	$\Delta R > 0.5$	4635(6.778%)	10008(15.490%)	9628(25.95%)	5843(35.46%)	2746(42.21%)	1685(50.94%)
	$\Delta E > 10\%$	25367(37.094%)	27338(42.313%)	17801(47.976%)	8716(52.89%)	3703(56.92%)	2009(60.73%)
	$\Delta E > 50\%$	1448(2.117%)	2885(4.465%)	2741(7.387%)	1763(10.70%)	889(13.7%)	548(16.6%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	3355(4.906%)	7430(11.50%)	7179(19.35%)	4516(27.40%)	2102(32.31%)	1292(39.06%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	892(1.30%)	2221(3.438%)	2261(6.094%)	1496(9.078%)	770(11.8%)	500(15.1%)

Table F.6: Count of Breit-Wigner tH ($m_t, m_{\bar{t}} < 150$ GeV) events with either one or both light quarks where the angular separation between the particle and the assigned jet is in excess of 0.5, the percent differences in energy between the two exceed 10% as well as 50%, and where both these angular separations and the energy differences exceed these limits. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events:	5-Jet Events:	6-Jet Events:	7-Jet Events:	8-Jet Events:	9+ Jet Events:
One q	$\Delta R > 0.5$	43124(40.698%)	47695(37.965%)	30361(35.028%)	14963(33.250%)	6358(31.54%)	3621(28.32%)
	$\Delta E > 10\%$	50897(48.034%)	55478(44.160%)	35456(40.906%)	17035(37.854%)	7243(35.93%)	4159(32.52%)
	$\Delta E > 50\%$	24336(22.967%)	30180(24.023%)	22379(25.819%)	12818(28.483%)	6247(30.99%)	4485(35.07%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	22387(21.128%)	23575(18.765%)	14740(17.006%)	6982(15.51%)	3005(14.91%)	1595(12.47%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	19625(18.521%)	20563(16.368%)	12488(14.408%)	5976(13.28%)	2501(12.41%)	1501(11.74%)
Both q	$\Delta R > 0.5$	3241(3.059%)	14710(11.709%)	18494(21.337%)	13956(31.012%)	7821(38.80%)	6392(49.98%)
	$\Delta E > 10\%$	33630(31.738%)	46673(37.151%)	36559(42.179%)	21312(47.358%)	10395(51.570%)	7303(57.11%)
	$\Delta E > 50\%$	1382(1.304%)	5123(4.078%)	6144(7.088%)	4442(9.871%)	2354(11.68%)	1858(14.53%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	2313(2.183%)	11451(9.1149%)	14551(16.788%)	10883(24.183%)	6088(30.20%)	4885(38.20%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	607(0.573%)	3924(3.123%)	5240(6.046%)	3952(8.782%)	2108(10.46%)	1722(13.47%)

Table F.7: Count of Breit-Wigner tt ($m_t, m_{\bar{t}} > 195$ GeV) events with either one or both light quarks where the angular separation between the particle and the assigned jet is in excess of 0.5, the percent differences in energy between the two exceed 10% as well as 50%, and where both these angular separations and the energy differences exceed these limits. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events:	5-Jet Events:	6-Jet Events:	7-Jet Events:	8-Jet Events:	9+ Jet Events:
One q	$\Delta R > 0.5$	38031(39.728%)	67077(51.705%)	49732(50.462%)	25083(45.866%)	10563(41.029%)	5782(34.95%)
	$\Delta E > 10\%$	45167(47.183%)	56826(43.803%)	39139(39.713%)	19713(36.047%)	8393(32.60%)	4814(29.10%)
	$\Delta E > 50\%$	18736(19.572%)	34705(26.752%)	29857(30.295%)	17458(31.923%)	8409(32.66%)	5565(33.64%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	19337(20.200%)	32894(25.356%)	23853(24.203%)	11680(21.358%)	4724(18.35%)	2479(14.99%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	14886(15.550%)	25143(19.381%)	18300(18.569%)	9100(16.64%)	3767(14.63%)	2143(12.95%)
Both q	$\Delta R > 0.5$	3180(3.322%)	17146(13.217%)	25777(26.155%)	20402(37.307%)	11821(45.916%)	9207(55.65%)
	$\Delta E > 10\%$	30728(32.099%)	53804(41.474%)	48476(49.187%)	30166(55.161%)	15464(60.066%)	10784(65.19%)
	$\Delta E > 50\%$	1160(1.212%)	6076(4.684%)	9103(9.237%)	7438(13.60%)	4351(16.90%)	3559(21.51%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	2203(2.301%)	13136(10.126%)	20068(20.362%)	15956(29.177%)	9325(36.22%)	7324(44.27%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	503(0.525%)	4390(3.384%)	7537(7.648%)	6448(11.79%)	3849(14.95%)	3221(19.47%)

Table F.8: Count of Breit-Wigner tH ($m_t, m_{\bar{t}} > 195$ GeV) events with either one or both light quarks where the angular separation between the particle and the assigned jet is in excess of 0.5, the percent differences in energy between the two exceed 10% as well as 50%, and where both these angular separations and the energy differences exceed these limits. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events:	5-Jet Events:	6-Jet Events:	7-Jet Events:	8-Jet Events:	9+ Jet Events:
One q	$\Delta R > 0.5$	23472(35.070%)	26396(36.482%)	16654(36.662%)	7906(36.17%)	3154(35.61%)	1657(33.67%)
	$\Delta E > 10\%$	31030(46.363%)	32116(44.388%)	19088(42.020%)	8739(39.98%)	3325(37.54%)	1698(34.51%)
	$\Delta E > 50\%$	12170(18.183%)	14973(20.694%)	10460(23.026%)	5646(25.83%)	2546(28.75%)	1628(33.08%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	11723(17.516%)	12897(17.825%)	8033(17.68%)	3642(16.66%)	1464(16.53%)	765(15.5%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	9549(14.27%)	10327(14.273%)	6368(14.02%)	2977(13.62%)	1208(13.64%)	674(13.7%)
Both q	$\Delta R > 0.5$	1944(2.905%)	6760(9.343%)	7552(16.62%)	5315(24.31%)	2858(32.27%)	2049(41.64%)
	$\Delta E > 10\%$	20692(30.916%)	25427(35.143%)	17856(39.308%)	9602(43.92%)	4343(49.03%)	2662(54.09%)
	$\Delta E > 50\%$	808(1.21%)	2213(3.059%)	2298(5.059%)	1565(7.159%)	826(9.33%)	596(12.1%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	1368(2.044%)	5131(7.092%)	5803(12.77%)	4043(18.49%)	2203(24.87%)	1586(32.23%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	318(0.475%)	1594(2.203%)	1899(4.180%)	1340(6.130%)	720(8.13%)	537(10.9%)

Table F.9: Count of Breit-Wigner tt (mixed-flank) events with either one or both light quarks where the angular separation between the particle and the assigned jet is in excess of 0.5, the percent differences in energy between the two exceed 10% as well as 50%, and where both these angular separations and the energy differences exceed these limits. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

		4-Jet Events:	5-Jet Events:	6-Jet Events:	7-Jet Events:	8-Jet Events:	9+ Jet Events:
One q	$\Delta R > 0.5$	37393(43.564%)	51827(52.082%)	33823(50.524%)	15988(46.424%)	5996(40.85%)	3190(37.22%)
	$\Delta E > 10\%$	40008(46.610%)	43502(43.717%)	26725(39.921%)	12555(36.456%)	4981(33.93%)	2614(30.50%)
	$\Delta E > 50\%$	18766(21.863%)	26707(26.838%)	19495(29.121%)	10588(30.744%)	4632(31.56%)	2809(32.78%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	18692(21.777%)	25247(25.371%)	16086(24.029%)	7357(21.36%)	2788(18.99%)	1364(15.92%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	14926(17.389%)	19138(19.232%)	11820(17.656%)	5524(16.04%)	2073(14.12%)	1107(12.92%)
Both q	$\Delta R > 0.5$	4185(4.876%)	14129(14.198%)	17197(25.688%)	12380(35.948%)	6632(45.18%)	4524(52.79%)
	$\Delta E > 10\%$	29376(34.224%)	41680(41.885%)	32527(48.588%)	18776(54.520%)	8610(54.52%)	5399(63.00%)
	$\Delta E > 50\%$	1454(1.694%)	4714(4.737%)	5601(8.367%)	4109(11.93%)	2321(15.81%)	1615(18.84%)
	$\Delta R > 0.5$, $\Delta E > 10\%$	2984(3.476%)	10701(10.754%)	13135(19.621%)	9542(27.71%)	5184(35.32%)	3535(41.25%)
	$\Delta R > 0.5$, $\Delta E > 50\%$	754(0.878%)	3479(3.496%)	4636(6.925%)	3532(10.26%)	2071(14.11%)	1458(17.01%)

Table F.10: Count of Breit-Wigner tH (mixed flank) events with either one or both light quarks where the angular separation between the particle and the assigned jet is in excess of 0.5, the percent differences in energy between the two exceed 10% as well as 50%, and where both these angular separations and the energy differences exceed these limits. The results are broken down by the number of valid jets counted for each particular event, and the percentage of n-jet events that the event count represents is given after in parentheses.

Bibliography

[1] S. Navas et al. (Particle Data Group). Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024 and 2025 Update. doi:10.1103/PhysRevD.110.030001.

[2] Biagio Di Micco, Maxime Gouzevitch, Javier Mazzitelli, and Caterina Vernieri. Higgs boson potential at colliders: Status and perspectives. *Reviews in Physics*, 5:100045, November 2020. doi:10.1016/j.revip.2020.100045.

[3] CMS Collaboration. Combination of searches for nonresonant Higgs boson pair production in proton-proton collisions at $\sqrt{s}$= 13 TeV, 2025. doi:10.48550/ARXIV.2510.07527.

[4] G. Aad et al. Combination of Searches for Higgs Boson Pair Production in pp Collisions at s=13 TeV with the ATLAS detector. *Physical Review Letters*, 133(10):101801, 2024. doi:10.1103/physrevlett.133.101801.

[5] P. Artoisenet, V. Lemaître, F. Maltoni, and O. Mattelaer. Automation of the matrix element reweighting method. 2010. doi:10.48550/ARXIV.1007.3300.

[6] Mary K. Gaillard, Paul D. Grannis, and Frank J. Sciulli. The Standard Model of Particle Physics. 1998. doi:10.48550/ARXIV.HEP-PH/9812285.

[7] ATLAS Collaboration. Study of Higgs boson pair production in the $HH \to b\bar{b}\gamma\gamma$ final state with 308 fb^{-1} of data collected at $\sqrt{s} = 13$ TeV and 13.6 TeV by the ATLAS experiment, 2025. doi:10.48550/ARXIV.2507.03495.

[8] Universität Zürich. URL: https://wiki.physik.uzh.ch/cms/latex:feynman.

[9] Michael Edward Peskin. *An introduction to quantum field theory*. The advanced book program. CRC Press, Taylor Francis Group, Boca Raton, first issued in paperback edition, 2019. Includes bibliographical references (pages 811-815) and index.

[10] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 Physics and Manual. 2006. doi:10.48550/ARXIV.HEP-PH/0603175.

[11] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet user manual. 2011. doi:10.48550/ARXIV.1111.6097.

[12] ATLAS Collaboration. Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. 2018. doi:10.48550/ARXIV.1805.01845.

[13] ATLAS Collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV. 2019. doi:10.48550/ARXIV.1902.04655.

[14] ATLAS Collaboration. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV. 2020. doi:10.48550/ARXIV.2012.00578.

[15] ATLAS Collaboration. Performance of b-Jet Identification in the ATLAS Experiment. 2015. doi:10.48550/ARXIV.1512.01094.

[16] The ATLAS Collaboration. Combined measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ decay channels with the atlas detector using $\sqrt{s} = 7$, 8 and 13 TeV pp collision data. 2023. doi:10.48550/ARXIV.2308.04775.

[17] ROOT Team. URL: https://root.cern.ch/doc/v632/classTLorentzVector.html.

[18] ROOT Team. URL: https://root.cern.ch/doc/v636/group__fitpanel.html.

[19] Hongda Guo. Determination of Top-Quark Mass and Studies of Cross-Section in the Semileptonic $t\bar{t}$ Decay Channel Using the Matrix Element Method. Bachelor's thesis, Ludwig-Maximilians-Univeristät München, November 2025.

[20] Petra Haefner. *Measurement of the Top Quark Mass with the Matrix Element Method in the Semileptonic Decay Channel at D0.* Phd thesis, Ludwig-Maximilians-Universität München, July 2008.

[21] Volker Blobel. An Unfolding Method for High Energy Physics Experiments, 2002. doi:10.48550/ARXIV.HEP-EX/0208022.

[22] Lydia Brenner, Pim Verschuuren, Rahul Balasubramanian, Carsten Burgard, Vincent Croft, Glen Cowan, and Wouter Verkerke. Comparison of unfolding methods using RooFitUnfold, 2019. doi:10.48550/ARXIV.1910.14654.

Declaration of Independent Work

I hereby declare that this thesis with the title

Investigating the Use of the Matrix-Element-Method in Reducing the Top-Antitop Quark Background to Higgs Pair Production Processes

is my own work, and that I have not used any sources and aids other than those stated in the thesis.

Patrick Rudolph Schumacher

München, on 13. July 2026