Studies for the measurement of the jet mass in fully merged hadronic top quark decays

Studien zur Messung der Jetmasse in vollständig überlagerten Top-Quark Zerfällen

von

Torben Dreyer

geboren am

14. März 1989

Master-Arbeit im Studiengang Physik

Universität Hamburg

2015

1. Gutachter: Prof. Dr. Johannes Haller
2. Gutachter: Prof. Dr. Peter Schleper

Abstract

This thesis presents studies for a measurement of the invariant mass of jets containing all decay products of a fully hadronic top quark decay given by the decay channel $t \to Wb \to qq'b$. At high transverse momentum p_T of the top quark, the decay products are collimated and can be reconstructed in one jet. The invariant mass of these jets is a fundamental jet substructure variable, used in top tagging algorithms to distinguish between jets from top quark decays and jets from light quarks or gluons. Jet substructure is well suited to test the modeling of parton showers and hadronization in Monte Carlo generators.

In the special case of jets from very collimated top quark decays the invariant mass of jets is sensitive to the mass of the top quark itself. Moreover, this jet mass is in principle calculable from first principle in an effective field theory, providing the opportunity to compare the measurement directly to those calculations. This could lead to a measurement of the top quark mass in a well defined renormalization scheme.

The main result of this thesis is a preliminary measurement of the jet mass of the leading jet in collimated top quark decays for $p_T > 400\,\text{GeV}$. The measurement is corrected to the level of stable particles. Corrections for detector effects are applied using a regularized unfolding. A second measurement has been done for jets with $p_T > 500\,\text{GeV}$, where jets with $400\,\text{GeV} < p_T < 500\,\text{GeV}$ are taken into account in the migration matrix of the unfolding. This measurement shows a reduced dependence on the simulation used to calculate the migration matrix.

Zusammenfassung

Diese Arbeit befasst sich mit Studien zur Messung der invariantem Masse von Jets, welche alle Zerfallsprodukte eines hadronischen Top-Quark-Zerfalls ($t \rightarrow Wb \rightarrow qq'b$) beinhalten. Für einen großen transversalen Impuls p_T des Top-Quarks sind dessen Zerfallsprodukte stark kollimiert und lassen sich in einem einzigen Jet rekonstruieren. Bei der invarianten Masse von Jets handelt es um eine fundamentale Jet-Substrukturvariable, welche unter anderem für Methoden wie das Top-Tagging verwendet wird, um Jets aus Top-Quark Zerfällen von Jets leichter Quarks oder Gluonen zu unterscheiden. Darüber hinaus ist Jet-Substruktur gut geeignet um Parton-Schauer- und Hadronisierungsmodelle in Monte-Carlo-Generatoren zu validieren.

Im Spezialfall eines Jets aus stark kollimierten Top-Quarks Zerfällen ist die Massenverteilung dieser Jets sensitiv auf die Masse des Top-Quarks selbst. Darüber hinaus ist die Jet-Masse berechenbar in einer effektiven Feldtheorie, wodurch ein Vergleich dieser Messung mit solchen Rechnungen möglich ist. Das könnte eine Bestimmung der Top-Quark Masse in einem wohldefinierten Renormalisierungsschema erlauben.

Das zentrale Ergebnis dieser Arbeit ist eine vorläufige Messung der Jet Masse in vollständig überlagerten Top-Quark Zerfällen für Jets mit $p_T > 400\,\mathrm{GeV}$. Die Messung wurde auf Teilchen-Ebene korrigiert, wobei die Detektoreffekte mit Hilfe einer regularisierten Entfaltung korrigiert wurden. In einer zweiten Entfaltung wurden Jets mit $p_T > 500\,\mathrm{GeV}$ gemessen, wobei Jets mit $400\,\mathrm{GeV} < p_T < 500\,\mathrm{GeV}$ in der Entfaltung berücksichtigt wurden. Diese Messung zeigt einen geringeren Einfluss des verwendeten Modells für die Berechnung der Migrationsmatrix, allerdings sind die statistischen Unsicherheiten größer.

Contents

1. Introduction

Substructure methods have become an imptortant tool in high energy physics. They give the opportunity to access information on processes that can not be resolved with standard reconstruction methods. Of special interest for many analysis is the substructure of jets. Jet substructure methods allow for example to study the distribution of energy within the jet or to separate contributions of different particle decays to one jet.

With increasing center of mass energy the jet substructure gets more and more important for a good description of jets and physics processes with high energy. In the last run of the LHC at a center of mass energy of $\sqrt{s} = 8\,\text{TeV}$, jet substructure has already been successfully used by the CMS collaboration and has proven to be a very successful method to gain more information on processes with very high momenta [1–4]. An important tool in this context and a good example for the use of jet substructure is top tagging, which has been studied in CMS extensively [5]. Being a very useful tool at $\sqrt{s} = 8\,\text{TeV}$ top tagging, as well as substructure in general, will be much more important in the next LHC runs with $\sqrt{s} = 13 - 14\,\text{TeV}$. Jet substructure is well suited to test the modeling of jets including hadronization models by measurements of jet substructure observables. A good description of jets by Monte Carlo methods is very important for many new physics searches, comparing Monte Carlo signal to real data.

Beside using jet substructure for searches for new physics and identification of boosted objects, it can be further used to perform standard model measurements. This thesis aims on a measurement of a basic jet substructure variable, the jet mass, which is, in case of jets containing a full hadronic top quark decay, connected to the mass of the top quark. Because of its high mass and the resulting large Yukava coupling to the Higgs field, the top quark has a direct impact on the Higgs sector of the standard model. Therefore the mass of the top quark is an important input parameter for consistency checks in the electroweak sector.

The most precise measurements measure the top quark mass by reconstruction of its decay products. A world combination in 2014 resulted in a top quark mass of $173.20 \pm 0.76\,\text{GeV}$ [6]. The calculation of the top quark mass from separated decay leads to a measurement of the top quark mass parameter in the Monte Carlo generator. This mass parameter, however, can not be compared unambiguously to the top quark mass parameter in the lagrangian density of the standard model in a well defined renormalization scheme, because it always includes effects like color reconnections, which can not be calculated perturbatively and have to be estimated using tuned models. For this reason this mass is often referred to as the Monte Carlo mass of the top quark.

A measurement of the jet mass, however, could lead to a determination of the top quark mass much less influenced by theoretical ambiguities, because the jet mass of a jet containing an hadronic top quark decay is in principle calculable from first principle in the framework of an effective field theory. This calculation has been done for lepton lepton collisions [7] and is planned also for hadron collisions. This provides the opportunity to compare a measurement to theoretical calculation from first principle, instead of Monte Carlo generated distributions and provides in this way a complementary determination of the top quark mass.

In this thesis studies are done to perform a measurement of the mass of jets containing all decay products of an hadronic top quark decay at the LHC. The measurement is corrected for detector effects and (in)efficiencies to particle level, to be able to compare it directly to theoretical calculations. The correction is performed using a regularized unfolding with the TUnfold framework [8]. The detector effects and reconstruction efficiencies are calculated from Monte Carlo simulation. Uncertainties arising from the choice of the Monte Carlo generator and the top quark mass parameter in the generator are evaluated. Finally a measurement in CMS data is performed using the 2012 dataset with a center of mass energy of $\sqrt{s} = 8\,\mathrm{TeV}$ and an integrated luminosity of $19.7\,\mathrm{fb}^{-1}$. The jet mass distribution is measured in two kinematic regions. The main measurement will be done for a jet $p_T > 400\,\mathrm{GeV}$. At the same time a second measurement of the jet mass for $p_T > 500\,\mathrm{GeV}$ together with events with $400\,\mathrm{GeV} < p_T < 500\,\mathrm{GeV}$ is performed to study the influence of a sideband region with lower p_T on the systematic uncertainties.

This thesis is organized in the following way. Chapter 2 starts with a short introduction of theoretical aspects about the standard model of particle physics and the top quark. The following chapter gives short overview of the experimental set-up introducing the Large Hadron Collider (LHC) and the CMS detector. Chapter 4 describes the event reconstruction as performed in the CMS experiment as well as some analysis techniques. The unfolding method used for this measurement is described in more detail in chapter 5. Chapter 6 describes the analysis. It begins with the definition of the measurement on the level of stable particles, continues with the reconstruction of events in data and finishes with an unfolding of the data and a determination of systematic uncertainties. The final results of the measurement can be found in chapter 7.

2. Theory

2.1. Standard Model of Particle Physics

The standard model of particle physics [9] describes all fundamental particles and forces except for gravity. The particles can be divided into two basic groups, the fermions, carrying a spin of one half, and gauge bosons with integer spin, responsible for the mediation of the fundamental forces. Furthermore there is a scalar boson called the Higgs boson, needed for the electroweak symmetry breaking and to include particle masses in the theory. In 2012, a new boson with a mass of about 125 GeV was observed by the CMS and ATLAS collaborations at the LHC [10, 11]. All studies of its properties show a good agreement with the standard model Higgs boson so far.

The fermions can be further divided into quarks and leptons. In contrast to leptons, quarks carry color charge and can therefore interact via the strong interaction. Both quarks and leptons occur in three generations mainly differing in their mass.

The three fundamental interactions described by the standard model are the electromagnetic interaction, mediated by the photon, the weak interaction, mediated by the W and Z bosons, and the strong interaction, mediated by eight gluons. These interactions are discussed in more detail in the following. Figure 2.1 shows the particle content of the standard model.

The Standard Model as a Quantum Field Theory

The standard model is a renormalizable quantum field theory, invariant under the local gauge group $U(1)_Y \times SU(2)_L \times SU(3)_C$. It basically combines three quantum field theories invariant under the local gauge groups $U(1)_Y$, $SU(2)_L$ and $SU(3)_C$, one for each fundamental interaction, described in the following.

Quantum Electro Dynamics (QED)

QED describes the electromagnetic interaction between electrically charged particles. It is invariant under the local gauge group $U(1)_{em}$. The requirement of this local gauge invariance leads to one massless gauge boson, called photon. QED is one of the most precise verified theories in physics. Due to small coupling constants over a large energy range, perturbation theory works well and QED is able to provide precise predictions.

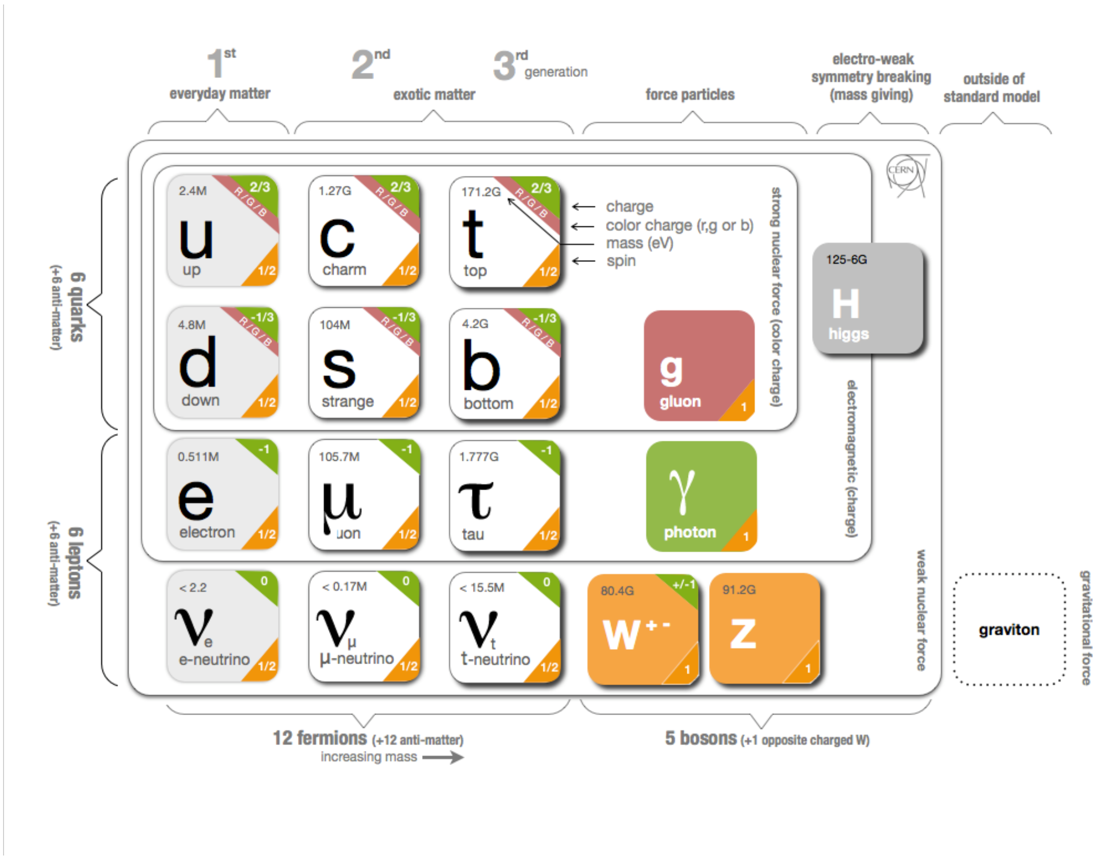

Figure 2.1.: The particle content of the standard model of particle physics, taken from reference [12].

Quantum Chromo Dynamics (QCD)

The strong interaction is described by QCD, a theory invariant under a $SU(3)$ group. It acts on fermions carrying one of three color charges. The invariance under $SU(3)$ leads to nine massless gauge bosons carrying color charge themselves. Only eight of them are contributing to the interaction since one is color neutral.

In case of QCD these gauge bosons are called gluons. The fact that gluons carry color charges themselves leads to the self-interaction between gluons and the confinement of quarks and gluons. This means the potential between two color charged particles increases approximately linearly with distance and therefore quarks and gluons cannot exist as free particles but only in color neutral multi-quark states called hadrons.

Weak Interaction

The weak interaction is obtained by a $SU(2)_L$ symmetry group. These transformations act only on left handed fermions (indicated by the "L"), leading to a coupling to the left handed isospin. Invariance under this group leads to three identical interacting mediator fields.

For the quark sector weak couplings between the different quark generations have been observed. This is implemented in the theory by a rotation of the down type isospin states of the quark fields. The left handed states coupling via the week interaction are then given by:

$$\begin{pmatrix} u \\ d' \end{pmatrix}_{\mathrm{L}} \quad \begin{pmatrix} c \\ s' \end{pmatrix}_{\mathrm{L}} \quad \begin{pmatrix} t \\ b' \end{pmatrix}_{\mathrm{L}} . \tag{2.1}$$

The rotation is described by a unitary matrix called the Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} . \tag{2.2}$$

Absolute values of the CKM matrix are [13] :

$$\mathbf{V}_{CKM} = \begin{pmatrix} 0.97427 \pm 0.00014 & 0.22536 \pm 0.00061 & 0.00355 \pm 0.00015 \\ 0.22522 \pm 0.00061 & 0.97343 \pm 0.00015 & 0.0414 \pm 0.0012 \\ 0.00886^{+0.00033}_{-0.00032} & 0.0405^{+0.0011}_{-0.0012} & 0.99914 \pm 0.00005 \end{pmatrix} . \tag{2.3}$$

The $SU(2)_L$ describes the interaction of the W bosons to the fermions, but not the coupling via the Z boson observed in [14]. A proper description is obtained by the unification of the weak and the electromagnetic interaction.

Electroweak Unification

The unification of the electromagnetic and the weak interaction can be realized by a definition of a new hypercharge Y inserted in the $U(1)$ symmetry group. This hypercharge is defined as:

$$Y = 2(Q - T_3), \tag{2.4}$$

with the electrical charge Q and the third component of the weak isospin T_3 (± 1 for the left handed isospin states).

Requiring local gauge invariance under both symmetry groups leads to four vector fields, three W^μ-fields from $SU(2)_\mathrm{L}$ and one B^μ from the $U(1)_Y$ group. A superposition of the first two W fields leads to the two physical W bosons. The photon and the Z boson fields A^μ and Z^μ can be obtained by a mixing of the W_3^μ with the B^μ field. This mixing is described by a rotation of these fields by the Weinberg angle θ_W, similar to the rotation in the weak interaction by the CKM matrix,

$$\begin{pmatrix} W_3^\mu \\ B^\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} Z^\mu \\ A^\mu \end{pmatrix}. \tag{2.5}$$

Spontaneous Symmetry Breaking

Electroweak unification with the resulting photon and the weak W and Z bosons requires local gauge invariance which leads to massless gauge bosons. However experimentally the W and Z bosons have been measured with masses of 80.4 GeV and 91.2 GeV [13].

The Solution is an introduction of a scalar field ϕ, the higgs field, and a symmetric potential with a finite vacuum expectation value, leading to spontaneous symmetry breaking. Including this field together with the potential in the electroweak theory leads to a new scalar boson, called the Higgs boson, mass terms for the W and Z bosons and provides a mechanism to include mass terms for all fermions in the standard model.

2.2. Top Quark

The top quark was first discovered in 1995 by the D0 [15] and CDF [16] collaborations at the Tevatron collider in proton anti-proton collisions, and has been studied extensively at the LHC and the Tevatron afterwards. The large cross section of $\sigma_{t\bar{t}} = 245.8^{+6.2+6.2}_{-8.4-6.4}$ pb for $\sqrt{s} = 8$ TeV [17] at the LHC leads to a large number of produced top quark pairs and precise measurements of its properties. With a mass around 175 GeV [6] the top quark is the heaviest particle of the standard model. Due to its high mass the top quark has a very small life time, smaller than the characteristic timescale for hadronization. This means the top quark decays before hadronization, making the top quark the only quark whose dynamics can be studied for a bare quark without being diffused by fragmentation effects.

2.2.1. Mass of the Top Quark

The mass of the top quark is a fundamental parameter of the standard model. It defines it's coupling strength to the higgs boson and has therefore a direct impact on the higgs sector and on the electroweak vacuum stability [18]. That makes it a very important input parameter for self-consistency checks of the standard model [19].

Definition and Measurement

Due to the importance of the top quark mass, a large effort has been made since its discovery to measure the mass of the top quark as precisely as possible. The most precise measurements of the top quark mass are direct measurements performed at the LHC and Tevatron, summarized in Fig. 2.2. The most precise single measurements have been done by CMS leading to a mass of $172.04 \pm 0.19(\text{stat.}) \pm 0.75(\text{syst.})$ GeV [20] and D0 with $m_{\text{top}} = 174.98 \pm 0.76$ GeV [21].

These measurements are obtained by taking the invariant mass of the top quark's products and have a very high experimental precision. However there is no unambiguous way to relate this measured top quark mass to the top quark mass in the lagrangian density of the standard model, using a well defined renormalization scheme, because a calculation of the top quark mass from the invariant mass of its decay products always includes processes that cannot be calculated perturbatively like color re-connections and hadronization (cf.e.g. [22]). Those effects are included in Monte Carlo event generators by phenomenological models. For this reason this mass definition is often referred to as the Monte Carlo mass $m_{\text{top}}^{\text{MC}}$. In reference [23] it is argued, that it is possible to relate the Monte Carlo mass to the pole mass of the top quark. However this is not possible from first principle and a theoretical uncertainty of the order of 1 GeV is assumed.

A possibility to suppress the influence of those effects, is a measurement of the jet mass of a jet containing all decay products of an hadronic top quark decay, proposed in

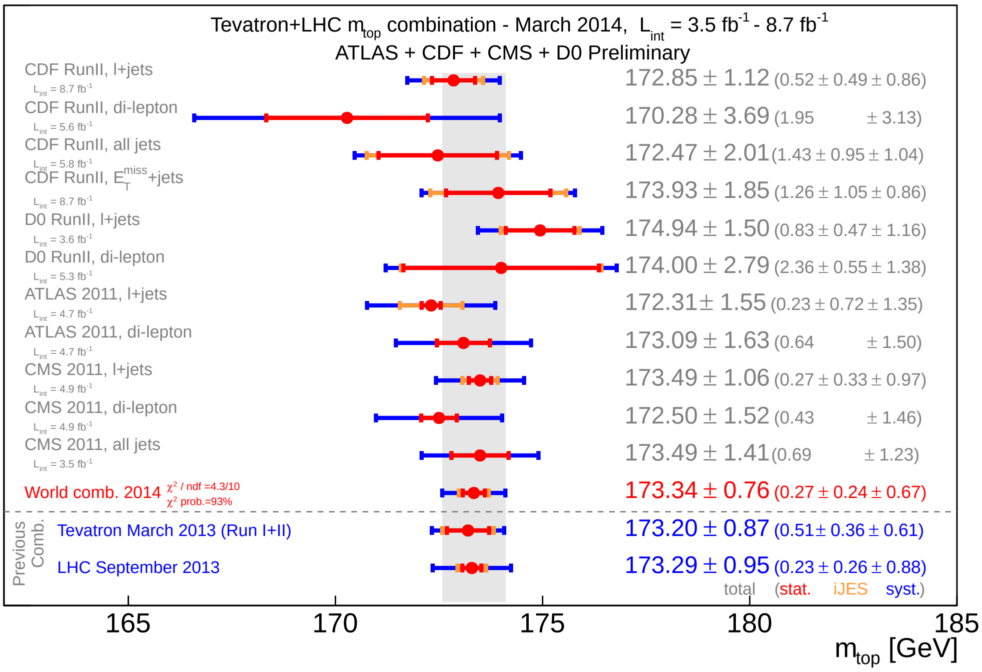

Figure 2.2.: Combination of top quark mass measurements at the LHC and Tevatron [6].

references [24, 25]. Leptonic top quark decays will not be studied in this thesis, because the neutrino from the W boson decay cannot be reconstructed directly, but only a missing transverse energy.

Calculations for the jet mass of jets containing all decay products of a top quark are done for lepton-lepton collisions [25]. These calculations will be extended for hadron collisions in the future. It is shown that the jet mass can be calculated in effective field theory and that the shape of the jet mass spectrum in the peak area is sensitive to the mass of the top quark. In this way a measurement of this jet mass spectrum could lead to a determination of the top quark mass in a theoretically well defined mass scheme. For lepton-lepton collisions it was verified, that this definition of the top quark mass can than be related to other important mass definitions like the $\overline{\mathrm{MS}}$-mass with theoretical uncertainties below Λ_{QDC} ($\approx 240\,\mathrm{MeV}$) [25].

2.2.2. Production and Decay

At the LHCTop quarks can be produced in top anti-top quark pairs via the strong interaction or as single top quarks by the weak interaction. The corresponding leading order Feynman diagrams for pair production are shown in Fig. 2.3.

Because of the strongly interacting initial state, top quarks are dominantly produced in pairs at the LHC. The dominant production channel is gluon-gluon fusion, because of a high density of gluons with low momentum in the proton, shown in Fig. 2.4, which

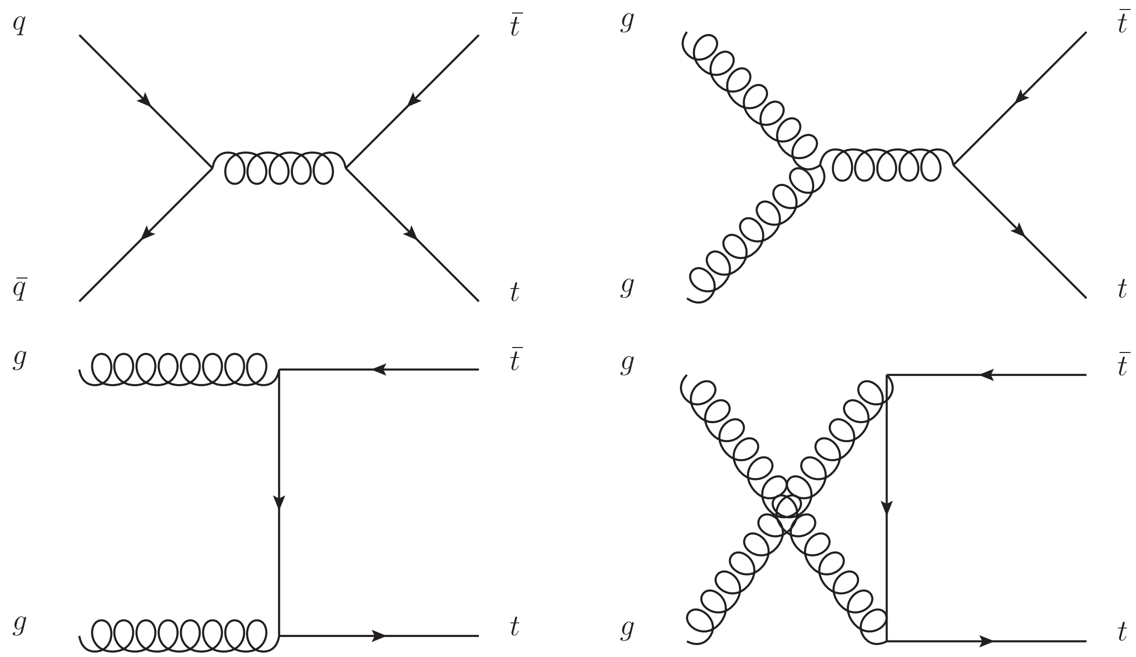

Figure 2.3.: Feynman diagrams for top anti-top quark production in leading order in hadron collisions. The top left diagram shows the production via quark anti-quark annihilation and the others show gluon-gluon fusion.

can contribute to the $t\bar{t}$-production due to the high momentum of the proton itself. The fraction of gluon-gluon fusion to the total $t\bar{t}$-production cross section is about 80 % for $\sqrt{s} = 7\,\mathrm{TeV}$ and 90 % for $\sqrt{s} = 14\,\mathrm{TeV}$ [13].

From the unitarity of the CKM-matrix, we know that the top quark decays with a probability of more than 99.9 % into a W boson and a b-quark. The W boson can further decay hadronically into two quarks, or leptonically into a charged lepton and a neutrino, leading to three different decay channels for a $t\bar{t}$-system, the all-hadronic channel with both W bosons decaying hadronically, the lepton+jets channel with one W boson decaying hadronically and one leptonically, and the dilepton channel with both W bosons decaying leptonically. The branching ratios are listed in Tab. 2.1. This analysis will concentrate on the lepton+jets channel visualized in Fig. 2.5.

decay channel	**branching ratio**
all-hadronic	45.7%
lepton+jets	43.8%
dilepton	10.5%

Table 2.1.: Branching ratios of the top quark decay channels [13].

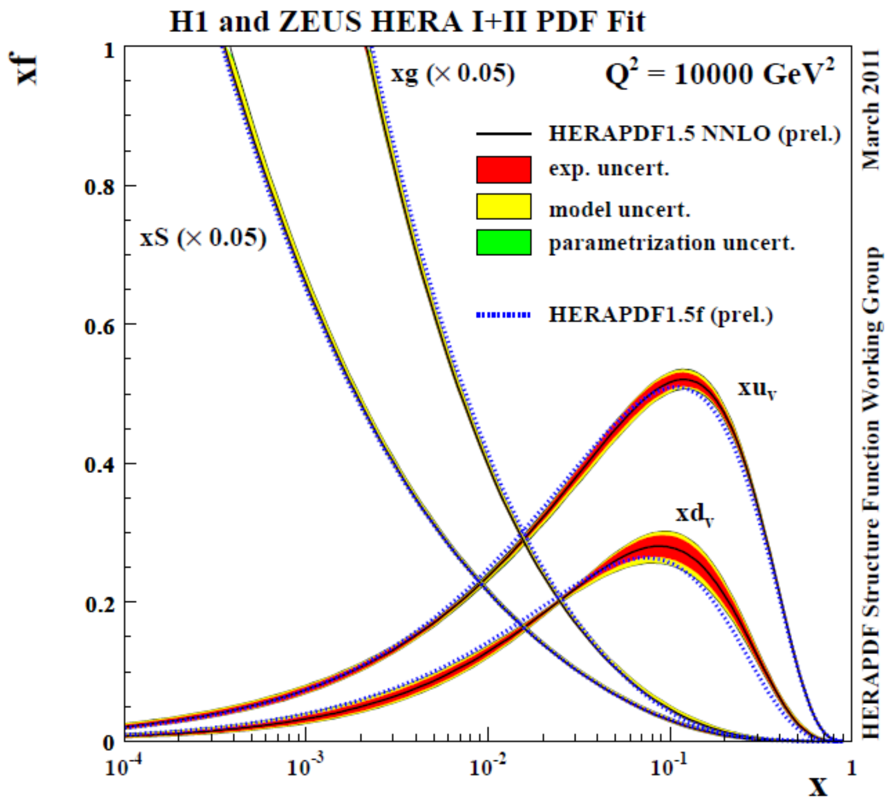

Figure 2.4.: Parton density functions (PDFs) in the proton from reference [26], measured by the H1 and ZEUS collaborations in electron-proton collisions at the HERA collider in Hamburg. On the x-axis is the fraction of the proton momentum and on the y-axis the parametrized PDFs: xg for the gluon distribution, xS for the sea-quark distribution and xu_v and xd_v for the valence quarks. The distributions of xg and xS are scaled by a factor of 1/20.

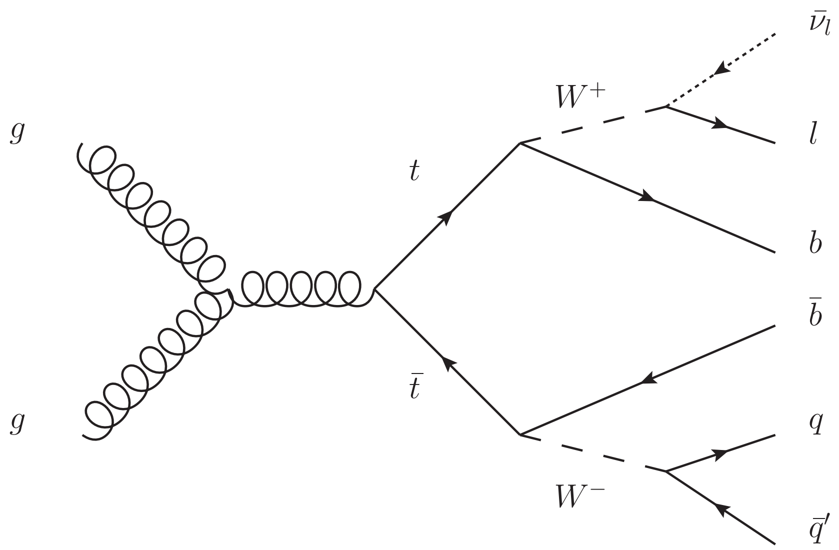

Figure 2.5.: Top anti-top quark decay in the lepton+jets channel.

2.3. Simulation

Because of the complexity of particle physics events and effects that cannot be calculated in perturbation theory, Monte Carlo simulation methods are crucial to compare the data to theory predictions. A full simulation of an event starts with the calculation of the matrix element of the hard process by a Monte Carlo generator like MADGRAPH [27] or POWHEG [28]. This matrix element is folded with the parton density function (PDF) of the proton.

Next, the radiation of additional partons is simulated using a parton shower like the ones implemented in PYTHIA [29] or HERWIG [30]. This is followed by the modeling of hadronization and particle decays, which leads to predictions on particle level. Detector effects are included by a full simulation of the detector implemented in GEANT4 [31, 32].

2.3.1. Monte Carlo Generators

In this analysis two different generators are used to calculate the hard matrix elements. The use of two different calculations is important to test the dependence of the measurement on the underlying assumptions. The two generators used differ most notably in the description of additional radiation to the production of top quark pairs. POWHEG is a next-to-leading order (NLO) matrix element generator, taking into account all real and virtual corrections at NLO, having up to one radiation additional parton to the hard process. MADGRAPH, on the other hand, is a leading order (LO) generator, which allows to add up to three additional partons at tree level to the leading order prediction.

2.3.2. Hadronization

Hadronization includes soft QCD splittings and bound states of colored particles that would lead to divergences in the calculation of the matrix elements. Therefore these processes can only be simulated using tuned models like string fragmentation [29] in case of PYTHIA or cluster fragmentation [30] in case of HERWIG. Both hadronization models have been used in this analysis, to test the influence of the hadronization model on the measurement.

3. Experiment

3.1. Large Hadron Collider

The Large Hadron Collider (LHC) is a proton-proton collider located at CERN (Consseil Européen pour la Recherche Nucléaire or European Organization for Nuclear Research) near Geneva in Switzerland. It is designed for proton-proton collisions at a center of mass energy of $\sqrt{s} = 14\,\text{TeV}$ and is further able to collide heavy ions. The main motivations for such a high energetic proton collider are the search for the higgs boson as well as new physics at the TeV scale, and studies of electroweak symmetry breaking. A key parameter of a collider is its Luminosity L defined as

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma}{4\pi \epsilon_n \beta^*} F, \tag{3.1}$$

with the number of bunches n_b, the number of protons per bunch N_b, the Lorentz factor γ, the normalized emittance ϵ_n, the betatron function in the interaction point β^* and a factor correction for effects related to the beam crossing F. Together with the cross section σ_{process} of a certain process one can calculate the expected event rate as:

$$\dot{N}_{\text{events}} = L\sigma_{\text{process}} \tag{3.2}$$

The design luminosity for 14 TeV is $L = 10^{34}\,\text{cm}^2\text{s}$. More interesting than this instantaneous luminosity is often the integrated luminosity, integrated over time, that is related to the total number of events expected in a dataset. Figure 3.1 shows the integrated luminosity delivered by the LHC to the CMS experiment in the years 2010 to 2012.

The LHC is a circular collider with a length of 26.7 km and four interaction points with experiments. ATLAS and CMS are general purpose detectors build for a full reconstruction of proton-proton collisions, LHCb is used for b-physics and ALICE is specially build for the reconstruction of heavy ion collisions. A detailed description of the LHC can be found in reference [34].

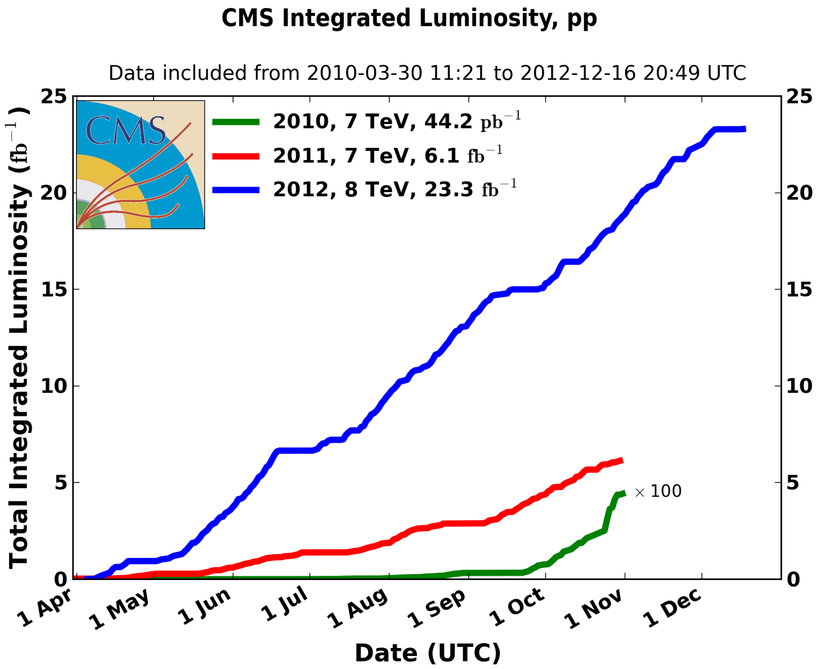

Figure 3.1.: The integrated luminosity in proton-proton collisions delivered by the LHC to the CMS experiment in the years 2010-2012 [33].

3.2. Compact Muon Solenoid (CMS)

The CMS detector is a general purpose detector at one of the LHC interaction points. It is designed to analyze proton-proton collisions at high energies and high repetition rates as well as heavy ion collisions. Its purpose is to reconstruct the transverse momentum and the energy of particles, as well as the trajectories and the electric charge of charged particles, coming from high energy proton-proton collisions at the LHC. To reconstruct a full collision event, several detector components are used and briefly described in the following. Additional information on the CMS-detector can be found in the technical design report [35, 36]. A schematic view is shown in Fig. 3.2.

3.2.1. Coordinate System

The coordinate system, used by CMS to describe the position of detector parts or physical objects, has its origin in the nominal interaction point of CMS. It is a right handed coordinate system with the y-axis pointing vertically in the upward direction, the x-axis pointing horizontally towards the center of the LHC and thus the z-axis pointing anti-clockwise in the direction of the beam. Usually the position of a particle or component is described by the azimuthal angle ϕ in the x-y-plane and the polar angle θ measured with respect to the beam axis. Usually θ is not used directly, but in form of the pseudo

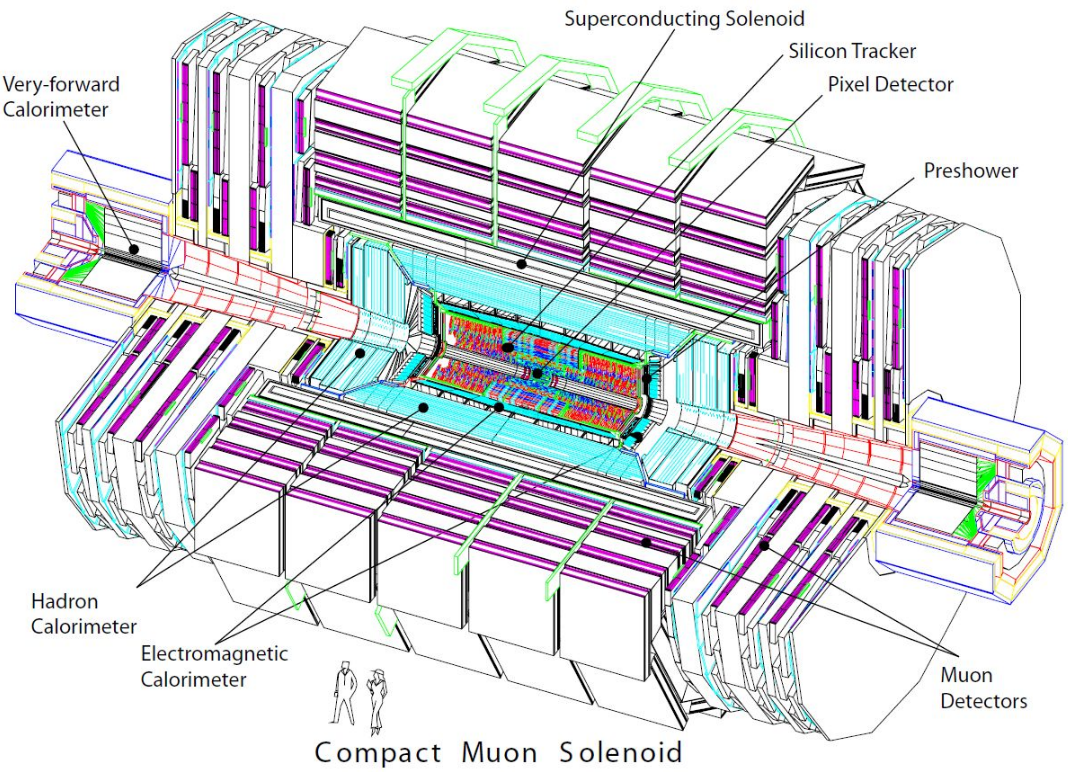

Figure 3.2.: Schematic view of the CMS detector [35].

rapidity η defined as

$$\eta = \ln \tan\left(\frac{\theta}{2}\right). \tag{3.3}$$

The absolute value of η runs from 0 (perpendicular to the beam axis) to infinity (parallel to the beam axis). Distances are calculated as the distance in the η-ϕ-plane $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, since this distance is invariant under longitudinal boosts.

3.2.2. Magnet

In order to measure the full kinematic properties of charged particles, a large and strong magnetic field parallel to the beam axis is needed to bend their trajectories by the Lorentz force. The curvature of these trajectories gives information about the momentum and the charge of the particle. The magnet of the CMS-detector is a superconducting solenoid with an inner radius of 3 m and a length of 12 m. It is able to produce magnetic fields of 4 T. Inside of the solenoid lies the tracker and the calorimetry. An iron yoke around the solenoid returns the magnetic field, providing enough field strength for the muon system outside the magnet.

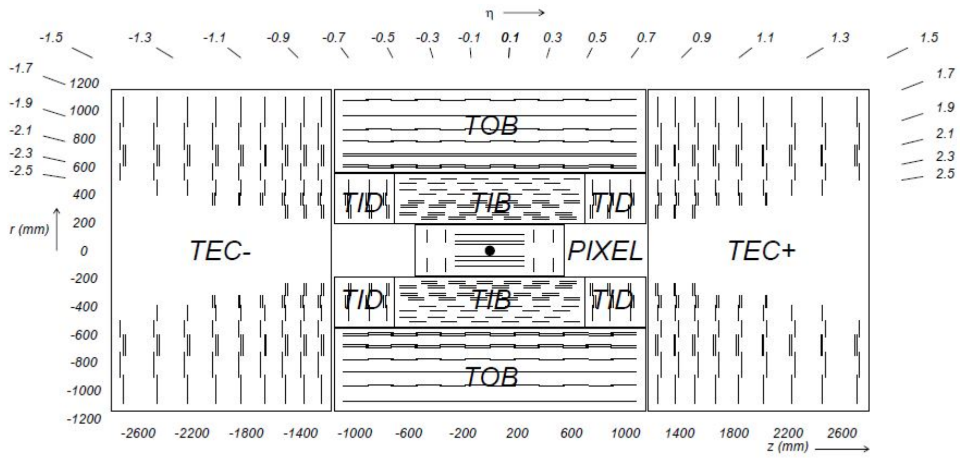

Figure 3.3.: Schematic view of the inner tracking system of CMS [35].

3.2.3. Inner Tracking System

The inner tracking system is used for precise reconstruction of tracks induced by charged particles and to provide a good reconstruction of primary and secondary vertices. A schematic view of the CMS tracking system can be found in Fig. 3.3. It covers a volume of 5.8 m in length and 2.5 m in diameter. Because of more than 20 interactions and over 1000 particles per bunch crossing taking place every 25 ns, CMS chose to build the inner tracking system of semi-conducting silicon layers, providing a good single point resolution and a very fast read out.

The tracking system is divided into two parts, a pixel and a silicon strip detector, each consisting of several layers of semi-conducting silicon detectors. Several layers in the barrel region coaxial around the beam pipe and some discs orthogonal to the beam, to cover the forward direction up to $|\eta| = 2.5$. Particles passing one of these layers create hits by exciting electrons in the conduction band of the silicon and therefore inducing a current, that can be measured. Tracks can be reconstructed by a fit to the hits in the different layers.

Pixel Detector

The pixel detector, schematically shown in Fig. 3.4, covers a rather small volume around the beam pipe. It consists of three layers of silicon pixel modules in the barrel region between 4.4 cm and 10.2 cm radial distance to the interaction point and two endcap discs. Each layer consists of many modules carrying hundreds of $100 \times 150\,\mu\mathrm{m}^2$ silicon pixels. These small pixels provide a good resolution of the tracks near the interaction point and thus improve the reconstruction of primary and especially secondary vertices, which is important for b-tagging.

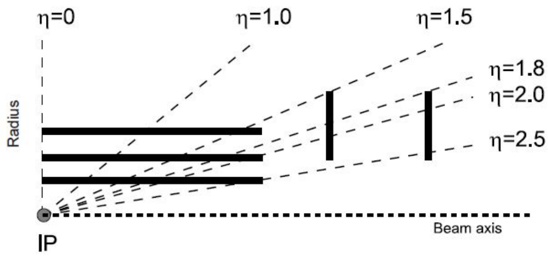

Figure 3.4.: Schematic view of the CMS pixel detector [36].

Silicon Strip Detectors

The silicon strip detectors have a worse resolution than the pixel detector, but at the same time much less read out channels. They can be used to efficiently cover a much larger volume than the pixel detector. The silicon strip tracker consists of ten layers in the barrel region covering a radius up to 1.1 meter and three plus nine endcap discs in the forward direction.

3.2.4. Calorimetry

The tracking system is surrounded by the calorimeter system, which is still inside of the magnet except for a small part of the hadronic calorimeter. Calorimeters are build to absorb particles and to measure the energy they deposit inside their active material.

Calorimeters are crucial for the detection of neutral particles, that will not leave tracks in the tracking system. In case of the CMS calorimeters the deposited energy is measured using scintillator materials. Particles passing a scintillator material excite atoms or molecules in the material, emitting light when returning to their ground state. This scintillation light is read out by photo detectors.

Electromagnetic Calorimeter

Electromagnetic calorimeters (ECALs) are used to measure the energy of electromagnetically interacting particles, like electrons and photons. Those particles start an electromagnetic shower inside the calorimeter and are absorbed completely within the ECAL. Muons, however have a higher mass and interact much less with the material. They deposit energy according to minimal ionizing particles (MIPs), but they are not absorbed completely within the ECAL.

A schematic view of the CMS ECAL is shown in Fig. 3.5. It is a homogeneous calorimeter consisting of lead tungstate ($PbWO_4$) crystals with a high density of $8.28\,\mathrm{g\,cm^{-3}}$. This homogeneous calorimeter provides a good energy resolution needed to be sensitive to the decay channel of the higgs boson into two photons, $H \to \gamma\gamma$.

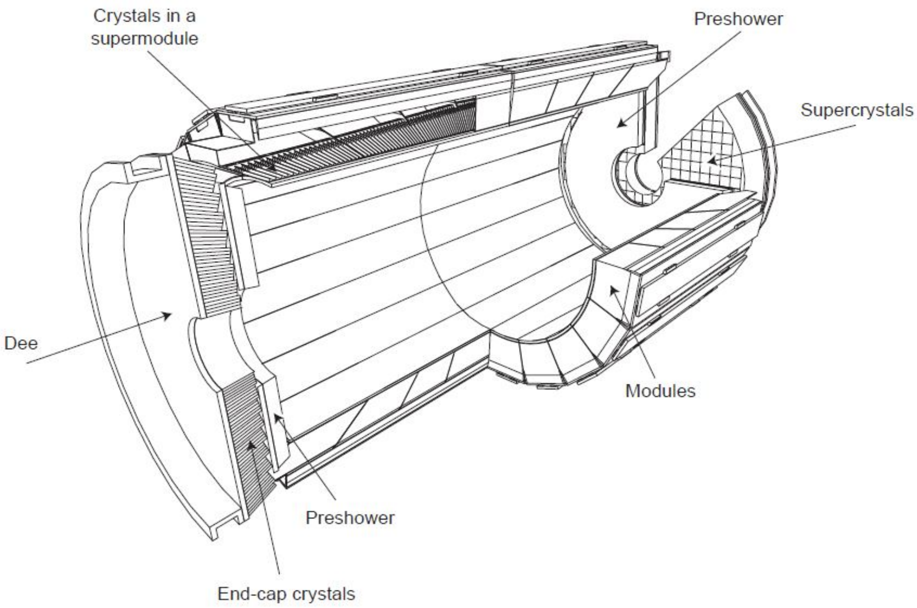

Figure 3.5.: Schematic view of the electromagnetic calorimeter of CMS [35].

Hadrons may start a shower in the ECAL but they will not be fully absorbed because of their much larger interaction length, compared to the radiation electrons and photons. For this reason, the ECAL is complemented by a hadronic calorimeter.

Hadronic Calorimeter

Considering the large interaction length of hadrons and the limited space in typical particle physics experiment, hadronic calorimeters (HCALs) can not be build homogeneous. Instead, HCALs always have a sampling structure with layers of dense material only used for absorption alternating with layers of active material for the read out. The CMS HCAL is a sampling calorimeter with absorption layers made of brass and plastic scintillator as active material. To get a better control of energy leakage form highly energetic showers, an additional layer of HCAL is located right outside of the solenoid, the hadronic outer (HO) calorimeter.

3.2.5. Muon system

As minimal ionizing particles (MIPs) muons are the only charged particles that are not absorbed by the calorimeters, allowing a separate reconstruction outside of the magnet. The muon system is located in several coaxial layers around the magnet and within the iron yoke, as well as in the forward region in form of two endcap discs. It is designed to measure efficiently the momentum and energy of muons and is included in the trigger system. For this purpose, the muon system consists of three different types of gaseous detectors, each with its advantages for it's specific region and it's dedicated purpose.

- Drift tubes are used in the central barrel region up to $|\eta| = 1.2$. In this region, the muon rate and the background is low.
- In the endcaps, where muon and background rates are high, cathode strip chambers are used with a faster response time and a finer segmentation.
- For the trigger additional, resistive plate chambers are integrated in the barrel and the endcap region.

3.2.6. Trigger

At design luminosity the bunch spacing is 25 ns, leading to a total event rate of 40 MHz. However, only a few hundred Hz can be stored. Fortunately interesting processes like $t\bar{t}$-production have much smaller event rates and most of the processes are soft QCD effects.

The purpose the trigger system is therefore to reduce the event rate from 40 MHz down to about 100 Hz that can be stored without loosing interesting physics events. The CMS trigger system consists of two parts, a hardware based level one trigger (L1) and a software based high level trigger (HLT). The L1 trigger must be able to make hardware based decisions within a few ns. Therefore it uses hardware components that provide a fast read out. It uses the calorimetry and the muon system of the CMS detector as well as global combinations of these systems like calculation of transverse energy E_T or missing transverse energy. It is able to write out events with a rate of about 100 kHz. The High-Level trigger is used further reduce the event rate from 100 kHz to 100 Hz for storage. It uses the full granularity of the detector and does not reconstruct the full event instantaneously, but only step by step, in order to be able to reject the event as soon as possible.

4. Event Reconstruction

4.1. Particle Flow Algorithm

The particle follow algorithm [37] is used in CMS to reconstruct and identify every reconstructible particle in the event using a combination of all detector components.

The identification is segmented in three stages. At the first stage muons are identified by reconstructed tracks in the inner and the muon tracker. In the second step, electrons are identified by associating a track to a cluster in the ECAL. As a last step the remaining tracks are matched to clusters in the HCAL or ECAL. The obtained particle candidates are called charged hadrons if the energy determined by the tracker is consistent within three sigma with the energy deposit measured with the calorimeter. In this case the energy of a charged hadron is taken only from the track information using a charged pion mass hypothesis. In this way the particle flow algorithm reduces the influence of the worse resolution of the hadronic calorimeter to the jet reconstruction.

Neutral particles like photons and neutral hadrons leave no track in the tracker and are identified by clusters in the calorimeters. These particle candidates are considerd as photons if they just deposit energy in the ECAL, otherwise they are as neutral hadrons.

4.2. Jets

As a result of the confinement of quarks and gluons, these particles cannot exist as free particles, but start to hadronize. This hadronization together with radiation of additional particles leads to a bunch of stable particles propagating in approximately the same direction as the initial particle and carrying in total the same momentum. This is called a jet. A good definition of jets is therefore crucial for most hight energy physics analyses.

4.2.1. Jet Algorithms

Jet algorithms are used to define jets from the reconstructed four momenta of particles in the event. Jet algorithms have to fulfill two conditions in order to obtain cross sections calculable beyond leading order. These two conditions are called infrared and collinear safety.

- Infrared safety means that the algorithm should be insensitive to soft radiation of a particle.
- Collinear safety means that a collinear splitting of particles should not change the jet. In other words: the jet algorithm should be insensitive to the replacement of one particle with momentum p by two collinear particles with momenta $p/2$.

Jet algorithms used in this thesis are sequential clustering algorithms. These algorithms start with four-momenta of all particles in the event and cluster them step by step to jets. All algorithms discussed in the following look for pairs of four momenta with the smallest defined distance parameter and merge them into one new four momentum if the distance is not bigger than a chosen maximal distance parameter R_0. Jets are build by a repetition of merging the four momenta until there is no pair four momenta with $\Delta R < R_0$ left. There are three commonly used algorithms working in this way. They differ in their definition of the distance parameter $d_{j_1 j_2}$.

- The Cambridge/Aachen (C/A) algorithm [38] is a purely geometrical jet algorithm in the sense, that its distance parameter is just the distance in the η-ϕ-plane
$$d_{j_1 j_2} = \Delta R^2_{j_1 j_2} \frac{1}{R_0}, \tag{4.1}$$
with $R = \sqrt{\eta^2 + \phi^2}$. This algorithm is often used for jet substructure studies.
- The k_T algorithm [39] weights the geometrical distance with the transverse momentum of the jet,
$$d_{j_1 j_2} = \Delta R^2_{j_1 j_2} \min(p^2_{T,j_1}, p^2_{T,j_2}) \frac{1}{R_0}. \tag{4.2}$$
In this way it combines low p_T four vectors first.
- The anti-k_T algorithm [40] is used in CMS for the standard jet reconstruction with a maximum distance parameter of $R_0 = 0.5$. Weighting the distance with p_T^{-1}, it combines the hardes particles first and ends with adding soft ones,
$$d_{j_1 j_2} = \Delta R^2_{j_1 j_2} \min(\frac{1}{p^2_{T,j_1}}, \frac{1}{p^2_{T,j_2}}) \frac{1}{R_0}. \tag{4.3}$$

4.2.2. Jet Energy Corrections

Jet energy corrections have to be applied to reconstructed jets to correct for differences between measured jets and particle level jets due to non linearities in the detector response, as well as differences between data an MC.

A detailed description of these corrections can be found in reference [41] and a brief overview is given below.

The total jet energy corrections are applied as a correction factor to the raw four-vector of the jet,

$$p_\mu^{\text{corrected}} = C\, p_\mu^{\text{raw}}. \tag{4.4}$$

This correction consists of four contributions combined by a factorization ansatz,

$$C = C_{\text{offset}}(p_T^{\text{raw}})\, C_{\text{MC}}(p'_T, \eta)\, C_{\text{rel}}(\eta)\, C_{\text{abs}}(p''_T), \tag{4.5}$$

where p'_T and p''_T represent the corrected p_T after the previous correction steps.

The contributions are:

- **offset** C_{offset}: The first correction factor is used to correct for an offset between reconstruction and particle level due to pile-up and electronic noise.
- **MC calibration** C_{MC}: The MC calibration correct the energy of reconstructed jets, such that it is on average equal to the energy of the jets on particle level.
- **relative corrections** C_{rel}: This factor corrects for an η dependence of the detector response returning a response flat in η.
- **absolute corrections** C_{abs}: Because of remaining residual effects in the response in data and simulation the absolute scale of the jet energy is derived from Z + jet and γ +jet events.

4.2.3. Pile-Up

At design luminosity the LHC produces on average more than 20 collisions per bunch crossing. After triggering typically only one of these processes includes a physically interesting process and the rest are soft QCD processes, that can overlap with the hard process and change event variables. Especially jet properties are influenced by the amount of pile-up in the event, because they consist of many particles covering rather large areas in the detector. In this way jets with a larger distance parameter cover a larger area and are therefore more effected by pile-up.

One possibility to reduce the influence of pile-up is **charged hadron subtraction** used by CMS and described in reference [42]. The particle flow algorithm reconstructs charged hadrons matching calorimeter clusters with tracks. The tracks can be related to the the primary vertices in the event. A leading primary vertex is defined as the vertex with the highest value of the sum of the squared track transverse momenta $\sum |p_T^{\text{track}}|^2$, all other primary vertices are considered to be pile-up vertices. The charged hadron subtraction subtracts all particles that can be associated with a pile-up vertex. Particles that can not be related to pile-up remain in the event.

4.2.4. Jet Grooming

Jet grooming methods are used to reduce the influence of pile-up and initial or final state radiation on single jets. The intention is to remove soft radiation and pile-up particles without changing the substructure of the jet. Three of these grooming methods are considered in this analysis:

1 **Filtering [43]:** The filtering algorithm takes all stable particles of a jet and re-clusters them using a clustering algorithm with a smaller distance parameter R_{filter}. This results in many smaller jets, ordered by their transverse momentum. In a next step only a number of N jets, the N jets with the highest p_T, are combined into a filtered jet. If there are less than N jets, all jets are kept. A disadvantage of this method is that it imposes a specific jet substructure by choosing a fixed number of N subjets. A typical distance parameter for the re-clustering is $R_{\text{filter}} = 0.3$

2 **Trimming [44]:** The ansatz of trimming is similar to the filtering algorithm. All stable particles of the jet are again clustered with a smaller parameter R_{sub}. Now all of these small jets that carry more than a certain fraction f_{cut} of the p_T of the entire jet are combined to the final trimmed jet. Typical parameters for trimming are $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 0.03$.

3 **Pruning [45]:** Pruning repeats the jet clustering with the same jet algorithm and the same R_0 and checks at every combination of object i and object j to object p the following two conditions:

$$\frac{\min(p_{Ti}, p_{Tj})}{p_{Tp}} < z_{\text{cut}} \qquad \text{and} \qquad R_{ij} > D_{\text{cut}}, \tag{4.6}$$

where z_{cut} and D_{cut} are predefined values. If these conditions are fulfilled objects i and j are not merged and the softer one is excluded from the clustering process. The idea behind these conditions is to reject soft radiation with large distances to the hard objects, that probably originate from pile-up or initial state radiation. Typical parameters are: $D_{\text{cut}} = m_{\text{jet}}/p_{T,\text{jet}}$ and $z_{\text{cut}} = 0.1$.

4.3. Boosted Top Quarks and Top Tagging

Due to the high center of mass energy of the LHC, top quarks can be produced with very high transverse momenta. For those quarks the decay products are boosted in the direction of flight of the decaying top quark due to the large Lorentz boost. The higher the momentum, the lower the distance between two of its decay products. This behavior is shown in Fig. 4.1, where the distance of two decay products of a top quark is shown on the y-axis against the p_T of the top quark on the x-axis. It is clearly visible, that the distance between two decay products decreases with the p_T of the top qurak.

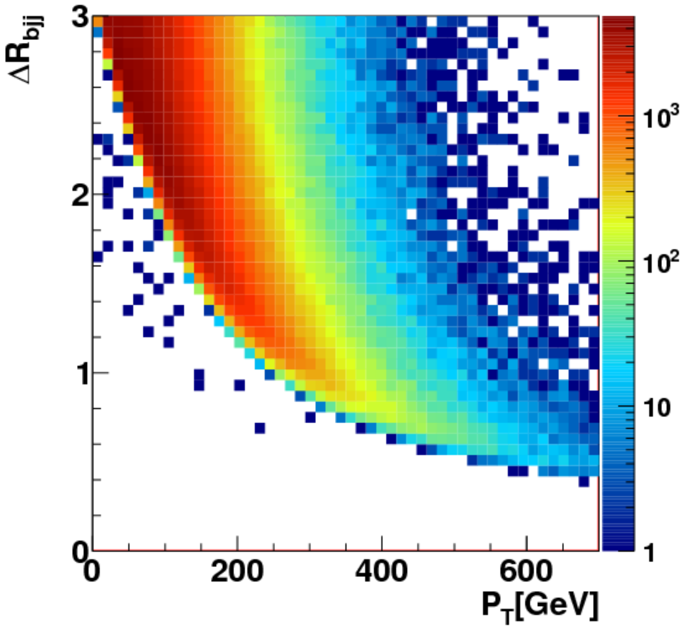

Figure 4.1.: Distance between two decay products of the top quark vs. the transverse momentum of the top quark [46].

Especially for hadronic top quark decays this effect gets very interesting, because all decay products form jets, which start to merge into one big or fat jet for approximately $p_T > 400\,\mathrm{GeV}$, assuming a distance parameter R_0 of the order 1. In this case the jets from the individual quarks cannot be reconstructed as separated jets anymore, but will be reconstructed as a single fat jet. The study of variables sensitive to the substructure of fat jets is important to identify them as top quark jets and distinguish them from QCD induced jets. This identification of top quark jets is called top tagging and plays an important role for the search of heavy particles that decay into top quarks. An overview of top tagging algorithms used in CMS can be found in reference [5].

Apart from the identification of top quarks, substructure variables can also be used to study standard model parameters. For example the shape of the jet mass distribution for boosted hadronic top quark jets is sensitive to the mass of the top quark, as described in section 2.2.1.

4.4. Missing Transverse Energy and $\mathrm{H_T}$

Missing transverse energy $\not{E}_T$ and the scalar sum of transverse momenta H_T are important variables to distinguish events including a leptonically decaying W boson, like $t\bar{t}$-events in the lepton+jets final state from QCD multi-jet events. $\not{E}_T$ originates from an imbalance with respect to the total momentum of all reconstructed particles in the event. Due to the absence of transverse momentum in the initial state and momentum conservation, the vectorial sum of all particle momenta should add up to zero. This is no longer the case if one particle leaves the detector undetected, like a weakly interacting neutrino.

Missing transverse energy is defined as the negative vectorial sum of reconstructed particle transverse momenta,

$$\not{E}_T = |\sum_{particles} -\vec{p}_T|. \tag{4.7}$$

A description of the reconstruction of $\not{E}_T$ in CMS can be found in reference [47]. Another important variable to reject processes not including a leptonically decaying W boson is H_T^{leptonic}, which is defined as the scalar sum of the transverse momenta of the reconstructed leptons in the event added with $\not{E}_T$,

$$H_T^{\mathrm{leptonic}} = \sum_{\mathrm{leptons}} |\vec{p}_T| + \not{E}_T. \tag{4.8}$$

4.5. Muons

Different methods are used within CMS to reconstruct muons. There is the possibility to reconstruct muons locally in the muon-system or the inner tracker, or globally using both trackers. A detailed description of the muon reconstruction can be found elsewhere [36, 48]. Here a short overview is given.

A local standalone muon reconstruction using the muon system only is used by the HLT trigger seeded by the L1 trigger. Global fits can be performed in two directions, starting from the muon-system or the inner tracker.

- The global muon reconstruction starts with a standalone muon reconstruction in the muon system. This standalone track is then matched to a track in the inner tracker and a global fit using the information of both tracks is performed.
- The tracker muon reconstruction works the other way around, starting from a track in the inner tracker. Tracks with a $p_T > 0.5\,\mathrm{GeV}$ and a momentum $p > 2.5\,\mathrm{GeV}$ are considered as muon candidates. The tracks are extrapolated to the muon system and defined as a muon if they match hits in the muon system. This method has a better efficiency as the global muon reconstruction for low transverse momenta $p_T < 5\,\mathrm{GeV}$, because only one muon segment is required.

The muon identification has three working points: loose, soft and tight. For this analysis the tight working point is used, that has the following criteria.

- The muon has to be reconstructed as a global and a tracker muon.
- The muon has to have a $p_T > 3\,\mathrm{GeV}$.
- The normalized χ^2 of the global fit is smaller 10.
- There has to be at least one muon chamber hit in two muon sections.
- At least 10 hist in the inner tracker including one hit in the pixel detector are found.
- The transverse impact parameter has to be smaller 2 mm.

4.6. Electrons

The reconstruction of electrons is more challenging than the reconstruction of muons. Because of the smaller mass, electrons produce more bremsstrahlung passing the tracker layers. Together with the magnetic field of 4 T this leads to an energy spread in the ϕ direction. This is a challenge for fitting the track, as well as defining the clusters in the ECAL. A detailed description of the electron reconstruction in CMS can be found in references [36, 49].

In the ECAL, so called super-clusters are defined consisting of several smaller clusters containing all the radiation emitted by the electron to reduce energy fluctuations. The reconstruction constructs primary electrons consisting of tracks, related to the interaction point, and matched to a super-cluster. Seeds for the tracking can be found using super-clusters in the ECAL. For high p_T electrons the tracks can then be fitted from the inner pixel layers to the outer layers of the tracker using a Kalman-Filter [50]. Low p_T electrons suffer non linear effects due to bremsstrahlung and the Kalman Filter gets inefficient. For those events a non linear generalization of the Kalman-Filter, the Gaussian Sum Filter [51] can be used for the fitting procedure.

Electrons can be identified by matching super-clusters to reconstructed tracks and can be divided into different classes applying certain criteria.

The electrons used in this thesis have to fulfill multivariate identification criteria described in reference [52]. They have to fulfill the following criteria.

- They have to pass a conversion veto.
- There should be not tracker layers before the first hit of the electron track.
- They have to pass a cut on the Triggering MVA discriminator (MVAtrig) dependent on the η of the super-cluster η_{SC}:
  - MVAtrig > 0.94 for $|\eta_{SC}| < 0.8$
  - MVAtrig > 0.85 for $0.8 < |\eta_{SC}| < 1.479$
  - MVAtrig > 0.92 for $1.479 < |\eta_{SC}| < 2.5$
- Their transverse momentum has to be greater than 35 GeV and $|\eta| <2.5$.

4.7. B-Tagging

B-tagging algorithms are used to distinguish jets form b-quark decays from those originating from lighter quarks and gluons. This is possible because the b-quark has a short lifetime ($\approx$ 1.5 ps) and therefore the B-meson inside of the jet decays after a short distance creating a secondary vertex within the jet that may be reconstructed. Tracks from this vertex can not be assigned to the primary vertex and are called displaced tracks. A description of b-tagging algorithms used in CMS can be found in reference [53].

In CMS secondary vertices are reconstructed using an adaptive vertex fitter. Vertices that share more than 65% of their tracks with the primary vertex are rejected, as well as vertices that have a distance $\Delta R > 0.5$ to the jet axis or vertices having a transverse distance to the primary vertex greater 2.5 cm and a mass compatible with the K^0 mass.

A very important variable to separate b-jets from light jets is the impact parameter. It gives the spacial distance between a track and the primary vertex at the point of closest approach. In contrast to the decay length of the b-meson, the impact parameter is Lorentz invariant.

Sometimes it is possible to combine tracks with a significant impact parameter to a so called pseudo vertex. This allows the calculation of vertex properties without a proper vertex reconstruction.

In this Analysis the standard CSV (Combined Secondary Vertex) tagger [53] is used. A CSV discriminator uses a combination of vertex reconstruction and track based lifetime information. Examples for these properties are the vertex category (real, pseudo or no vertex), the vertex mass, the number of tracks assigned to the vertex and the jet, to name a few. A detailed list of all parameters can be found in reference [53]. The combination of these properties to one discriminator is done using a likelihood based multivariate method. This combination allows to tag a jet, even if just a pseudo or no vertex is found.

The cut on the CSV discriminator defines the working point of this tagger. The three official CMS working points are loose, medium and tight with a mis-identification probability of jets induced by lighter quarks of about 10%, 1% and 0.1%.

5. Unfolding

5.1. Basic Problem

An important aspect of particle physics is the measurement of observables and a comparison with theoretical predictions. The measured quantities should be independent of the experimental set up to be comparable among different experiments and to our theoretical expectations.

For this reason detector effects, like a finite detector resolution or acceptance have to be corrected for. This problem is addressed by unfolding algorithms, constructed to unfold detector effects from measured distributions. The aim of the unfolding is to get an estimator for a true distribution $\mathbf{x}$[1] with a given measurement of a reconstructed distribution $\mathbf{y}$. Due to detector effects an event truly lying in bin x_i can be reconstructed in a different bin y_j of the reconstructed distribution. Those Migrations can be described using a matrix of probabilities $\mathbf{A}$, that gives the probability that an event in bin i is reconstructed in bin j. Using this approach the problem can be written as:

$$\tilde{y}_i = \sum_{j=1}^{m} A_{ij}\tilde{x}_j \quad , \quad 1 \leq i \leq n, \tag{5.1}$$

where the tilde indicates that this is the statistical mean, which is not necessarily the true quantity due to statistical fluctuations. Furthermore is n the number of bins of the reconstructed distribution and m the number of bins of the true distribution $\mathbf{x}$.

In a typical application the reconstructed distribution is measured, the migration matrix is determined from simulation and the true distribution is unknown. A schematic view is given in Fig. 5.1, where a migration matrix $\mathbf{A}$ introduces detector effects to the true distribution $\mathbf{x}$ to obtain a distribution $\tilde{\mathbf{y}}$ that can be measured. A real measurement further includes statistical fluctuations leading to a distribution $\mathbf{y}$. The aim of the unfolding is now to obtain the true distribution from a measured distribution $\mathbf{y}$, by solving equation (5.1) for $\mathbf{x}$.

[1] $\mathbf{x}$ is a vector of the bins of the true distribution.

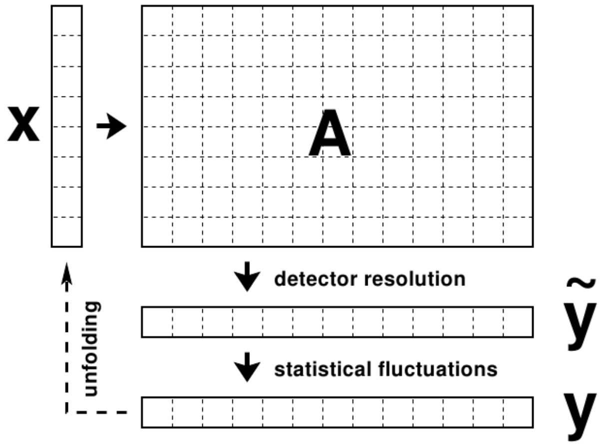

Figure 5.1.: Schematic representation of the unfolding problem, taken from reference [8], with the true distribution $\mathbf{x}$, the reconstructed distribution $\mathbf{y}$ and the response matrix $\mathbf{A}$

5.2. Regularized Unfolding in TUnfold

The unfolding method used in this thesis is a regularized unfolding method implemented in TUnfold. A short overview of the principle and the basic formulas is given in the following. More information about TUnfold can be found in reference [8].

In principle the problem described before in equation (5.1) can be solved using a least squares method. This would lead to a minimization of

$$\chi^2 = (\mathbf{y} - \mathbf{A}\mathbf{x})^{\mathrm{T}} \mathbf{V_{yy}}^{-1} (\mathbf{y} - \mathbf{A}\mathbf{x}) \tag{5.2}$$

where $\mathbf{V_{yy}}$ is the covariance matrix of y.

From equation (5.1) to equation (5.2) a replacement of $\tilde{\mathbf{y}} \to \mathbf{y}$ and $\tilde{\mathbf{x}} \to \mathbf{x}$ is done. This replacement of values without statistical fluctuations by the ones including the fluctuations introduces large statistical fluctuations to the measurement amplified by the unfolding. To damp these fluctuations it is necessary to introduce a regularization term to the minimization.

$$\tau^2 (\mathbf{x} - f_b \mathbf{x_0})^{\mathrm{T}} (\mathbf{L}^{\mathrm{T}} \mathbf{L}) (\mathbf{x} - f_b \mathbf{x_0}), \tag{5.3}$$

where τ is a constant parameter giving the strength of the regularization and $\mathbf{L}$ is a matrix containing the regularization conditions. These parameters are explained in the following section. The regularization term introduces a small bias $f_b \mathbf{x_0}$ to the measurement of $\mathbf{x}$ consisting of a normalization factor f_b and a vector $\mathbf{x_0}$. At the same time it reduces the influence of statistical fluctuations to the result. The matrix $\mathbf{L}$ determines if the bias is introduced to the the single values in each bin, to the first derivative between the bins or to the curvature (second derivative).

The unfolding method in TUnfold looks for the minimum of the Lagrangian

$$\begin{aligned}\mathcal{L}(x,\lambda) =& \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 \\ =& (\mathbf{y} - \mathbf{A}\mathbf{x})^{\mathrm{T}} \mathbf{V}_{\mathbf{yy}}{}^{-1} (\mathbf{y} - \mathbf{A}\mathbf{x}) \\ &+ \tau^2 (\mathbf{x} - f_b \mathbf{x_0})^{\mathrm{T}} (\mathbf{L}^{\mathrm{T}} \mathbf{L}) (\mathbf{x} - f_b \mathbf{x_0}) \\ &+ \lambda (Y - \mathbf{e}^{\mathrm{T}} \mathbf{x}),\end{aligned} \tag{5.4}$$

where the two first terms are already known from the least squares method and the regularization. The third term is an optional area constraint, which ensures, that the integral of the input is the same as the integral of the output. In this term Y is the sum over all components of $\mathbf{y}$, $Y = \sum_i y_i$ and the vector $\mathbf{e}$ is defined as $e_j = \sum_i A_{ij}$.

The minimum is found by using the first derivatives.

$$\frac{\partial \mathcal{L}(\mathbf{x},\lambda)}{\partial x_j} = 0 \qquad \text{and} \qquad \frac{\partial \mathcal{L}(\mathbf{x},\lambda)}{\lambda} = 0 \tag{5.5}$$

The exact solution of $\mathbf{x}$ and the determination of its covariance matrix can be found in reference [8].

5.3. Regularization Strength and Conditions

The choice of the regularization strength and conditions should be done carefully to get a result with small fluctuations, that is at the same time not dominated by the regularization.

Regularization Strength

There are two methods implemented in TUnfold to determine the optimal regularization strength:

- L-Curve Scan

  The L-Curve scan defines two Variables $L_x^{\text{curve}} = \log \mathcal{L}_1$ and $L_y^{\text{curve}} = \log \frac{\mathcal{L}_2}{\tau^2}$ that quantify the influence of the first two terms of the lagrangian to the result. The τ value giving the best compromise with respect to the influence of the terms $\mathcal{L}_1$ and $\mathcal{L}_1$ to the result is determined, building a graph in the $L_x^{\text{curve}} - L_y^{\text{curve}}$ plane for many τ values. The obtained graph is typically l-shaped and the τ value in the point of largest curvature is the one giving the best compromise. Figure 5.2 shows a schematic view of a possible graph with L_x^{curve} on the x-axis and L_y^{curve} on the y-axis. Every point on the blue curve corresponds to a specific value of τ.

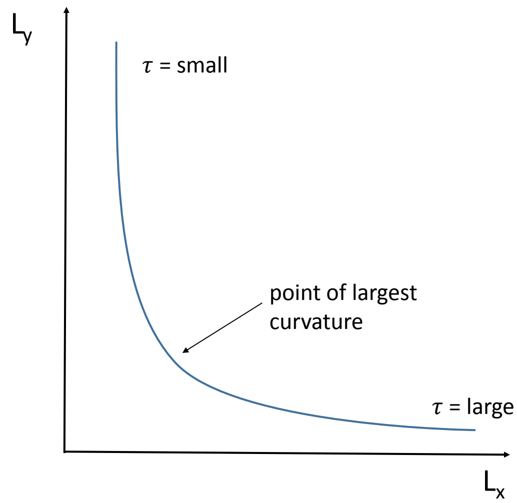

Figure 5.2.: Schematic representation of a possible L-curve graph in the $L_x^{\text{curve}} - L_y^{\text{curve}}$ plane, with $L_x^{\text{curve}} = \log \mathcal{L}_1$ and $L_y^{\text{curve}} = \log \frac{\mathcal{L}_2}{\tau^2}$. Every point on the cure belongs to a specific value of τ. The point of largest curvature determines the best compromise with respect to the influence of the terms $\mathcal{L}_1$ and $\mathcal{L}_1$ to the result.

- Minimization of global correlation coefficients

  A second method to determine the best value of τ is a minimization of the global correlation coefficients defined as:

$$\rho_i = \sqrt{1 - \frac{1}{(\mathbf{V_{xx}}^{-1})_{ii}(\mathbf{V_{xx}})_{ii}}} \quad , \tag{5.6}$$

  with the covariance matrix of the output $\mathbf{V_{xx}}$.

  The first term of the lagrangian causes large positive correlations, whereas the regularization term introduces negative correlations. So in principle there should be a value of τ, for which the correlations cancel and the mean global correlation coefficient reaches a minimum. This value of τ is chosen as the best choice.

Regularization Conditions

The Matrix $\mathbf{L}$ can be chosen in such way that the regularization acts on the absolute values of $\mathbf{x}$ (size regularization), on the first or on the second derivative of $\mathbf{x}$. Which condition performs best, depends on the unfolded distribution and on the chosen method to determine the τ-parameter. This will be discussed further for the case of this analysis in the Analysis chapter.

6. Analysis

The goal of this analysis is a measurement of the jet mass on particle level of jets containing a fully-hadronic top quark decay. The particle level is the level of stable particles after hadronization, but without detector effects. A measurement to particle level provides the possibility to compare the results with other experiments and to theoretical calculations.

This chapter is divided into three parts. In the first part the particle level phase space is defined in such a way, that it contains mostly fully merged hadronic top quark decays and can be compared to theoretical calculations. The second part includes a similar phase space definition on reconstruction level, that further suppresses background processes like QCD multi-jet production, W+jets and single top quark production. In the third part a regularized unfolding is set up and tested for model dependence and the dependence of the top quark mass, before it is applied to the 2012 CMS data set.

6.1. Data and simulation samples

Data

The used dataset includes an integrated luminosity of $19.7\,\mathrm{fb}^{-1}$, collected in proton-proton collisions with a center of mass energy of $\sqrt{s} = 8\,\mathrm{TeV}$ at the LHC by the CMS detector in the years 2010 to 2012.

Simulation Samples

Signal samples are top anti-top quark pair production samples produced with different Monte Carlo Generators and different top quark masses. A full list of samples used is listed in Tab. 6.1. The dominant backgrounds for this analysis are QCD multi-jet production, W boson production in association with an additional jets and single top quark production. The samples used are listed in Tab. 6.2. To determine the systematic uncertainty due to scale variations of the renormalization and factorization scales, four samples with varied scales and high mass of the $t\bar{t}$-system have been used, listed in Tab. 6.3.

characteristic	**generator**	**number of events**	**integrated luminosity [pb^{-1}]**
inclusive	POWHEG	21675970	88185
$700\,\text{GeV} < m_{t\bar{t}} < 1000\,\text{GeV}$	POWHEG	3082812	169478
$m_{t\bar{t}} > 1000\,\text{GeV}$	POWHEG	1249111	363114
	MADGRAPH	62240059	253214
$m_{top} = 166.5\,\text{GeV}$	MADGRAPH	27284436	111003
$m_{top} = 169.5\,\text{GeV}$	MADGRAPH	40952045	166607
$m_{top} = 171.5\,\text{GeV}$	MADGRAPH	24655025	100305
$m_{top} = 173.5\,\text{GeV}$	MADGRAPH	26807385	109062
$m_{top} = 175.5\,\text{GeV}$	MADGRAPH	40401115	164366
$m_{top} = 178.5\,\text{GeV}$	MADGRAPH	26752453	108838

Table 6.1.: Simulated $t\bar{t}$ signal samples used in this analysis.

process	**characteristic**	**generator**	**number of events**	**integrated luminosity [pb^{-1}]**
W+jets	+ 1 jet	MADGRAPH	23141598	3473.2
	+ 2 jets	MADGRAPH	34044921	15767
	+ 3 jets	MADGRAPH	15539503	24266
	+ 4 jets	MADGRAPH	13382803	50683
single top	s-channel	POWHEG	259961	68591
	t-channel	POWHEG	3758227	66635
	tW-channel	POWHEG	497658	44834
	anti top s-channel	POWHEG	139974	79531
	anti top t-channel	POWHEG	1935072	63032
	anti top tW-channel	POWHEG	493460	44456
QCD	muon enriched	PYTHIA6	79261757	-
	electron enriched	PYTHIA6	214438367	-

Table 6.2.: Simulated samples of background processes, used in this analysis. For QCD different samples for different p_T regions are used and therefore no luminosity is given here.

characteristic	**variation**	**generator**	**number of events**	**integrated luminosity [pb^{-1}]**
$700\,\text{GeV} < m_{t\bar{t}} < 1000\,\text{GeV}$	up	POWHEG	1241650	123346.5
	down	POWHEG	1308090	119300.4
$m_{t\bar{t}} > 1000\,\text{GeV}$	up	POWHEG	2243672	360944
	down	POWHEG	2170074	380258.7

Table 6.3.: Simulated $t\bar{t}$ samples with varied renormalization and factorization scales used in this analysis to estimate systematic uncertainties.

6.2. Studies on Particle Level

The particle level is the level of stable particles after hadronization, but before the detector simulation. On this level the measurement will be compared with theoretical calculations from first principle. For this reason a precise definition of the phase space on this level is crucial for this analysis.

This section includes studies of events simulated with MADGRAPH and POWHEG, showered and hadronized with PYTHIA before the detector simulation. The phase space for the measurement is defined and differences between MADGRAH and POWHEG are studied.

6.2.1. Phase Space Definition

One of the most important steps of this analysis is the definition of the phase space on particle level, because it defines the result of the measurement. This phase space definition should just include theoretically well defined and calculable objects and should comprise a selection that can be reconstructed in the experiment with reasonable background suppression.

The measurement will be done in top anti-top quark decays with one top quark decaying hadronically and one leptonically including an electron or muon. The electron and muon+jets channels ensure a reasonable suppression of backgrounds from processes not involving top quarks on reconstruction level. The all hadronic decay channel is not used as it would have too much QCD-multi-jet background, and the di-leptonic channel provides no hadronic top quark decay. Tau leptons are excluded because their reconstruction is difficult. Because of their decay into pions they can easily be reconstructed as jets in the detector, making their identification and in this way the background subtraction inefficient.

Minimal Transverse Momentum of the Jets

In order to measure the mass of jets containing a fully merged hadronic top quark decay, the transverse momentum of the top quark should be high enough to ensure that its decay products are boosted in the direction of flight of the top quark and merge into a single jet. Moreover the phase space should be defined in such a way, that the cross section is calculable to any order in perturbation theory. That means the transverse momentum of the top quark has to be much higher than its mass $p_{T,\text{top}} >> m_{\text{top}}$.

The transverse momentum p_T of the hadronic top quark jets is correlated to the transverse momentum of the top quark itself. For theory calculations the p_T should be as high as possible, while for an experimental measurement there have to be enough events left in data after reconstruction.

To decide on the minimum transverse momentum of the hadronic top quark jets, the

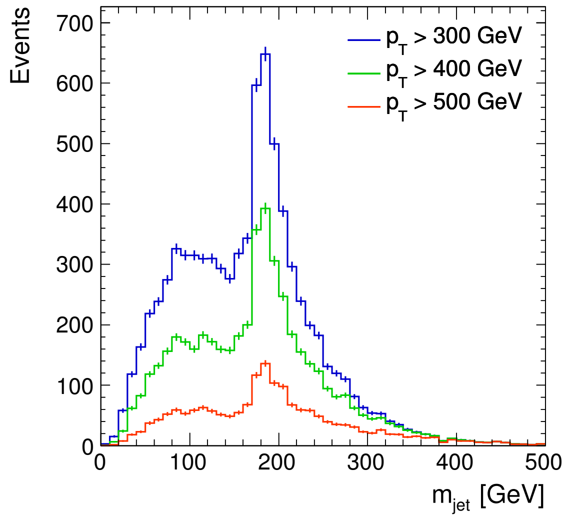

Figure 6.1.: Distribution of the invariant mass of the leading jet in top anti top quark decays in the lepton+jets channel for different transverse momenta of the leading jet. The transverse momentum of the second jet is greater 100 GeV.

mass of the leading jet in $t\bar{t}$ events in the lepton+jets channel including an electron or muon is shown for a transverse momentum of the leading jet greater 300, 400, 500 GeV in Fig. 6.1. The jet algorithm used for the jet clustering is a Cambridge/Aachen algorithm with an R parameter of 1.2. Because of a second jet in the event formed by the bottom quark from the leptonically decaying top quark, a cut on the transverse momentum of the second jet to be greater 100 GeV is applied. The figure shows the distribution of the jet mass for events simulated with POWHEG + PYTHIA with a center of mass energy of $\sqrt{s} = 8$ TeV and normalized to an integrated luminosity of 19.7 fb^{-1}.

The distribution shows a sharp peak at about 180 GeV, originating from jets including all decay products of a top quark decay. The value of 180 GeV is higher than the mass of the top quark because of initial and final state radiation. Jets with lower masses originate from events in which not all top quark decay products are clustered into the leading jet. In one example only the decay products of the hadronic W boson decay are clustered into the jet, leading to a mass around 80 GeV. In another example the leading jet is formed by the bottom quark only, or from the leptonically decaying side, or by radiation.

Assuming a reconstruction efficiency of 10%, a cut on the leading jet p_T greater 500 GeV is not expected to provide enough events in data to perform a measurement with small enough statistical uncertainties. The amount of events around the top quark mass peak for a leading jet p_T >400 GeV could include enough events for the measurement after reconstruction. Taking into account that theoretical calculations need a high transverse momentum of the top quark [54], the transverse momentum of the leading jet is chosen to be greater 400 GeV for the rest of this thesis.

Basic Selection and Distance Parameter of the Jets

The transverse momentum of the top quark has a direct impact on the distance ΔR between two of its decay products. The lower the p_T of the top quark, the less boosted are its decay products and the larger is the distance between two decay products. The distance between two decay products of the top quark against its transverse momentum has been discussed in C. 4.3. The distance parameter R of the jets needs to be chosen in such a way that in most of the events all three decay products of the top quark are reconstructed in one jet, but not too large to avoid large effects from pile-up on reconstruction level.

Figure 4.1 shows that a distance parameter $R > 1.0$ is needed to get most of the hadronic top quark decays merged in one jet. In the following, different distance parameters are studied using the following basic selection:

- 1 jet with $p_T > 400\,\mathrm{GeV}$ and $|\eta| < 2.5$
- 1 jet with $p_T > 150\,\mathrm{GeV}$ and $|\eta| < 2.5$
- veto on more jets with $p_T > 150\,\mathrm{GeV}$ and $|\eta| < 2.5$
- 1 electron or muon with $p_T > 45\,\mathrm{GeV}$ and $|\eta| < 2.5$

The leading jet with $p_T > 400\,\mathrm{GeV}$ is supposed to be the jet containing the hadronic top quark decay, the second cut is used to select the b-jet from the leptonic top quark decay. The jet veto on additional jets is needed for theory calculations [54], and the lepton cut should ensure an efficient reconstruction in data. The veto affects the number of selected events but has no significant influence on the shape of the jet mass spectrum.

Figure 6.2 shows the jet mass distribution of the leading jet on particle level for events simulated with POWHEG + PYTHIA using different distance parameters for the jets of $R = 0.8, 1.2$ and 1.5. All of these jets are clustered with the Cambridge/Aachen algorithm. For a R-parameter of $R = 0.8$, the leading jet has a tendency to small jet masses. Two peaks beside the top quark mass peak are visible, one at approximately $90\,\mathrm{GeV}$ coming from semi merged top quark decays and one peak at $50\,\mathrm{GeV}$ that probably consists of jets from bottom quarks or radiation. Those jets are a challenge for the theory calculations and should be efficiently suppressed by the selection. Using larger R-parameters leads to more fully merged top quark decays within the distance R. However, the R-parameter should not be chosen too large, because influence of radiation on the jet mass increases with the radius of the jet, leading to a large tail toward high masses, which can be seen in Fig. 6.2 for $R = 1.5$. An R-parameter of $R = 1.2$ shows the best reconstruction of the top quark mass peak with this basic selection and is used for the rest of this thesis.

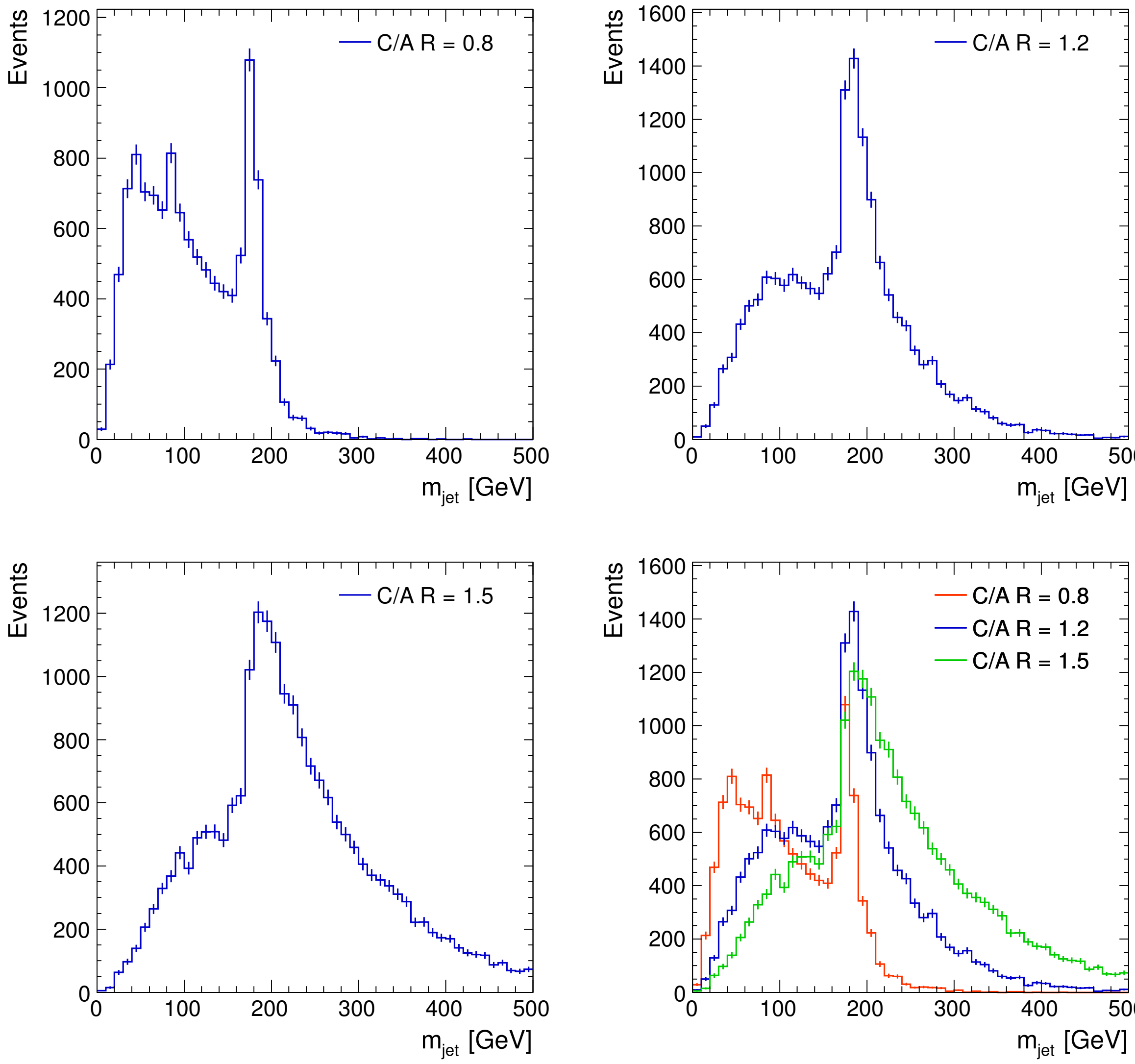

Figure 6.2.: Distribution of the invariant mass of the leading jet for different R-parameters used for the jet clustering in simulated $t\bar{t}$ events. The jets are clustered with the Cambridge/Aachen algorithm. The basic selection described in the text is applied.

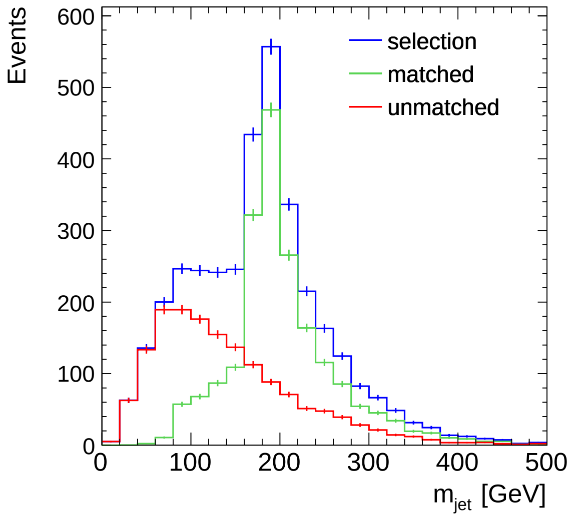

Figure 6.3.: Distribution of the invariant mass of the leading jet after the basic selection compared with the matched and unmatched fraction of this selection.

Further Selection

It is important to test, how many of the jets can actually be matched to the hadronically decaying top quark. This is possible using the information of the MC generator. A jet is matched if its distance to the hadronically decaying top quark ΔR is below the distance parameter of the jet, $\Delta R < 1.2$. Figure 6.3 shows the distribution of the leading jet mass with the full basic selection applied in comparison with the matched and unmatched fractions of the selected events. The distribution of the full selection shows a peak near the top quark mass, that is dominated by fully merged decays, matched to a hadronically decaying top quark. However, there is also a large shoulder to lower jet masses dominated by unmatched events that should be suppressed for the measurement.

The jet mass of the unmatched jets shows the typical shape for QCD-jets (jets induced by light quarks[1] or gluons). In these events the leading jet is formed by one of the two bottom-quarks in the event or by initial or final state radiation. Since these jets are not included in first-principle calculations, a way to suppress their contribution is needed. The matched jets show a peak a bit above than the top quark mass (higher because of additional radiation). Events in the shoulder to smaller jet masses down to 80 GeV are semi-merged events giving masses comparable to the W mass of 80 GeV. These events are also not calculable in fixed-order calculations and need to be suppressed.

To suppress the semi-merged and unmatched jets, two additional cuts are added to the basic selection:

- $\Delta R_{\text{2nd jet, lepton}} < 1.2$
- $m_{\text{leading jet}} > m_{\text{2nd jet + lepton}}$

[1] light quarks at this point means all quarks except for the top quark

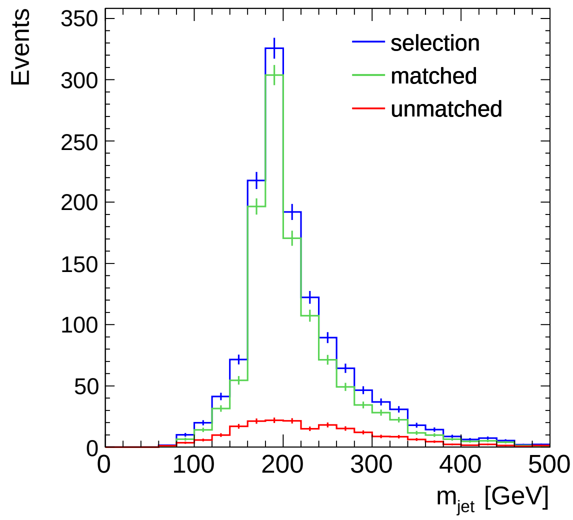

Figure 6.4.: Distribution of the invariant mass of the leading jet after the full selection together with the matched and unmatched fraction of this selection.

The motivation for the first cut is to assume that the leptonically top quark is boosted, to increase the fraction of fully merged hadronically top quarks. It is also experimentally motivated, to get a better description of the particle level on reconstruction level, which will be discussed in more detail in the section 6.3.2. A top quark is boosted if the distance between two decay products of the top quark (in this case the b-quark and the lepton) is smaller than 1.2.

The second cut is used to suppress the semi-merged and unmatched events. Just requiring $m_{\text{leading jet}} > m_{\text{2nd jet}}$ would suppress the unmatched but not the semi-merged events because a W-jet has in most cases still a bigger mass than the b-jet. Adding the lepton to the second jet returns the leptonic top quark decay without the neutrino, which should have a lower mass than the full hadronic top quark decay, but a higher mass than the hadronic W-jets.

The distribution of the jet mass on particle level is shown in Fig. 6.4 after the full selection in comparison with the matched and unmatched fraction of the selection. It becomes clear that the tow additional cuts suppress the influence of the unmatched jets as well as the shoulder due to semi-merged decays for the matched events. The result is a selection including mostly events, where the leading jet contains a full hadronic top quark decay.

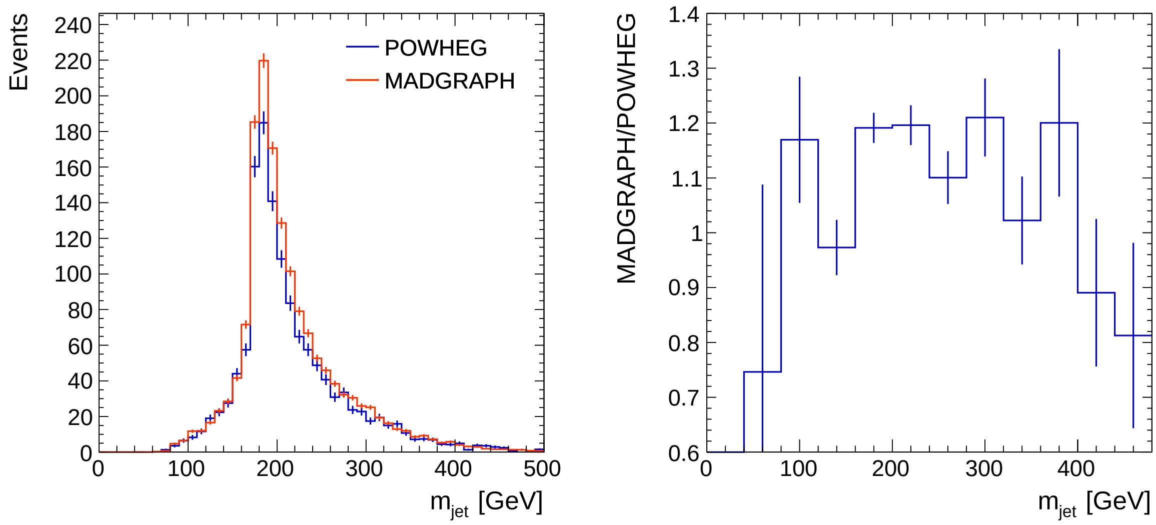

Figure 6.5.: Invariant mass of the leading jet after the full selection simulated with MADGRAPH and POWHEG on the left and the ratio between both distributions on the right.

6.2.2. Different Monte Carlo Generators

Because the final measurement should be independent of the chosen simulation model and therefore on the choice of the Monte Carlo generator, it is interesting to look at differences in the jet mass spectrum for different generators. In this section differences in the jet mass between events generated with POWHEG and with MADGRAPH are analyzed. For this comparison the full selection of the previous section is applied.

Figure 6.5 shows the jet mass of the leading jet after the full selection for POWHEG and MADGRAPH and a ratio between them. The observation is that there are about 10% more MADGRAPH-events than POWHEG-events in the peak region. The excess of MADGRAH in the peak area can be explained by two effects. First, the p_T spectrum of the top quarks is harder in MADGRAPH than in POWHEG, as shown in Fig. 6.6. For a $p_T > 400\,\mathrm{GeV}$ this results in about 10% more MADGRAPH- than POWHEG-events.

Secondly, the p_T of the third jet is softer in MADGRAPH, witch leads to a smaller influence of the jet veto. This second effect is shown Fig. 6.7, where the distribution of the third jet p_T is shown after the full selection except for the veto for $t\bar{t}$ events simulated with POWHEG and MADGRAPH.

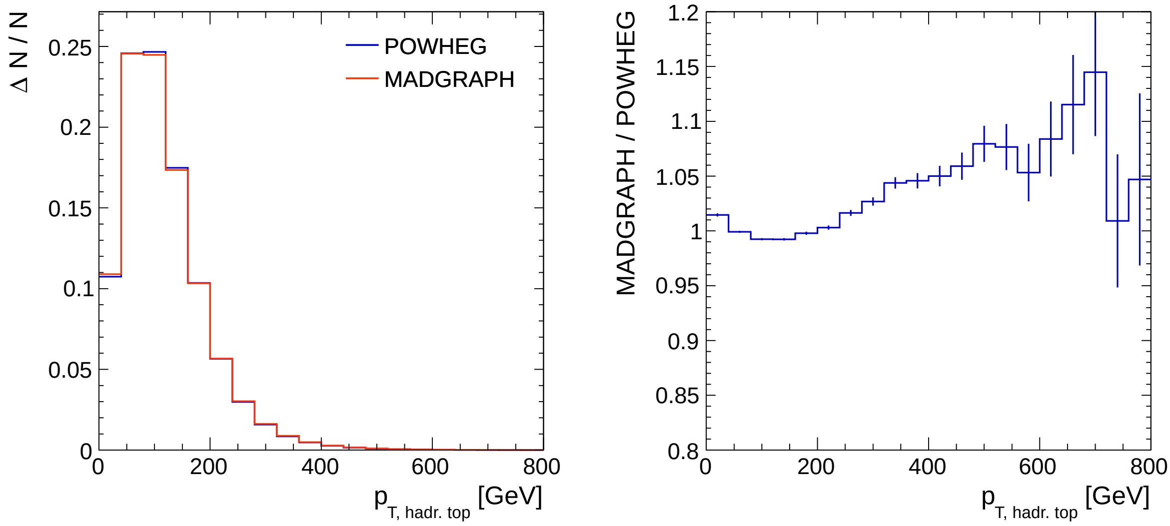

Figure 6.6.: Distribution of the transverse momentum of the hadronically decaying top quark in $t\bar{t}$ events in the lepton+jets channel without selection simulated with MADGRAPH and POWHEG (left) and the ratio between these distributions (right). The distributions are normalized to the total number of events.

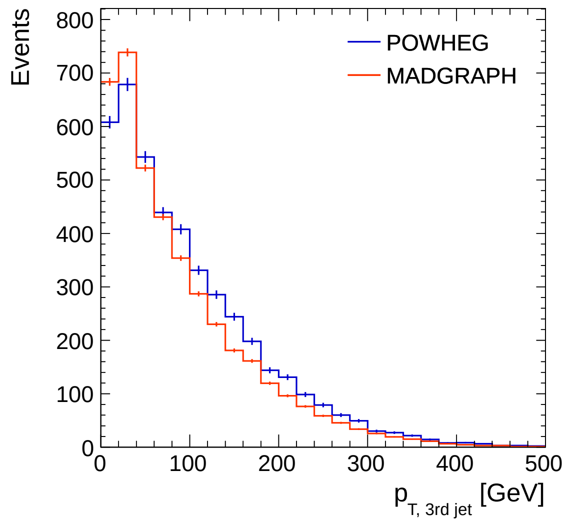

Figure 6.7.: Distribution of the transverse momentum of the third leading jet in p_T after the basic selection without the jet veto applied, simulated with MADGRAPH and POWHEG.

6.3. Reconstruction

In data it is important to get a phase space similar to the particle level phase space, as well as an efficient suppression of the dominant backgrounds. The dominant backgrounds for this analysis are QCD-multi-jet production, production of W-bosons in association with additional jets and single top quark production.

6.3.1. Basic Selection

The basic selection is used to suppress the backgrounds. It is derived from a search for $t\bar{t}$-resonances in reference [55], and was adjusted to this analysis, such that it provides a clean $t\bar{t}$ selection in the lepton+jets channel.

The used data is recorded using non isolated lepton triggers. The trigger for the muon channel requires one muon with $p_T > 40\,\text{GeV}$ and $|\eta| < 2.1$, while the trigger used for the electron channel requires one electron with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.5$. The selection cuts on reconstruction level are listed in Tab. 6.4. At this stage, jets refer to anti-kt jets

electron channel	**muon channel**
at least one good primary vertex	
one electron ($p_T > 35\,\text{GeV}$, $\lvert\eta\rvert < 2.5$)	one muon ($p_T > 45\,\text{GeV}$, $\lvert\eta\rvert < 2.1$)
no muon	no electron
at least two jets with $p_T > 50$ and $\lvert\eta\rvert < 2.5$	
2D cut	
$H_{T,\text{lep}} > 150\,GeV$	
missing $E_T > 20\,GeV$	
triangular cut	
one b-tag (CSV tight working point)	

Table 6.4.: Selection criteria of the basic selection on reconstruction level for the electron and the muon channel.

with a distance parameter of $R = 0.5$. These jets are used for the triangular and 2D, cut described below.

The cut on the lepton is applied to select the lepton+jets channel and suppress QCD multi-jet background. The veto on muons in the electron and electrons in the muon channel avoids an overlap between the two channels. The $H_{T,\text{lep}}$ cut is again used for suppression of QCD as well as the missing E_T cut on $20\,\text{GeV}$, with a non-zero value expected because of the neutrino in the final state. The 2D-cut is used to ensure that the lepton originates form a hard process and not from semi-leptonic decays of charm- and bottom mesons. This helps to suppress QCD-background. The 2D-cut consists of two cuts:

$$\Delta R_{\text{min}}(\text{lepton, jets}) > 0.5\,,$$
$$\text{or}\;\; p_{\text{T, rel}}(\text{lepton, jets}) > 2.5$$

with the minimal distance of the lepton to the closes jet $\Delta R_{\text{min}}(\text{lepton, jets})$ and the fraction of the lepton p_T perpendicular to the nearest jet p_T, $p_{\text{T, rel}}$.

In the electron channel an additional triangular cut is applied to further suppress QCD-background,

$$-\frac{1.5}{75\,\text{GeV}}\,\not{E}_T + 1.5 < \Delta\Phi(\text{lepton},\ \not{E}_T) < \frac{1.5}{75\,\text{GeV}}\,\not{E}_T + 1.5$$
$$-\frac{1.5}{75\,\text{GeV}}\,\not{E}_T + 1.5 < \Delta\Phi(\text{jet}_{\text{leading}},\ \not{E}_T) < \frac{1.5}{75\,\text{GeV}}\,\not{E}_T + 1.5$$

and ensures that $\vec{\not{E}}_T$ does not point along the transverse direction of the leading jet or the lepton. The last selection step is the requirement of a b-tagged jet, which is an important cut to suppress the W+jets background.

6.3.2. Post Selection

After the basic selection a post-selection is applied to obtain a similar phase space on reconstruction level as on particle level In this selection jets are considered to be Cambridge/Aachen jets with $R = 1.2$. The post selection on reconstruction level is similar to the one on particle level. It includes the following cuts.

- one jet with $p_T > 400\,\text{GeV}$ and $|\eta| < 2.5$.
- one jet with $p_T > 150\,\text{GeV}$ and $|\eta| < 2.5$.
- veto on more jets with $p_T > 150\,\text{GeV}$ and $|\eta| < 2.5$.
- one electron (muon) with $p_T > 45\,\text{GeV}$ and $|\eta| < 2.5$ ($|\eta| < 2.1$).
- $\Delta R_{\text{2nd jet, lepton}} < 1.2$.
- $m_{\text{leading jet}} > m_{\text{2nd jet}}$.

The last cut does not include the lepton as on particle level and compares just the masses of the two leading jets. These cuts have already been discussed in detail in section 6.2. Just the ΔR cut between the second jet and the lepton is motivated on reconstruction level and therefore mentioned again in this section.

It requires a boosted leptonically decaying top quark in order to obtain more boosted hadronically decaying top quarks without any direct requirements on the hadronic side, which could bias the measurement. Figure 6.8 shows the distribution of the leading jet

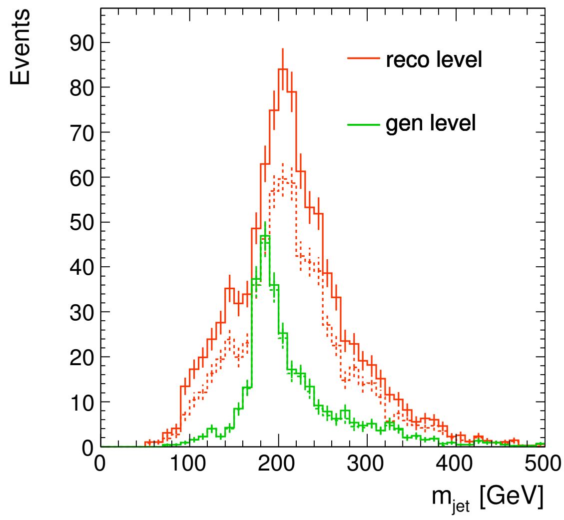

Figure 6.8.: Jet mass distribution on particle level for events passing the particle level and reconstruction level selection, compared to events on reconstruction level passing just the reconstruction level selection. The solid lines indicate the full selection without the cut on $\Delta R_{\text{2nd jet, lepton}} < 1.2$. For the distributions displayed with dotted lines the full selections with this cut are applied.

mass for events on particle level, passing the particle and the reconstruction level selection, and on reconstruction level passing just the reconstruction level selection. It is visible, that this cut has just a minor influence on the particle level selection after requiring $m_{\text{leading jet}} > m_{\text{2nd jet}}$, but a bigger influence on reconstruction level. It increases the purity of the selection defined as $\frac{N_{\text{rec+gen}}}{N_{\text{rec}}}$, where $N_{\text{rec+gen}}$ is the number of events passing the particle level and the reconstruction level selection and N_{rec} the number of events just passing the reconstruction level selection. A high purity is important for a stable unfolding and therefore this cut improves the performance of the unfolding.

Control distributions for the electron channel after the full selection comparing data and Monte Carlo simulation are shown in Fig. 6.9. The number of b-jets and the distributions of p_T of the first three jets in data agree with simulation within the statistical uncertainties. The $H_{T,\text{lep}}$ and the missing E_T distributions show sightly bigger differences between data and simulation for small values of $H_{T,\text{lep}}$ and missing E_T, but still a reasonable agreement. Simulation and data mostly agree within the statistical uncertainties. All MC samples are normalized to an integrated luminosity of 19.7 fb^{-1} and the $t\bar{t}$-samples are additionally scaled by a factor of 0.75. This scaling is needed, because the p_T spectrum of the top quark is expected to be softer in data compared to simulation. This effect has been seen for a top quark p_T lower 400 GeV in references [56, 57].

Similar control distributions for the muon channel can be found in the appendix A.

The jet mass distribution in data compared to simulated events for $t\bar{t}$-samples generated with POWHEG and MADGRAPH after the full selection is shown in Fig. 6.10. It is clearly visible that the selection developed in simulation performs equally well in data.

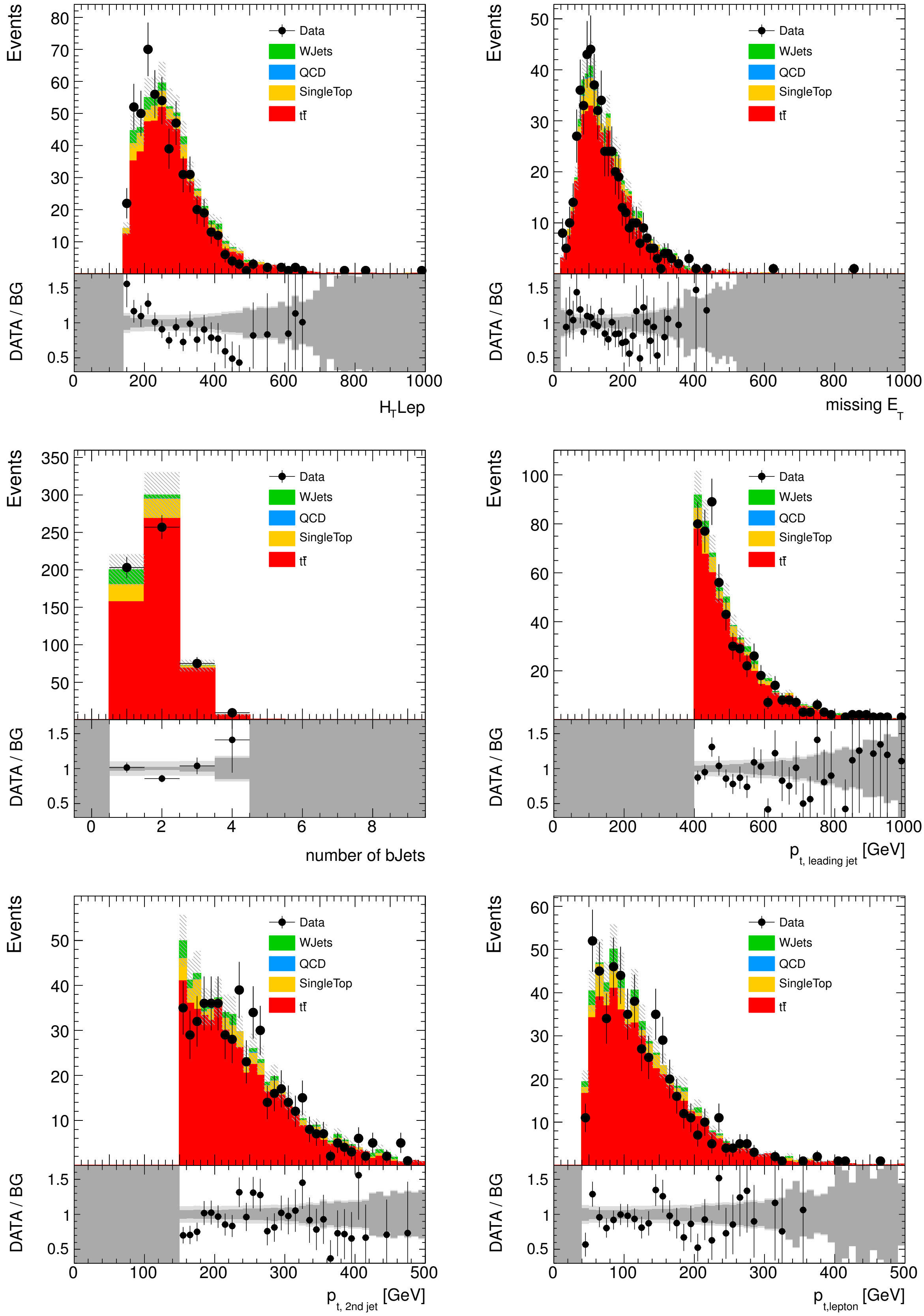

Figure 6.9.: Control distributions after the full selection in the electron+jets channel. The $t\bar{t}$-samples are simulated with POWHEG plus PYTHIA and scaled by a factor of 0.75.

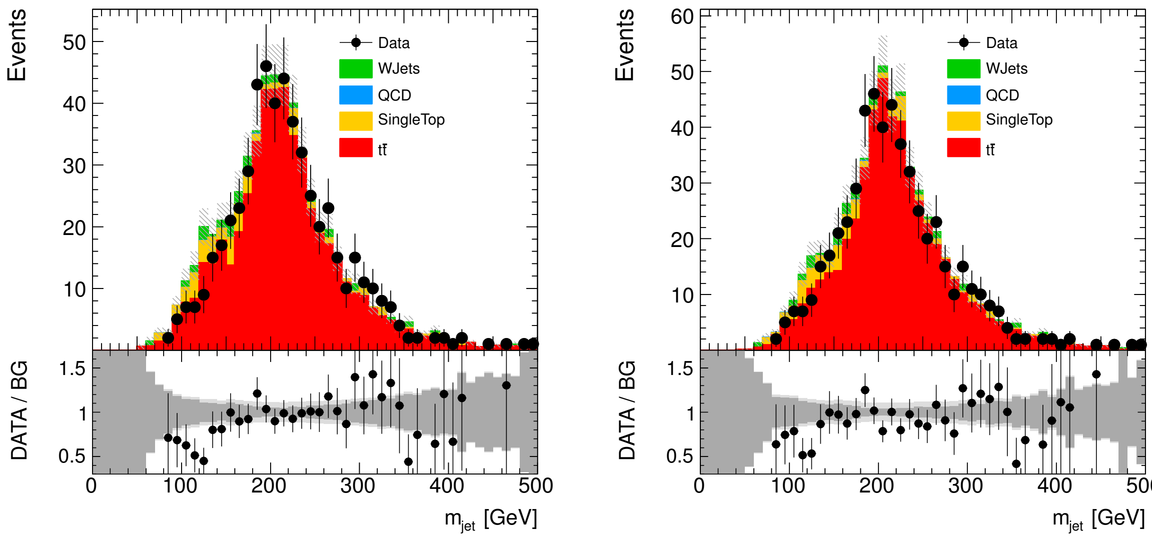

Figure 6.10.: Distribution of the invariant mass of the leading jet after full selection for $t\bar{t}$ simulated with POHEG on the left and MADGRAPH on the right.

It selects a phase space dominated by events in which the leading jet includes all decay products of a hadronic top quark decay. After the full selection, the QCD-background is negligible and will not be included in the unfolding. Overall the jet mass is equally well described by POWHEG and MADGRAPH. POWHEG is chosen as the default sample for the unfolding, since the p_T distribution of the top quark is better described.

A comparison between the jet mass in simulated events on reconstruction and particle level is shown in Fig. 6.11, for events passing the particle and reconstruction level selection. The peak of the jet mass spectrum on reconstruction level is shifted to higher masses compared to the particle level. The reason for this is pile-up, which is included on reconstruction level but not on particle level. The large distance parameter $R = 1.2$ of the jets result in a large pile-up dependence due the clustering of particles from pile-up processes into the jets.

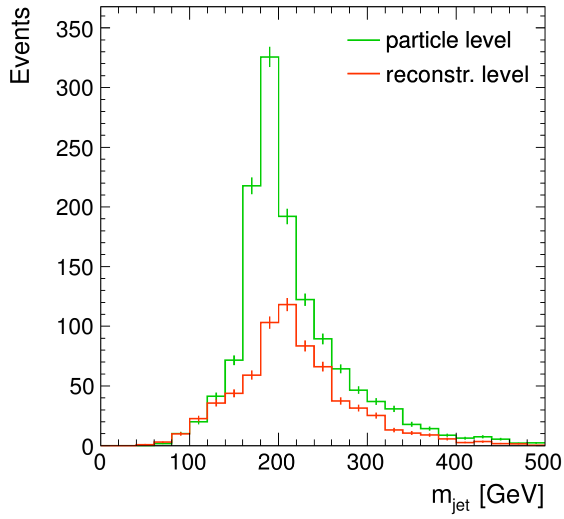

Figure 6.11.: Comparison between the jet mass distribution of the leading jet on particle and reconstruction level for events passing both particle and reconstruction level selections.

6.3.3. Grooming Studies

In order to get a better matching between reconstruction and particle level and to damp the influence of pile-up, the three grooming techniques described in chapter 4.2.4 have been studied, using the recommended values in the references [36, 43, 44].

- pruning: $D_{\text{cut}} = m_{\text{jet}}/p_{T,\text{jet}}$ and $z_{\text{cut}} = 0.1$
- filtering: $R_{\text{filter}} = 0.3$ and $N_{\text{filter}} = 3$
- trimming: $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 0.03$

Figure 6.12 shows the invariant mass of the leading jet in simulated $t\bar{t}$ events passing the full selection for jets groomed with one of the the methods mentioned above and ungroomed. As expected, the groomed distributions show a more narrow peak around the top quark mass at lower mass values.

Using pruning, some events reconstructed near the top quark mass before grooming are now reconstructed at low masses. This means that reconstructed particle candidates, originating from the decay of the top quark have been filtered out by the grooming algorithm. The shape of the mass spectrum changes dramatically, and in an undesired way. Trimming was chosen for the following optimization.

For a good measurement a good reconstruction of the events used is important. Therefore the trimming was optimized to get a narrow pull distribution $(m_{rec} - m_{gen})/m_{gen}$, peaking at zero. The best trimming parameters found are $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 0.003$. The pull distributions are shown in Fig. 6.13 and a comparison between reconstruction and particle level for events passing both particle and reconstruction level selection is

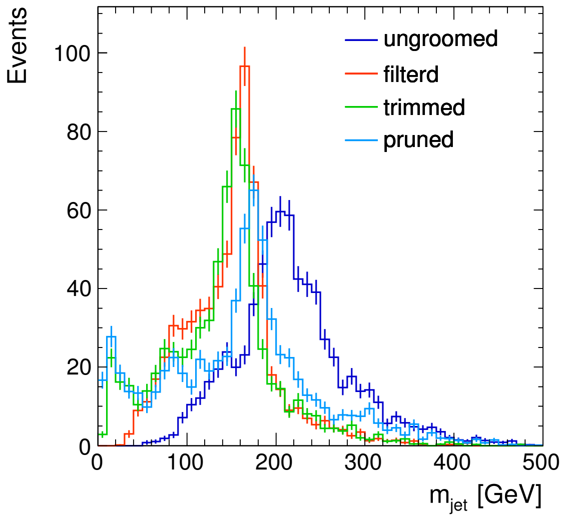

Figure 6.12.: Comparison of the ungroomed jet mass of the leading jet after full selection on reconstruction level with distributions obtained using different grooming methods

shown in Fig. 6.14 with and without trimming. For the last plots the trimming was applied before the post selection.

However, it has been found that trimming leads to model dependent effects in the unfolding. For this reason trimming is not used in the final measurement, but should be studied for the final set up with respect to all systematic uncertainties again in future studies.

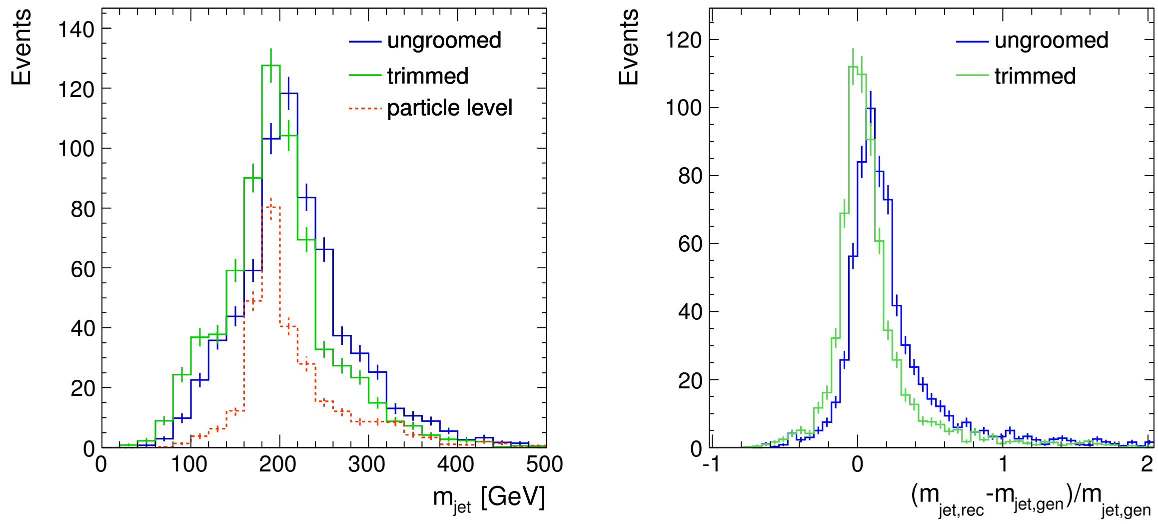

Figure 6.13.: Distribution of the invariant mass of the leading jet for optimized trimming compared with the ungroomed distribution and a particle level distribution. The particle level distribution includes events passing particle level and reconstruction level selection (left). Pull distributions between reconstruction and particle level jet masses for particle and reconstruction level selections applied for groomed and ungroomed reconstruction level (right).

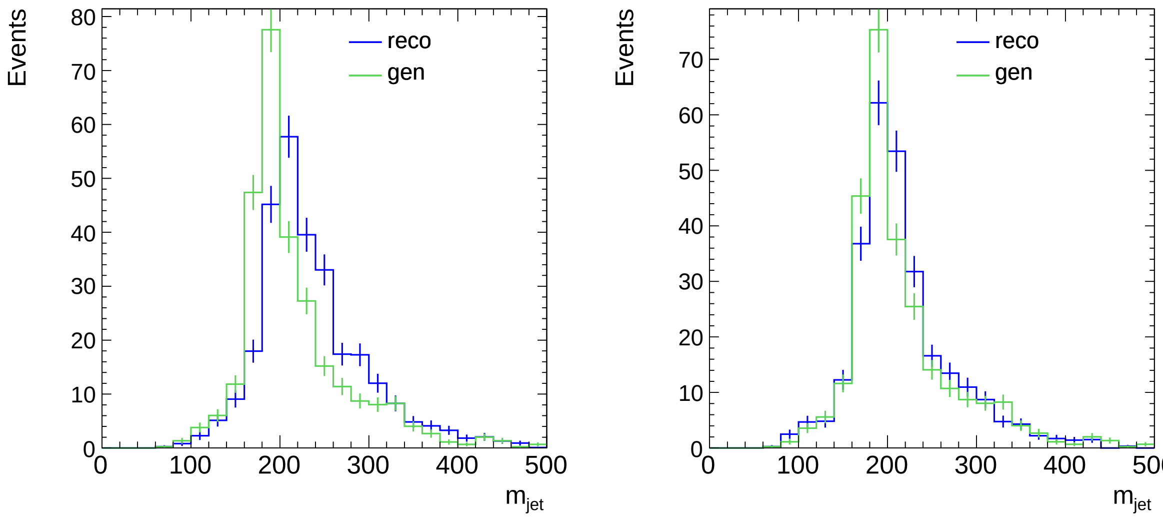

Figure 6.14.: Distribution of the invariant mass of the leading jet for events passing particle level and reconstruction level selection for groomed (right) and ungroomed (left) jets on reconstruction level, compared to the particle level jet mass. In this case the grooming was applied before the post selection.

6.3.4. Mass of light jets in Top Anti-Top decays

The reconstruction of the jet mass is essential for this analysis. Since jets with a distance parameter of $R = 1.2$ have not been studied in CMS, a study of the jet mass reconstruction is performed here. If the jet mass reconstruction is not well described by simulation, an uncertainty on the jet mass scale would have to be added to the final result.

A method to study the jet mass scale selects jets containing just the hadronic W-decay within the $t\bar{t}$ decays on reconstruction level and compares data to simulation. The mass of these jets should be centered around the W boson mass. Several attempts have been made to obtain the W-peak in $t\bar{t}$-decays.

The first selection is inspired from a selection in reference [58] providing a good selection of hadronic W-jets in $t\bar{t}$-decays for jets with $R = 0.8$. It selects at least three jets with $p_T > 150\,\text{GeV}$ and denotes the jet with the highest mass to be the W-jet candidate. The following additional criteria have to be fulfilled,

- one b-tag (CSV medium working point),
- $\Delta R(\text{jet}, b - \text{jet}) > 1.2$,
- $\Delta R(\text{jet}, \text{lepton}) > \frac{\pi}{2}$,
- $\Delta\phi(\text{jet}, \not{E}_T) > 2.0$.

Figure 6.15 shows the selection and the matched fraction of jets for the jets with $R = 1.2$ used in this analysis. The fraction of matched jets at the W-mass is below 50 %. The selection bases on the assumption, that the W-jet is the jet with the highest mass in the event. However for a distance parameter of $R = 1.2$, jets originating from light quark decays gain mass by collecting particles from initial state radiation, final state radiation and pile-up. Together with the p_T cut on the leading jet of $p_T > 400\,\text{GeV}$ this leads to the effect, that the mass of jets including a hadronic W-decay does not differ significantly from the mass of a jet originating from a light quark decay (including bottom-quarks), making this selection inefficient for jets of this size. Hence the masses of light quark jets and W-jets are peaking at almost the same mass, making it impossible to separate them efficiently.

A second selection is based on the assumption that there should be at least three jets in the event, one from the hadronic W-decay and two b-jets from the top quark decays. After demanding at least three jets with p_T> 150 GeV and at least two b-tags with a loose CSV working point the not b-tagged jet should be the hadronic b-jet. The selection is shown in Fig. 6.16 and suffers from the same problems as the previous selection. The high mass of light quark jets with a distance parameter of $R = 1.2$ makes it difficult to reconstruct the W-peak in $t\bar{t}$-decays on a falling background, which would allow an estimation of the jet mass scale. Therefore it can not be done in this thesis.

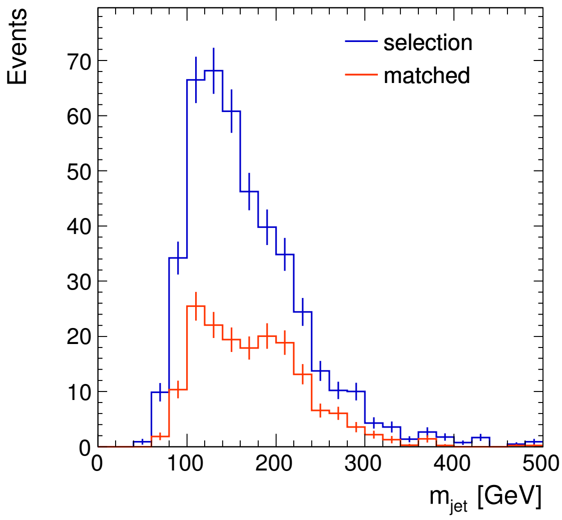

Figure 6.15.: Distribution of the invariant mass of the jet with the highest mass for events selected by the W-tagging selection described in reference [58] in comparison with the fraction of jets matched to a hadronic W boson.

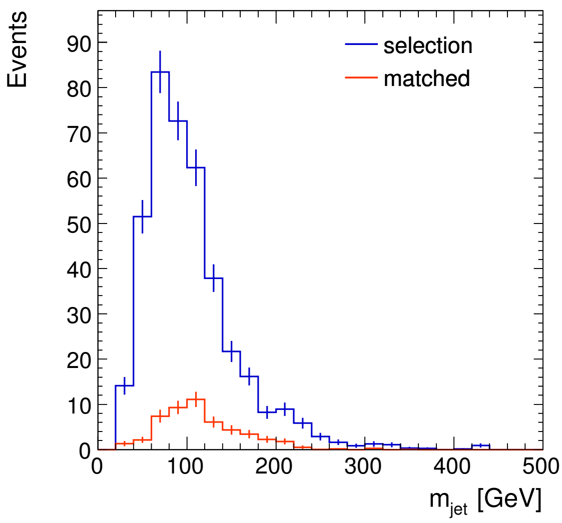

Figure 6.16.: Distribution of the invariant mass of the one of the tree leading jets which has not been b-tagged in comparison with the fraction of this selection matched to a hadronic W boson.

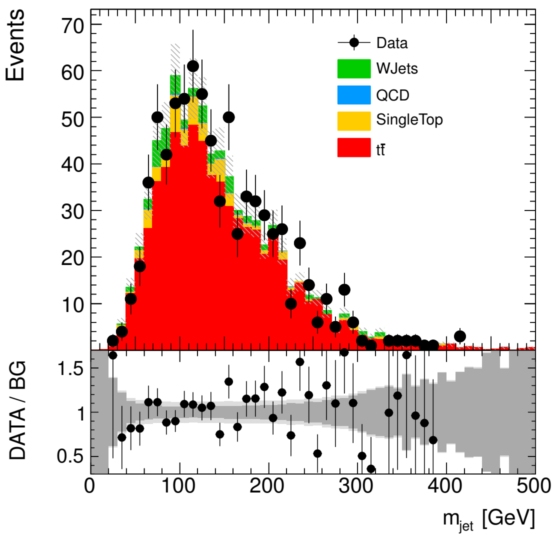

Figure 6.17.: Distribution of the invariant mass of the leading jet after the basic selection and requiring three jets with $p_T > 150\,\text{GeV}$.

In order to test the reconstruction of the jet mass of light jets in $t\bar{t}$-decays, data is compared to simulation in Fig. 6.17. For this comparison three jets with $p_T > 150\,\text{GeV}$ are required in addition to the basic selection. This results in a statistically independent sample from the main analysis. Data and simulation agree within the statistical uncertainties, and no further uncertainty is applied due to the jet mass scale.

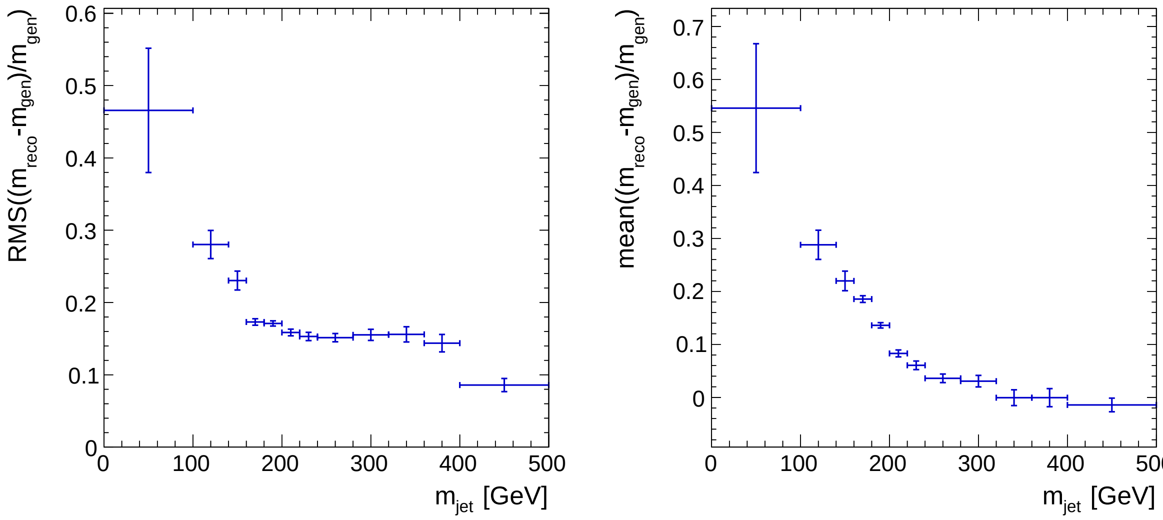

Figure 6.18.: Resolution (left) and mean relative difference between the mass of jets on reconstruction and particle level (right) for the leading jet as a function of the jet mass on particle level.

6.4. Unfolding

6.4.1. Binning

The choice of the binning on particle level for the unfolding is driven by the reconstruction resolution of the jet mass. The resolution is determined as the root mean squared (RMS) of the pull distribution $(m_{\text{rec}} - m_{\text{gen}})/m_{\text{gen}}$ and shown in Fig. 6.18 as well as the mean of this distribution as a function of m_{jet}.

For small masses the resolution is only about 50 %, because the mass of the jets is dominated by radiation and highly influenced by pile-up on reconstruction level. The same effects lead to a reconstruction of events with small masses on particle level at higher jet masses on reconstruction level. The bin width on particle level is chosen to be roughly one standard deviation of the pull distribution for a given mass. Smaller bin widths would lead to larger correlations between neighboring bins of the final measurement. The binning on reconstruction level is chosen to avoid overlap of bin borders between particle and reconstruction level to suppress border effects and to have a reasonable number of events entering the unfolding. Figure 6.19 shows the distribution of the leading jet mass on particle level and on reconstruction level with the chosen binning in $t\bar{t}$ events simulated with POWHEG+PYTHIA. Both distributions show a peak around the top quark mass and a reasonable number of events in each bin for the unfolding.

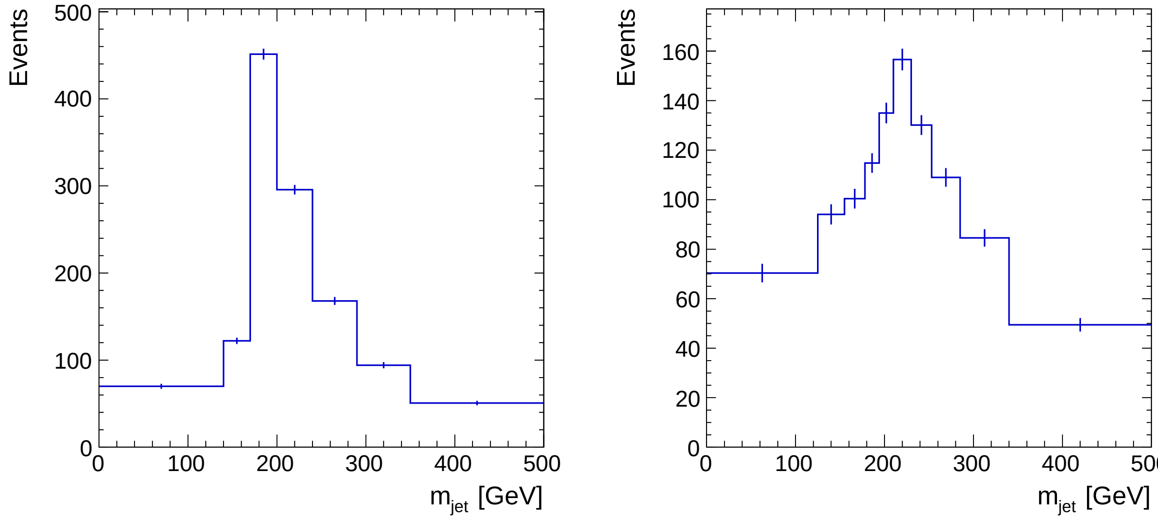

Figure 6.19.: Distribution of the invariant mass of the leading jet on particle level (left) and reconstruction level (right) after the full selection.

6.4.2. Determination of the Migration Matrix

The migration or response matrix includes the possibilities that an event that has a mass of the leading jet on particle level in a certain bin on particle level is reconstructed in one of the bins on reconstruction level. It is used to correct the reconstructed distribution for detector effects. In this analysis the migration matrix is determined from MC simulation. Rows (x-axis) contain the generator distributions and columns (y-axis) the reconstructed distributions. A whole row should contain 100 % of the generated events in this generator level bin. This is important when applying reconstruction weights to correct for differences between efficiencies in MC and data (for example because of b-tagging, jet energy resolution, and more).

Events passing the reconstruction selection but not the particle level selection can be seen as background for this measurement because they do not belong to the desired phase space. These events are included in the migration matrix instead of being subtracted before the unfolding like other background processes, because subtracting those would lead to a dependence of the shape of the generated distribution and in this way of the top quark mass that was used in the simulation. This influence can be avoided by including the $t\bar{t}$-background in the unfolding.

Normalized response matrices determined in POWHEG and MADGRAPH plus PYTHIA are shown in Fig. 6.20. Both response matrices are very similar and show only small off-diagonal entries. The effect of small masses reconstructed too high, as already seen in Fig. 6.18, is visible in the first bin of the response matrices.

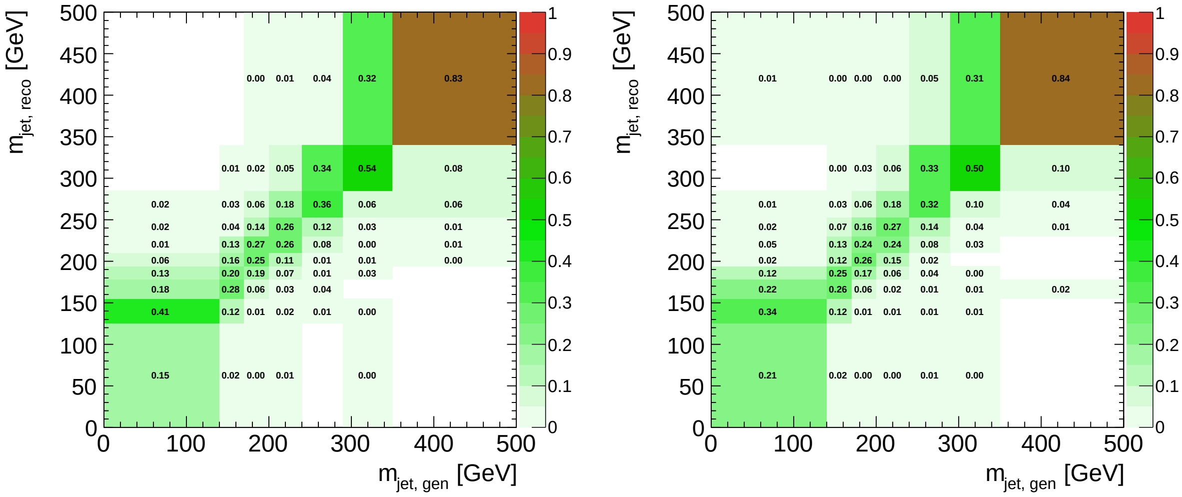

Figure 6.20.: Normalized response matrices obtained with POWHEG (left) and MADGARPH (right). The contents are normalized to the number of events passing the reconstruction and the particle level selection.

6.4.3. Purity, Stability and Efficiency

Before performing the unfolding a few quantities are derived that allow to estimate the average size of corrections due to efficiencies and migrations. The three quantities are defined as:

$$\text{purity} = \frac{N_{\text{rec+gen}}}{N_{\text{rec}}} \tag{6.1}$$

$$\text{stability} = \frac{N_{\text{rec+gen}}}{N_{\text{gen}}} \tag{6.2}$$

$$\text{efficiency} = \frac{N_{\text{rec}}}{N_{\text{gen}}}, \tag{6.3}$$

where N_{rec} is the number of events passing the reconstruction selection, N_{gen} the number of events passing the particle level distribution and $N_{\text{rec+gen}}$ the number of events passing both distributions.

These three quantities should be sufficiently high to ensure a stable unfolding. All three quantities are evaluated for the selection using POWHEG + PYTHIA generated events and shown in Fig. 6.21 as a function of the invariant mass of the leading jet using the binning that is used on particle level. They are sufficiently high, except for the purity in the first bin, which is so low, that the unfolding cannot be trusted in this bin. The measurement in this bin will therefore be omitted for the final result. The bin is still included in the unfolding to account for migrations into other bins.

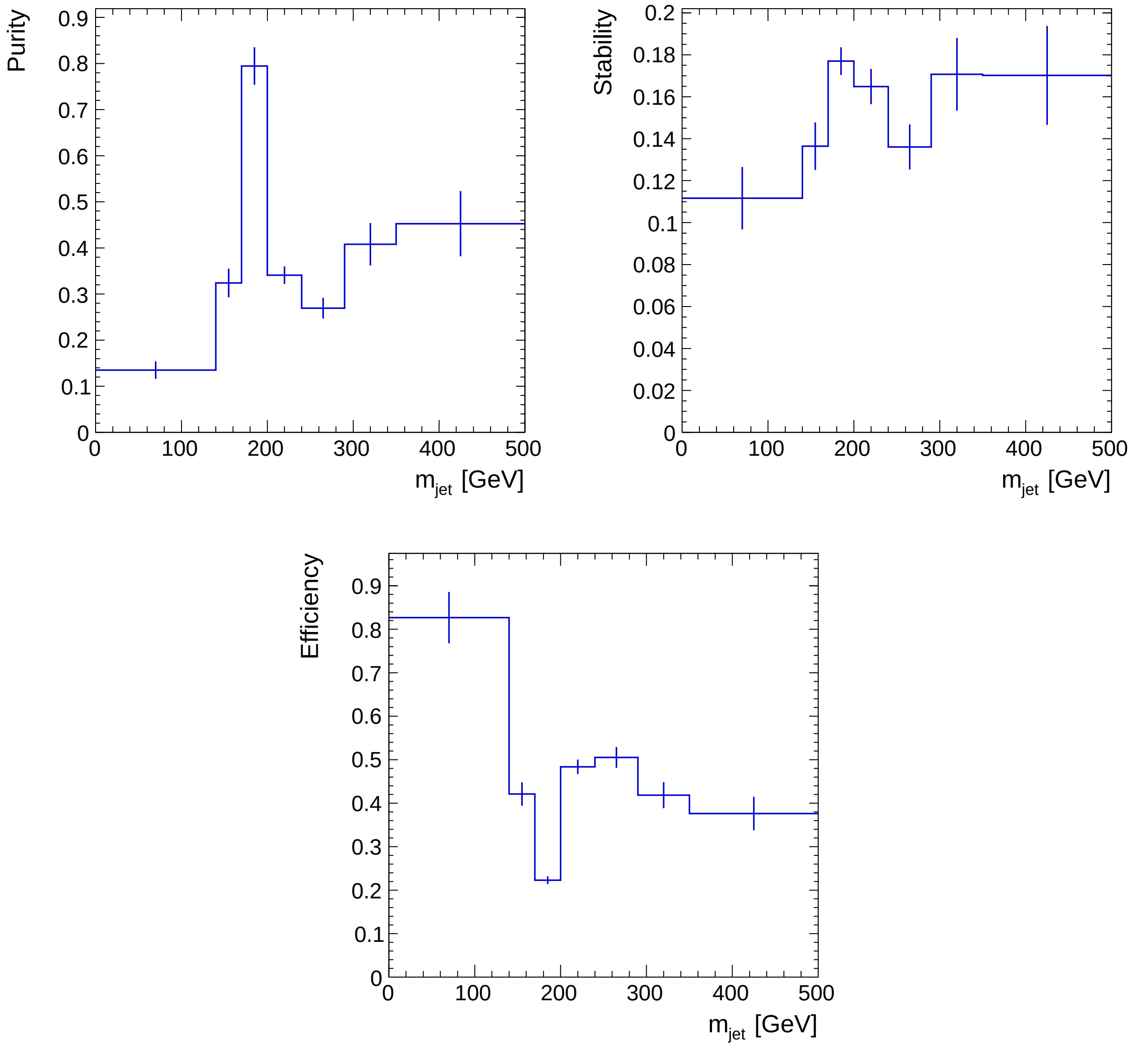

Figure 6.21.: Purity (top left), stability (top right), and efficiency (bottom center) as a function of m_{jet}, calculated for events simulated with POWHEG.

6.4.4. Regularization

Regularization is an important method to damp fluctuations amplified by the minimization process within the unfolding. The regularization conditions and the regularization strength τ has to be chosen carefully, to get an efficient damping. The form of the L matrix determines if the regularization is applied with respect to the size of each single bin, to the first derivative or the curvature (second derivative) of the distribution. For the first tests a size regularization is used. In this case the L matrix is just a unity matrix for all bins inside the particle level phase space.

Determination of the τ-Parameter

Two methods to determine the ideal τ parameter have been introduced in section 5. These methods are now tested for this specific unfolding. Figure 6.22 shows an unfolding of events generated with MADGRAPH, and unfolded with the migration matrix determined from POWHEG. Two different τ-parameters, determined with an L-curve scan and a minimization of global correlation coefficients, are used. Both scans are shown in the same figure below the distributions.

For size regularization the regularization term acts just on single bins, while regularization using derivatives acts on differences between different bins. For this reason the usage of derivatives is expected to induce larger correlations to the result and thus leads to a better performance of the minimization of the average global correlation coefficient. However, for a peaking spectrum size regularization should lead to correlations, large enough to get a minimum in the scan over the average global correlation coefficient. Figure 6.22 shows indeed, that this scan has a minimum and returns lower statistical errors and less statistical fluctuations than the L-Curve scan. Another advantage of this method is the possibility to include systematic uncertainties in the determination of the global correlation coefficients and therefore in the determination of the τ-parameter. Systematic uncertainties as described in section 6.5, up to the background subtraction, are already included in the scan shown in Fig. 6.22. The L-curve scan does not provide this possibility.

The minimization of the average global correlation coefficient is chosen to be the default method for all following tests and for the unfolding of the data. It is always checked, if the scan has a well defined minimum.

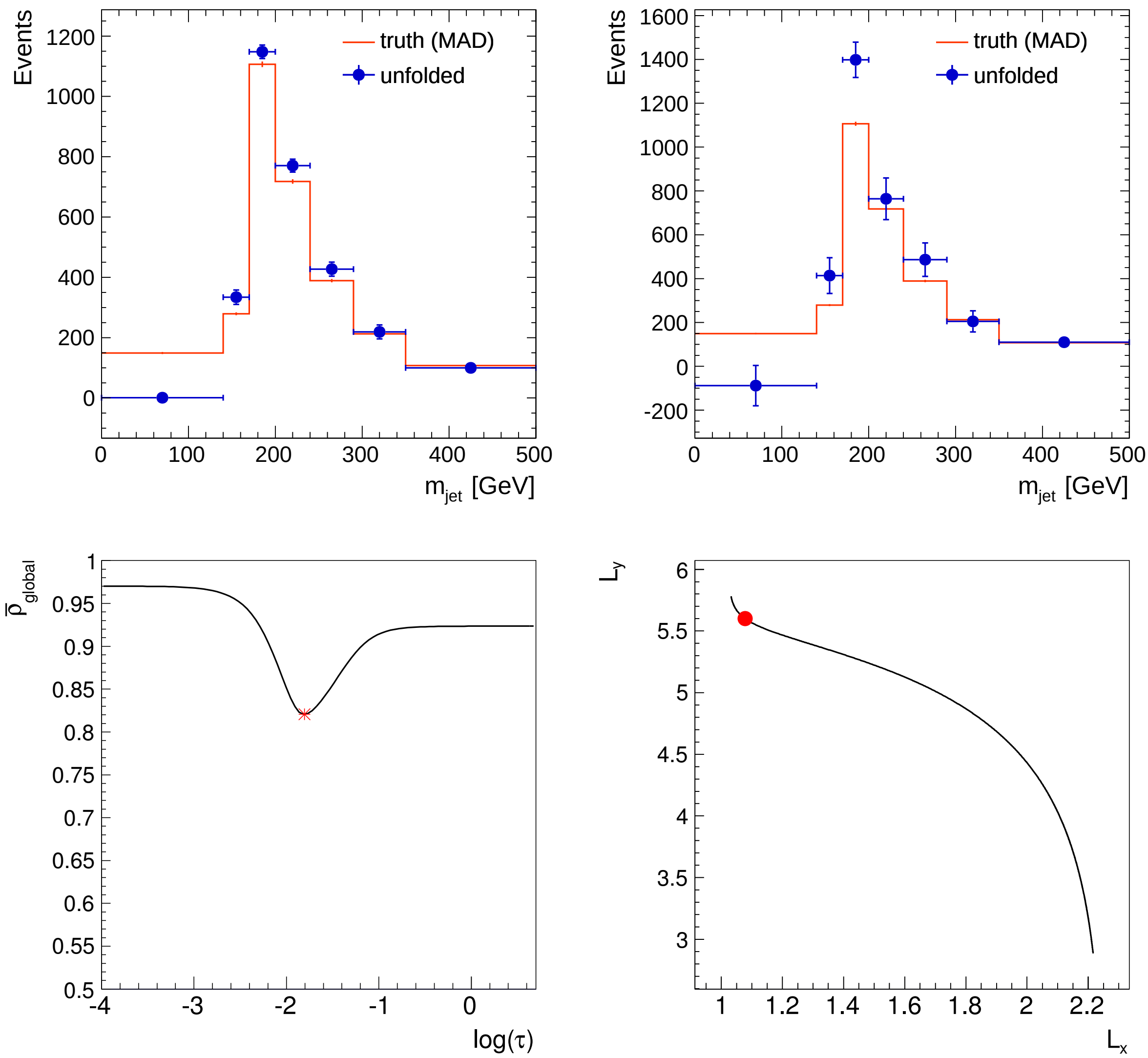

Figure 6.22.: Unfolded distribution of the jet mass in events simulated with MADGRAPH, unfolded with migrations taken from POWHEG and compared to the particle level in MADGRAPH. On the top left a scan over global correlation coefficients was used and on the top right an L-curve scan. The corresponding scan curves can be found below the distributions. The red markers shows the tau value that has been determined by the scans.

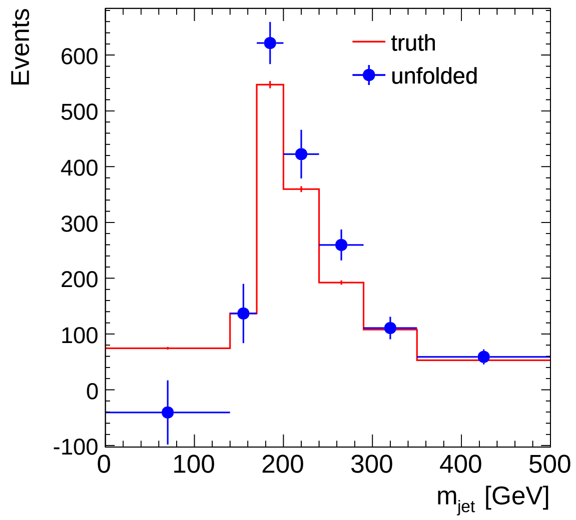

Figure 6.23.: Distribution of the jet mass in events simulated with MADGRAPH and unfolded with migrations of POWHEG using a regularization to the first derivative with a scan over the average global correlation coefficient. For the peak bin only the first derivative with respect to the right bin is used.

Regularization conditions

A scan of global correlation coefficients should perform better with first or second derivative regularization, because the regularization acts on derivatives between bins and should induce negative correlations between different bins leading to a minimum in the average global correlation coefficient. Regularization using the first or second derivative is not a good choice for a peaking spectrum like in this analysis. For the bin in the peak, the first derivative is not well defined numerically. In Fig. 6.23 the unfolding output using the first derivative is shown. For the peak bin just the first derivative with respect to the right bin is considered. Regularization using the first derivative introduces a bias. The first derivative is much lager near the peak compared to the tails of the distribution, leading to a strong dependence of the result in the peak area on the regularization strength. Figure 6.23 shows that this leads to a good matching between output and truth in the tails where the low first derivatives are numerically small and large difference in the peak area, where they are large. The same arguments hold for regularization using the second derivative. For this reason the size regularization has been chosen for this analysis.

6.4.5. Unfolding Tests

The first and most basic test is an unfolding of a reconstruction level jet mass spectrum with a migration matrix determined using the same sample. In this case the unfolding should return exactly the initial distribution. Any differences would indicate basic problems in the unfolding set up. For this test migrations and the reconstruction distribution have been determined with POWHEG + PYTHIA. The result in Fig. 6.24 shows perfect

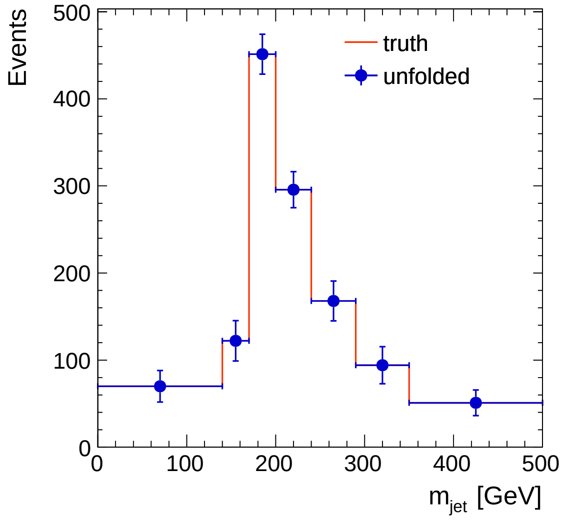

Figure 6.24.: Jet mass distribution in events simulated with POWHEG unfolded with migrations of POWHEG in comparison with the particle level in POWHEG

agreement between the true distribution and the unfolded output. The statistical uncertainties shown are not correct, because the migration matrix is fully correlated to the reconstructed input distribution and the particle level distribution it is compared with.

Dependence on the MC generator

To determine that the unfolding is independent of the underlying assumptions, reconstructed distributions, simulated with a specific generator, are unfolded using migrations determined with a different generator. In this way the dependence of the unfolding on the used Monte Carlo generator can be tested. Differences between the output and the particle level distribution have to be considered as a systematic uncertainty. Figure 6.25 shows unfolded distributions of the jet mass in events simulated with MADGRAPH or POWHEG and unfolded with migrations determined with the other generator respectively. The unfolded distributions are shown together with the jet mass distributions on particle level simulated using the same generator. Except for the first bin, which has low purity, the deviations between the truth distributions and the unfolded ones are small, showing that the unfolding is independent on the assumption of the underlying distribution. Differences in the case of MADGRAPH unfolded with POWHEG are added as a systematic uncertainty in the final result.

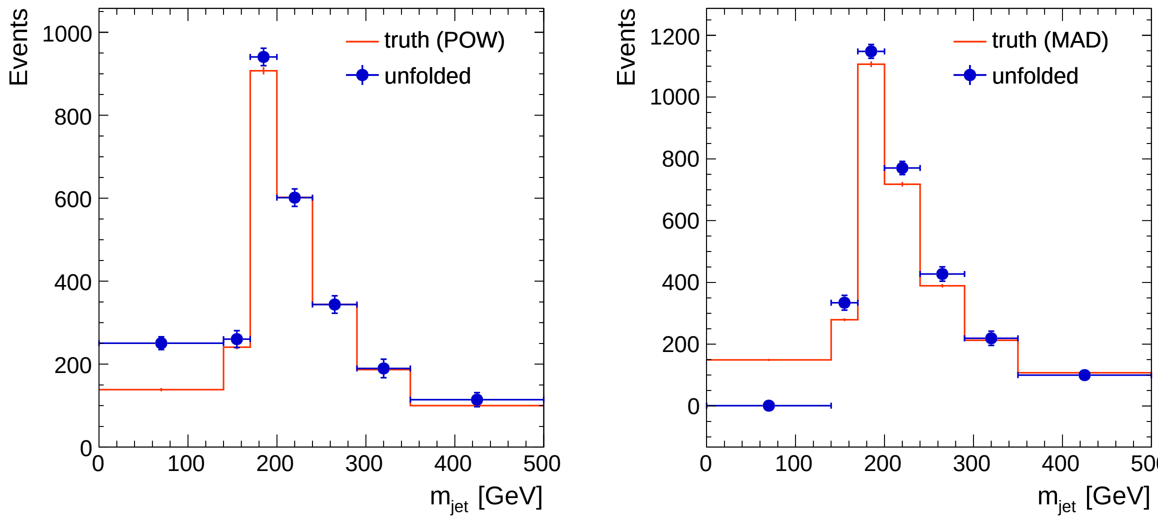

Figure 6.25.: Unfolded jet mass distributions in events simulated with MADGRAPH with migrations of POWHEG in comparison with the particle level in MADGRAPH on the right and the other way around on the left.

Influence of renormalization and factorization scale variations

It is further important to test the influence of renormalization and factorization scales on the unfolding. Variations of these scales change the kinematics of the signal processes and thus the jet mass distribution. In the ideal case the unfolding should be insensitive to the choice of the scale, otherwise changes in the result will be considered as a systematic uncertainty. Samples produced with the factorization and renormalization scales varied by factors of 2 and 0.5 are unfolded by using migration matrices obtained from samples using the default scale choices. These samples were produced only for $m_{t\bar{t}}$ >700 GeV and are therefore unfolded with migrations in POWHEG for $m_{t\bar{t}}$ >700 GeV. Figure 6.26 shows unfolded distributions of samples with varied scales unfolded with the default POWHEG sample for $m_{t\bar{t}} > 700$ GeV. A possible reason for a large bias for the down variation can be found in Fig. 6.27, where the distributions of the first three jet p_Ts and the jet mass of the leading jet is shown. While these distributions for the up variation are very similar to the default sample, the distributions for the down variation differ much more. This leads to different corrections between particle and reconstruction level and in this way to a bias in the unfolding. Like before, differences between truth and unfolded distribution have to be considered as an uncertainty.

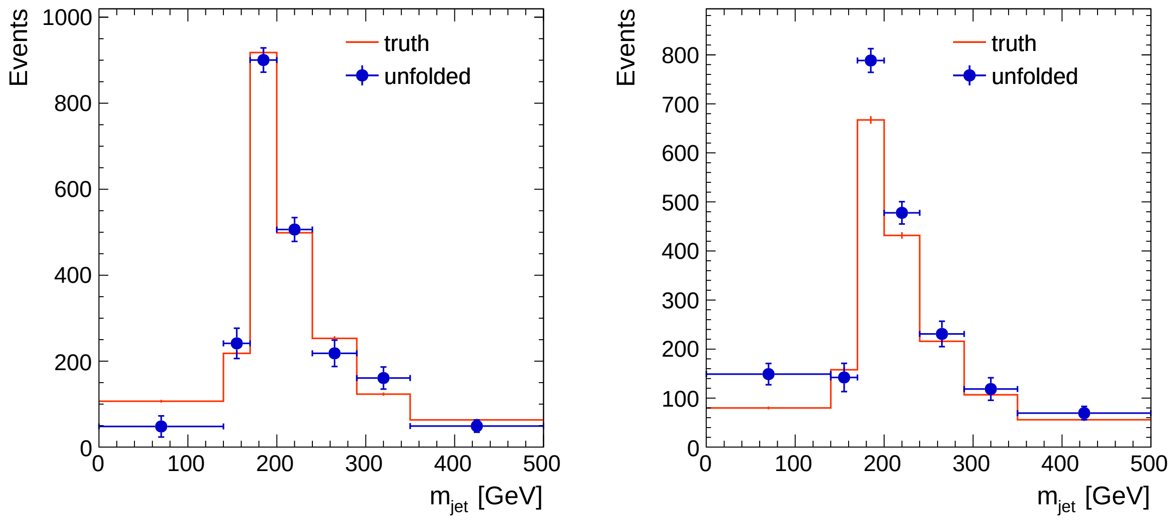

Figure 6.26.: Jet mass distributions in events simulated with POWHEG with factorization and renormalization scales varied by factors of 2 (left) and 0.5 (right) unfolded with migrations determined with the default POWHEG sample for $m_{t\bar{t}} > 700\,\text{GeV}$. The unfolded distributions are shown together with the particle level distributions simulated with the same generator.

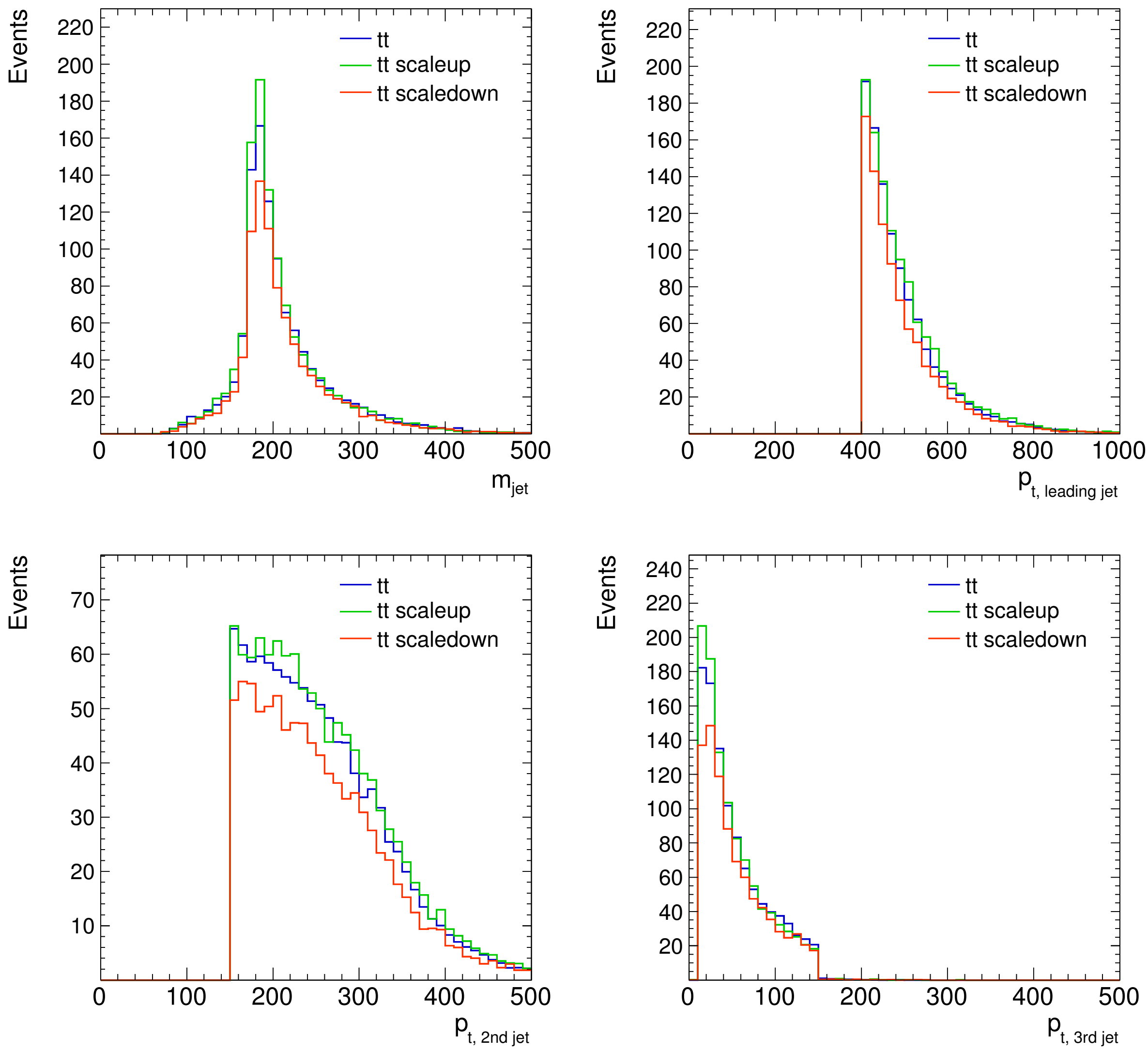

Figure 6.27.: Control distributions for POWHEG samples produced with up (left) and down (right) varied renormalization and factorization scales compared to the default POWHEG sample.

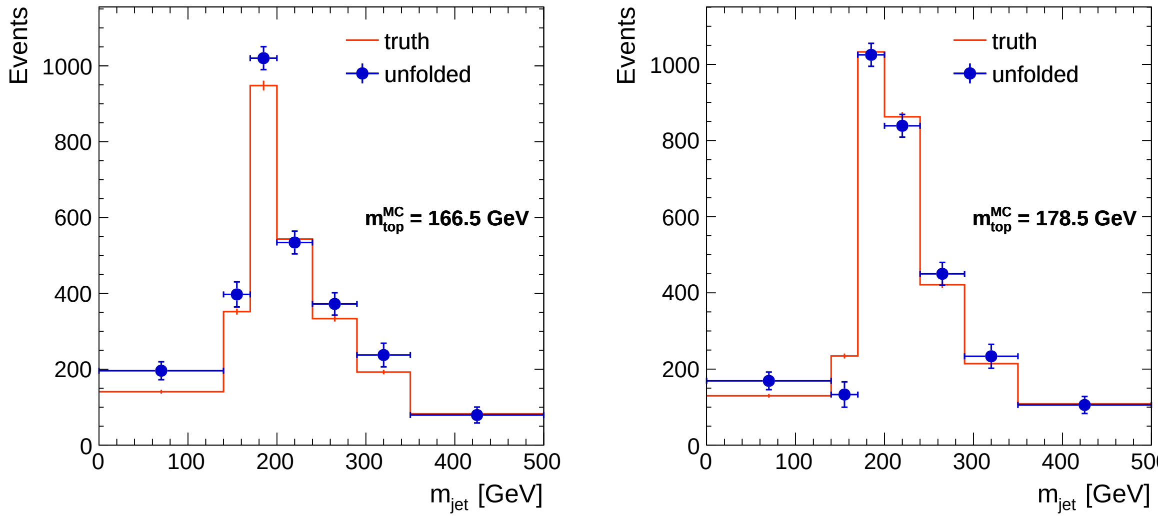

Figure 6.28.: Jet mass distributions in events simulated with MADGRAPH with different values of $m_{\mathrm{top}}^{\mathrm{MC}}$ unfolded with migrations determined with the central MADGRAPH sample with $m_{\mathrm{top}}^{\mathrm{MC}} = 172.5\,\mathrm{GeV}$. The unfolded distributions are shown together with the particle level distributions simulated with the same values of $m_{\mathrm{top}}^{\mathrm{MC}}$.

Dependence on the top quark mass

Because of the sensitivity of the jet mass to the top quark mass it is important to verify that the unfolding is unbiased with respect to the choice of $m_{\mathrm{top}}^{\mathrm{MC}}$ used to obtain the migration matrix. To perform this test, samples generated with different choices of $m_{\mathrm{top}}^{\mathrm{MC}}$ are unfolded using a value of 172.5 GeV to obtain the migration matrix. Figure 6.28 shows distributions produced with MADGRAPH and values of 166.5 GeV and 178.5 GeV for $m_{\mathrm{top}}^{\mathrm{MC}}$, unfolded with the central MADGRAPH sample with $m_{\mathrm{top}}^{\mathrm{MC}} = 172.5\,\mathrm{GeV}$. The same distributions for the masses of 169.5 GeV, 171.5 GeV, 173.5 GeV and 175.5 GeV can be found in appendix B. The unfolding performs well for a mass difference up to ± 3 GeV. For mass differences of ± 6 GeV the agreement between the unfolded and the true distribution gets slightly worse, with at least one bin being off by about 2σ. Control distributions for the highest and lowest value of $m_{\mathrm{top}}^{\mathrm{MC}}$ compared to the central value are shown in Fig. 6.29, where the distributions of the first three jet p_Ts and the mass of the leading jet are shown for all three samples. Similar to the down variation of the renormalization and factorization scales the low mass distributions show a significant difference to the other two samples, leading to a bias in the unfolding.

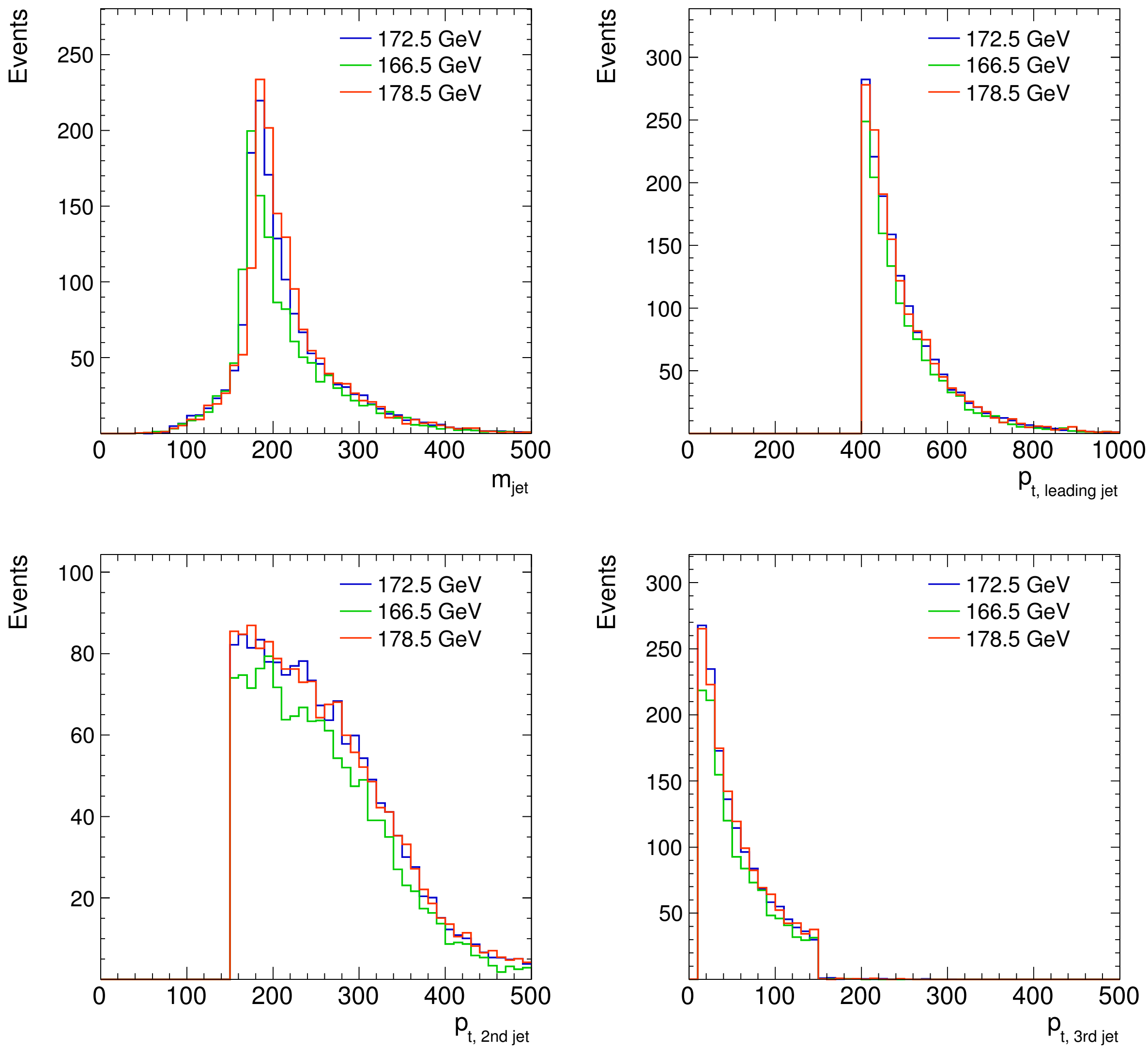

Figure 6.29.: Control distributions for MADGRAPH samples produced with different top quark masses.

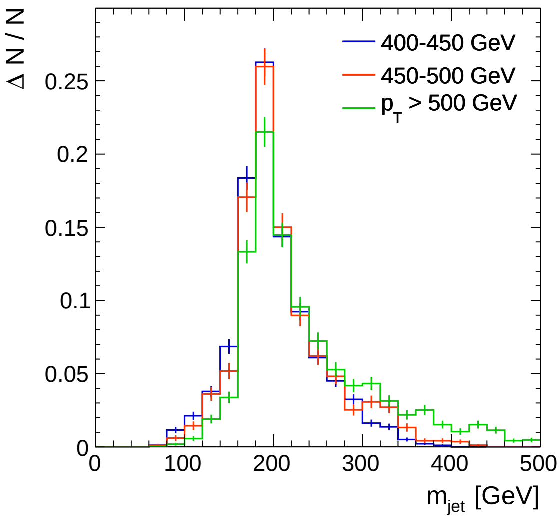

Figure 6.30.: Jet mass distribution of the leading jet on particle level after the full selection for different p_T regions.

6.4.6. Devision of the phase space

In order to study possible improvements on the unfolding, a devision of the phase space in two regions on both particle and reconstruction level is studied in the following. Previously it has been observed that a variation of the renormalization and factorization scales as well as a variation of $m_{\mathrm{top}}^{\mathrm{MC}}$ leads to differences in the p_T spectra of the jets and the jet mass distribution. Figure 6.30 shows the jet mass distribution for different p_T bins. For $p_T > 500\,\mathrm{GeV}$ the shape of the mass spectrum differs significantly compared to the shape in the lower p_T bins. In order to take this p_T dependence into account in the unfolding the phase space is divided in two p_T regions, one from 400 GeV to 500 GeV and one for 500 GeV and above. Both of these regions are unfolded at the same time to account for migrations and correlations between the two p_T regions. The migration matrix is divided in four parts, describing migrations within the two p_T regions and between them. Figure 6.31 shows a response matrix simulated with POWHEG + PYTHIA in the electron channel. In order to show both sup phase-spaces in the matrix at the same time, events with a leading jet $p_T > 500\,\mathrm{GeV}$ are shown with a jet mass increased by a constant value of 500 GeV. The response matrix shows only small migrations from the high to the low p_T regions.

After the unfolding the two p_T regions are added up to obtain on single distribution of m_{jet} with $p_T > 400\,\mathrm{GeV}$. The covariance matrix of the output is given by two covariance matrices for the two sub-phase spaces, and two covariance matrices, holding the covariances between the sub-phase spaces. The covariance matrix for the combination of both sub phase spaces is obtained by adding up the four covariance matrices.

For this set up all the tests of the previous section are repeated to check for improvement. All the tests can be found in the appendix for the divided phase space in C.1 and for the

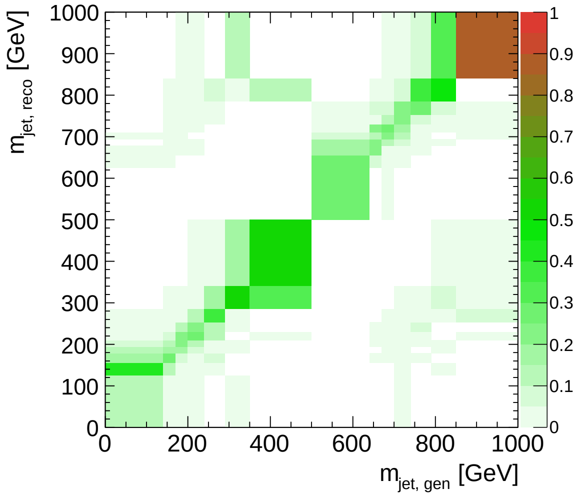

Figure 6.31.: Migration matrix in POWHEG for the unfolding using two regions in p_T. Events with a leading jet $p_T > 500\,\text{GeV}$ are shown with a jet mass of the leading jet increased by a constant value of 500 GeV.

combination in C.2.

The combined tests show no improvement to the undivided set up. The statistical uncertainties are slightly higher and the model dependence increases, leading to larger systematic uncertainties. Before the combination of the two p_T-regions however, one can see that the main contribution of the model dependence, due to the choice of the Monte Carlo generator and the scale variations, comes from the lower p_T-region. A possible explanation for this effect is, that as a result of different p_T-spectra in the various samples, migrations from lower p_T-regions, which are not included in the unfolding can lead to different corrections, and in this way to the large bias for the scale variations and different MC generators. The shape of the p_T-spectra of the MADGRAPH sample and the sample produced with renormalization and factorization scales varied down by a factor of 0.5, differ to the one of the POWHEG sample. This could lead to a difference in the jet mass spectrum. In principle it would be interesting to test this with an additional p_T-region with, for example $300\,\text{GeV} < p_T < 400\,\text{GeV}$, to see if it's possible to reduce the model dependence on the phase space with $p_T > 400\,\text{GeV}$. This is not within the scope of this work, but will be implemented at a later stage. However, a measurement for $p_T > 500\,\text{GeV}$ will be performed, which has reduced systematic uncertainties albeit larger statistical ones. The main result, however, will be obtained by an unfolding for jets with $p_T > 400\,\text{GeV}$ without a division in two regions.

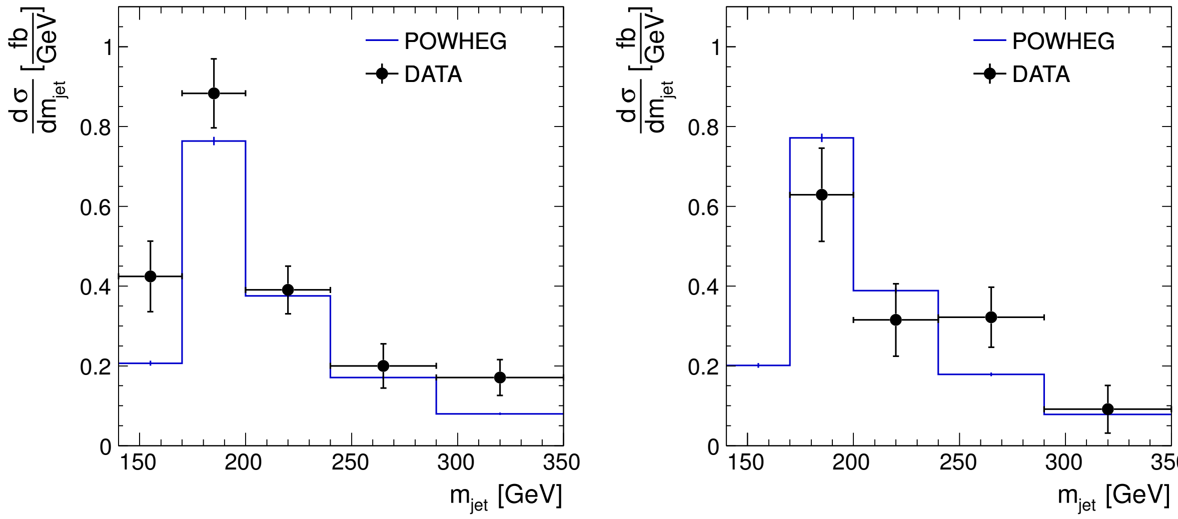

Figure 6.32.: Measured differential cross section as a function of the leading jet mass in data, unfolded with migrations obtained from POWHEG, and compared to the particle level distribution in POWHEG. The electron channel is shown on the left and the muon channel on the right. The error bars include statistical errors only.

6.4.7. Unfolding real data

The unfolding procedure is applied to real data. The electron and the muon channels are unfolded separately, and added up afterwards, because they are statistically independent. Purity, stability and efficiency have been shown above and do not differ between the two channels. In the first bin the purity is so low that it is excluded from the final result. Results for the unfolding of jets with $p_T > 400\,\text{GeV}$ are presented in Fig. 6.32. It shows the unfolded m_{jet} distributions from data in both channels compared to the truth distribution simulated with POWHEG. The results for the unfolding of jets with $p_T > 500\,\text{GeV}$, together with the low p_T sideband region can be found in appendix D.

Figure 6.33 shows the scans over the global correlation coefficient used for the unfolding of the data in both channels as a function of the τ-parameter. Both scans show a well defined minimum and that the chosen value of τ lies in both cases directly in the minimum. Correlations between the output bins are also be calculated. The absolute value of the correlation coefficients between different bins should not be greater than 50 % to ensure a stable unfolding. Figure 6.34 shows the correlation matrices for both channels. All off-diagonal elements are below 50 %.

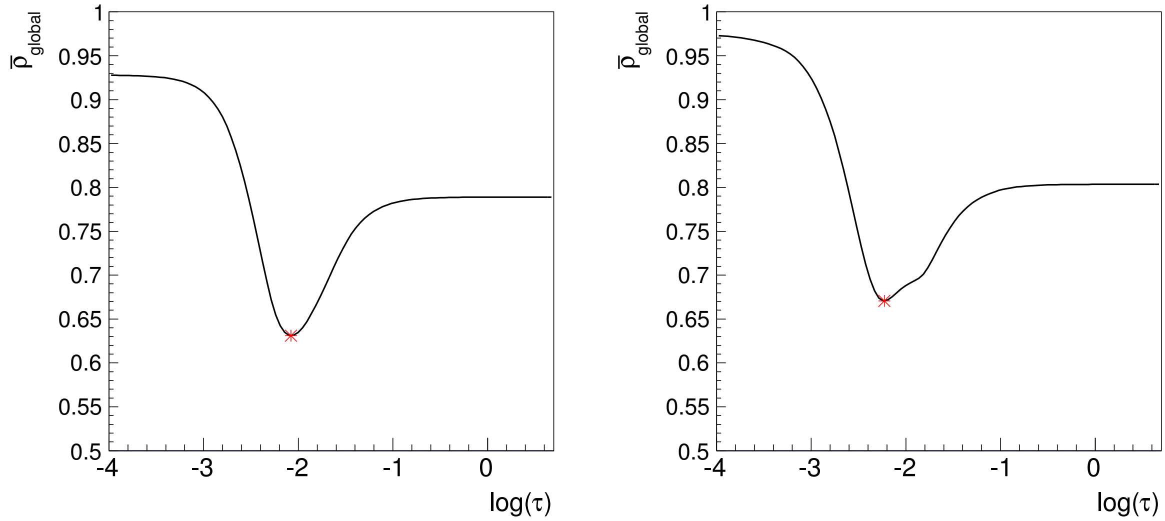

Figure 6.33.: Scans of the average global correlation coefficient for the unfolding of the data in the electron channel on the left and the muon channel on the right, as a function of the τ-parameter. Systematic uncertainties are included in the calculation of global correlation coefficients. The red marker shows the τ value that has been chosen.

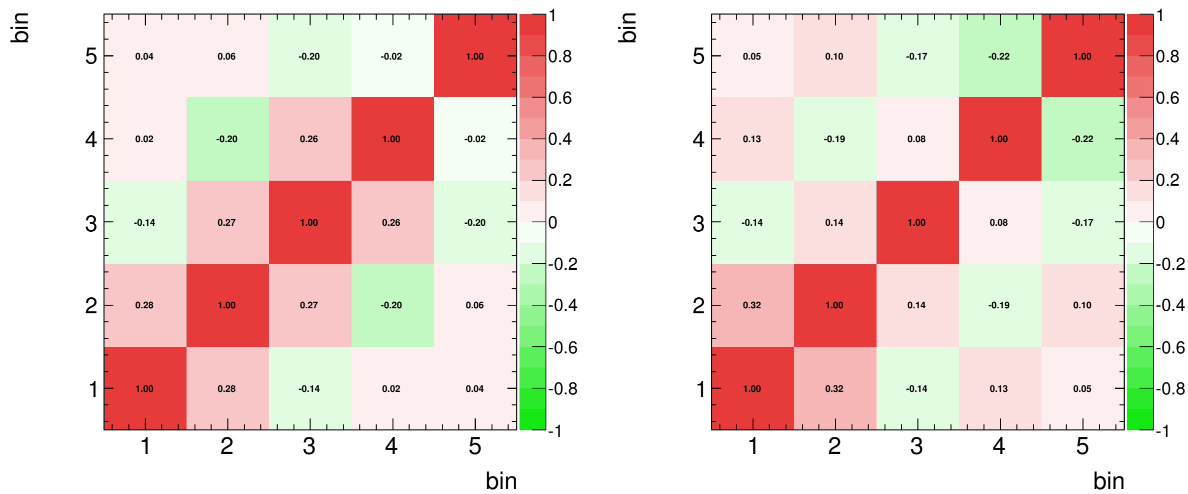

Figure 6.34.: Correlation coefficients of the output in the electron channel on the left and the muon channel on the right, including statistical uncertainties only.

6.5. Systematic Uncertainties

Apart from statistical uncertainties, a set of systematic uncertainties has to be taken into account for this measurement. A list of systematic uncertainties that are included in the final result is given in the following.

Uncertainties included in the unfolding process

The uncertainties included in the unfolding process are obtained by a variation of correction factors by ± 1 standard deviation. Correction factors are used to correct for differences between Monte Carlo simulation and data. The data is unfolded with migration matrices determined with up and down varied correction factors and the unfolded distributions are compared to the ones obtained with a non varied migration matrix. The mean difference of the up and down varied output to the unvaried output is added quadratically as systematic uncertainty to the total uncertainty. The considered uncertainties are:

- The jet energy resolution (JER) corrections are applied to MC samples as an η dependent factor. The size of the uncertainties depends on η as well.

- The jet energy corrections (JEC) as described in section 4.2.2. In this case the size of the uncertainties depends on p_T and η.

- Pile-up correction factors are applied to match the number of primary interactions to the instantaneous luminosity profile in data. The variation is done by varying the minimum bias cross section used for the matching.

- B-tagging scale factors due to efficiencies and mis-tag rates as evaluated in reference [53]. The uncertainties depend on p_T and η.

Uncertainties due to background processes

Events originating from background processes are estimated from simulation and subtracted from data before the unfolding. Uncertainties due to the normalization of the background samples are estimated by changing the cross section of the various processes.

- The uncertainty of the $W+$jets cross section is chosen conservatively as the uncertainty for W+heavy flavor production in reference [59]. An uncertainty of 23 % for all $W+$jets events is used.

- The uncertainty on single top cross section is also taken to be 23 % as measured in reference [60] for tW-production.

Normalization Uncertainties

Uncertainties affecting the overall normalization are added quadratically to the total uncertainty after the unfolding. The considered uncertainties are:

- An uncertainty to the luminosity determination of 2.6 %, as descried in reference [61].
- An uncertainty for the single electron trigger efficiency of 1 % [55].
- An uncertainty for the single muon trigger of 1 % [55].

Uncertainties due to model dependence

- Reconstructed events simulated with MADGRAPH are unfolded with a migration matrix determined in POWHEG, to get an estimation on the uncertainty of the Monte Carlo simulation used. The difference between the output and the particle level distribution in MADGRAPH is considered as systematic uncertainty.
- The uncertainty due to the choice of the renormalization and factorization scales are estimated by using, samples with varied scales. These are unfolded with the default POWHEG sample for $m_{t\bar{t}} > 700\,\text{GeV}$. Differences between output and particle level distributions are again considered as systematic uncertainties.
- The uncertainty due to the choice of $m_{\text{top}}^{\text{MC}}$ used to obtain the migration matrix is estimated using samples with different values of $m_{\text{top}}^{\text{MC}}$. It is estimated by unfolding samples produced with masses, that differ up to $\pm 6\,\text{GeV}$ from the sample used to obtain the central result.

Because the deviations between unfolded and true distribution for the variations of $m_{\text{top}}^{\text{MC}}$ and the scale variations have similar reasons, only the largest deviation in a given bin is applied as a systematic uncertainty to the final result. This is done in order not to overestimate the systematic uncertainty.

The uncertainties due to the variations are shown for both measurements in Fig. 6.35. Uncertainties due to model dependence can be found in Fig. 6.36. These uncertainties are added quadratically to the statistical uncertainties in each bin.

Figure 6.36 shows that a significant reduction of the uncertainty due to the model dependence is obtained for the measurement with $p_T > 500\,\text{GeV}$ in the peak region ($170 < m_{jet} < 200\,\text{GeV}$). In the other regions the reduction of the model dependence is not as large, but it can be seen from Fig. C.2 and C.3 in appendix C.1, that these could be statistical fluctuations. A study if these are real systematic differences is beyond the scope of this work and will be addressed in the future.

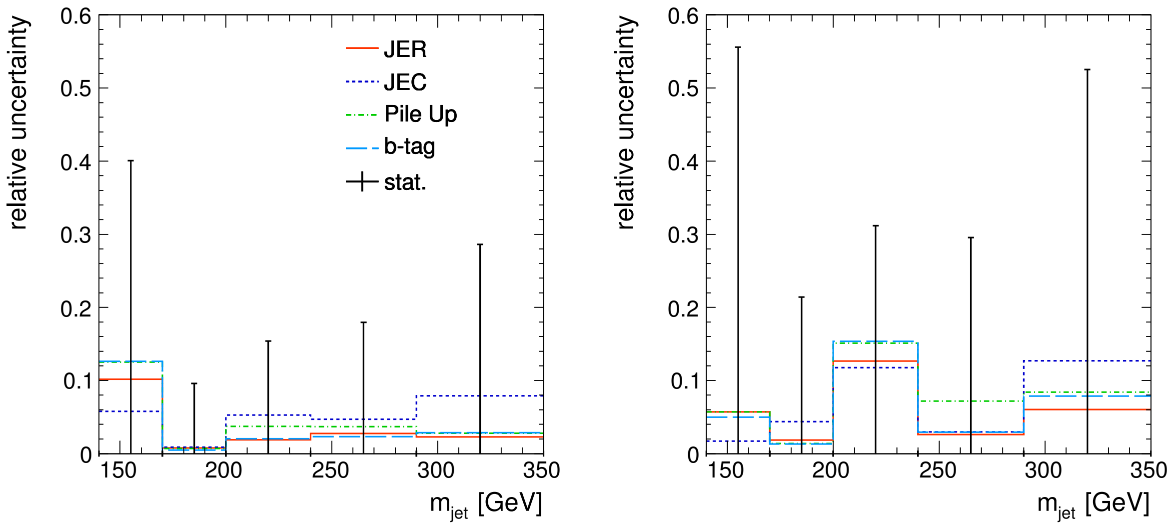

Figure 6.35.: Systematic uncertainties due to experimental effects for the unfolding of jets with $p_T > 400\,\mathrm{GeV}$ (left) and $p_T > 500\,\mathrm{GeV}$ (right).

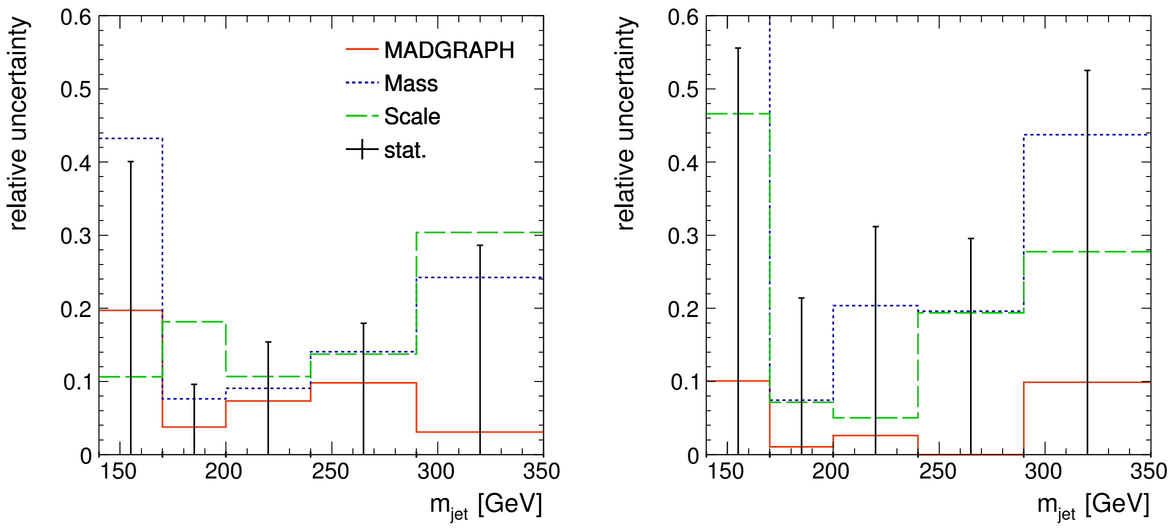

Figure 6.36.: Systematic uncertainties due to the model dependence for the unfolding of jets with $p_T > 400\,\mathrm{GeV}$ (left) and $p_T > 500\,\mathrm{GeV}$ (right).

7. Results

The main result of this thesis is an unfolded measurement on the level of stable particles of the differential cross section for top quark production in the lepton+jets channel as a function of the leading jet mass. The measurement is performed in a phase space defined by the following criteria discussed in section 6.2:

- 1 jet with $p_T > 400\,\mathrm{GeV}$ and $|\eta| < 2.5$
- 1 jet with $p_T > 150\,\mathrm{GeV}$ and $|\eta| < 2.5$
- veto on more jets with $p_T > 150\,\mathrm{GeV}$ and $|\eta| < 2.5$
- 1 electron or muon with $p_T > 45\,\mathrm{GeV}$ and $|\eta| < 2.5$
- $\Delta R_{\text{2nd jet, lepton}} < 1.2$
- $m_{\text{leading jet}} > m_{\text{2nd jet + lepton}}$

The result is obtained by combining the electron and muon channels, as given in Fig. 6.32. The event numbers are scaled with the luminosity and the bin width to obtain the differential cross section. All systematic uncertainties listed in section 6.5 are added quadratically to the statistical uncertainties. The measured differential cross section as a function of the leading jet mas can be found in Fig. 7.1, in comparison with particle level distributions obtained from POWHEG and MADGRAPH. In Fig. 7.2 the results are compared to MADGRAPH produced with different values of the top quark mass $m_{\text{top}}^{\text{MC}}$.

Figure 7.1 shows a better description of the data by POWHEG than by MADGRAPH. The systematic uncertainties are quite large, especially in the first two bins. These large uncertainties arise form the dependence on the scale variation, shown in Fig. 6.26, and from the mass dependence of the unfolding shown in Fig. 6.28. A detailed list of all systematic uncertainties for each bin is listed in Tab. 7.1. A list of cross sections for each bin together with the uncertainties applied can be found in Tab. 7.2.

The matrix of correlation coefficients including all statistical uncertainties for the data is shown in Fig. 7.3. All off-diagonal correlation coefficients lie well below 50 %. The systematic uncertainties are added fully uncorrelated to the covariance matrix after the unfolding.

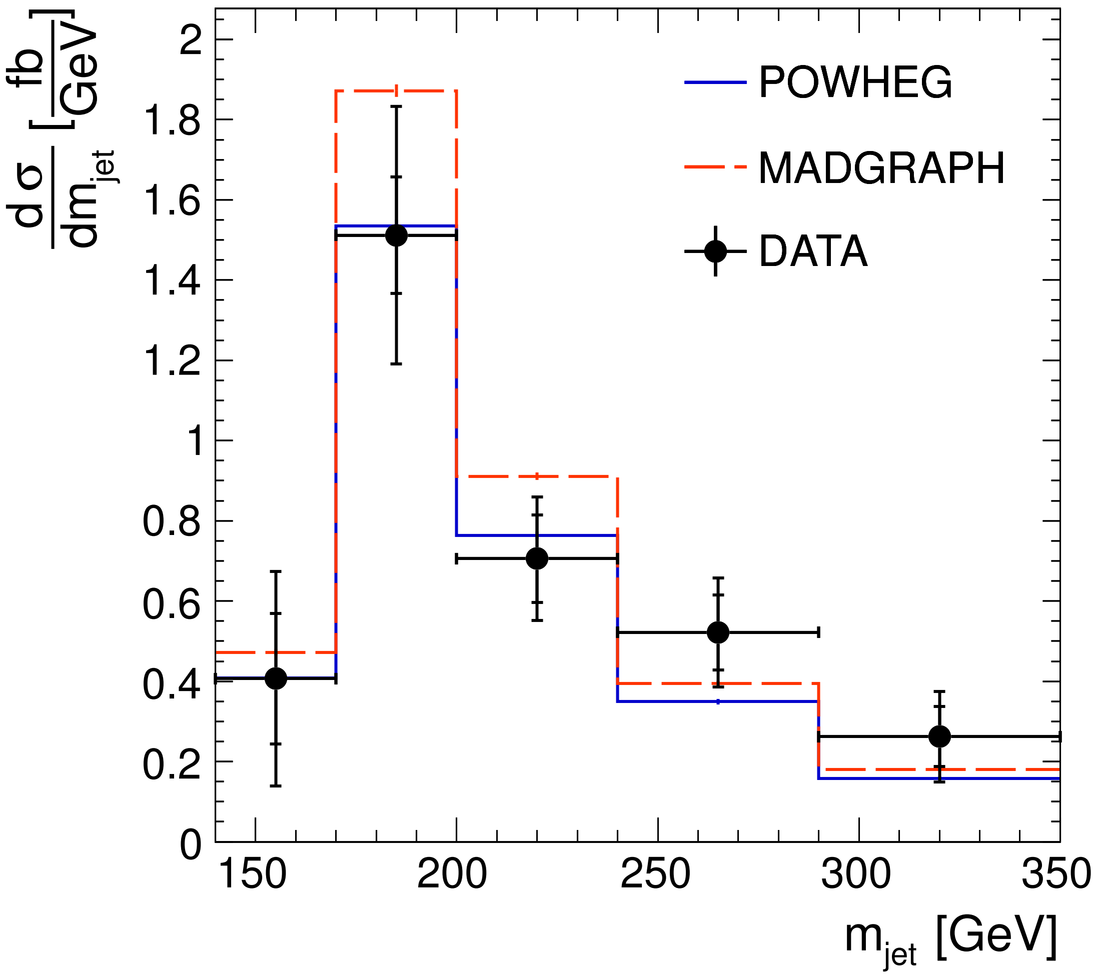

Figure 7.1.: Measured differential cross section for the electron and muon channels as a function of the leading jet mass. The measurement to the particle level distributions in MADRAPH and POWHEG. The unfolding was performed with migrations determined with POWHEG. The inner error bar shows to the statistical uncertainties. Systematic uncertainties are added quadratically to obtain the total uncertainty (outer error bar).

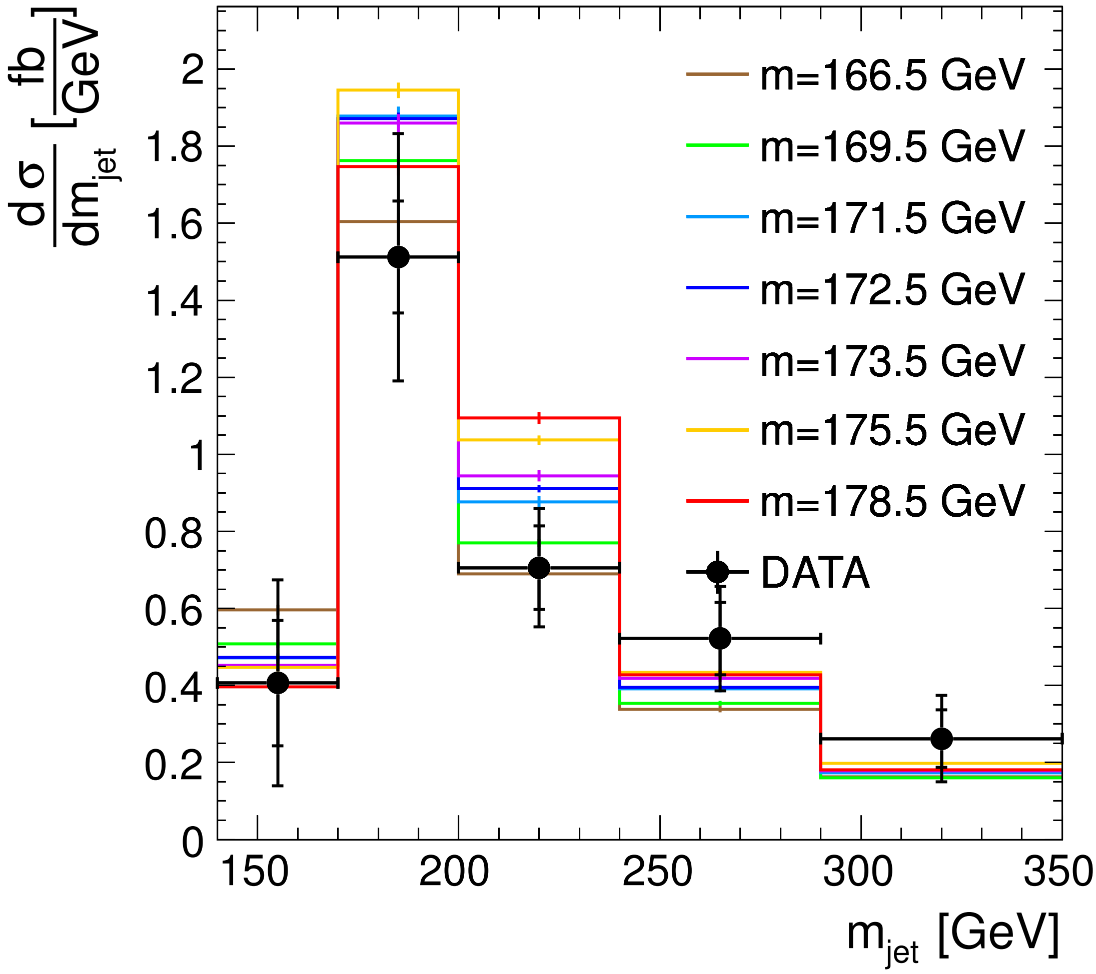

Figure 7.2.: Measured differential cross section for the electron and muon channels as a function of the leading jet mass. The measurement is compared to the particle level distributions in MADRAPH with different values of the top quark mass $m_{\text{top}}^{\text{MC}}$. The unfolding was performed with migrations determined with POWHEG. The inner error bar shows to the statistical uncertainties. Systematic uncertainties are added quadratically to obtain the total uncertainty (outer error bar).

bin	1	2	3	4	5
input	± 35.9 %	± 8.8 %	± 14.0 %	± 16.7 %	± 26.4 %
background subtraction	± 10.6 %	± 2.1 %	± 2.9 %	± 2.6 %	± 4.4 %
stat. on the migr. matrix	± 14.3 %	± 3.3 %	± 5.7 %	± 6.0 %	± 10.1 %
total statisical error	± 40.1 %	± 9.6 %	± 15.4 %	± 17.9 %	± 28.6 %
background cross setion	± 0.2 %	± 2.2 %	± 3.7 %	± 1.8 %	± 0.1 %
JER	+ 11.7 %	+ 0.0 %	+ 0.0 %	+ 0.0 %	+ 0.0 %
	- 0.0 %	- 1.1 %	- 2.7 %	- 2.8 %	- 2.6 %
JEC	+ 10.5 %	+ 0.3 %	+ 3.7 %	+ 1.0 %	+ 5.4 %
	- 1.1 %	- 0.0 %	- 6.9 %	- 8.4 %	- 10.4 %
pile up	+ 16.5 %	+ 0.2 %	+ 1.7 %	+ 1.3 %	+ 0.0 %
	- 0.0 %	- 1.2 %	- 5.8 %	- 6.1 %	- 3.5 %
b-tagging scale factors	± 12.6 %	± 0.5 %	± 2.0 %	± 2.3 %	± 2.9 %
MADGRAPH-POWHEG	± 14.1 %	± 3.7 %	± 7.3 %	± 9.8 %	± 3.1 %
scale and mass	± 43.2 %	± 18.2 %	± 10.7 %	± 14.1 %	± 30.4 %
Luminosity	± 2.6 %	± 2.6 %	± 2.6 %	± 2.6 %	± 2.6 %
Trigger	± 1.0 %	± 1.0 %	± 1.0 %	± 1.0 %	± 1.0 %

Table 7.1.: Detailed list of systematic uncertainties and their contribution to the total uncertainty for each measurement bin ($p_T > 400\,\mathrm{GeV}$).

bin	1	2	3	4	5
Range [GeV]	140-170	170-200	200-240	240-290	290-350
Cross section [fb]	12.2	45.4	28.2	26.1	15.7
statistical uncertainty	40.1 %	9.6 %	15.4 %	17.9 %	28.6 %
systematic uncertainty	21.5 %	3.8 %	8.4 %	7.7 %	9.6 %
model uncertainty	47.5 %	18.5 %	12.9 %	17.1 %	30.5 %
total uncertainty	65.8 %	21.2 %	21.8 %	26.0 %	42.9 %

Table 7.2.: List of measured cross sections and systematic uncertainties for each bin ($p_T > 400\,\mathrm{GeV}$).

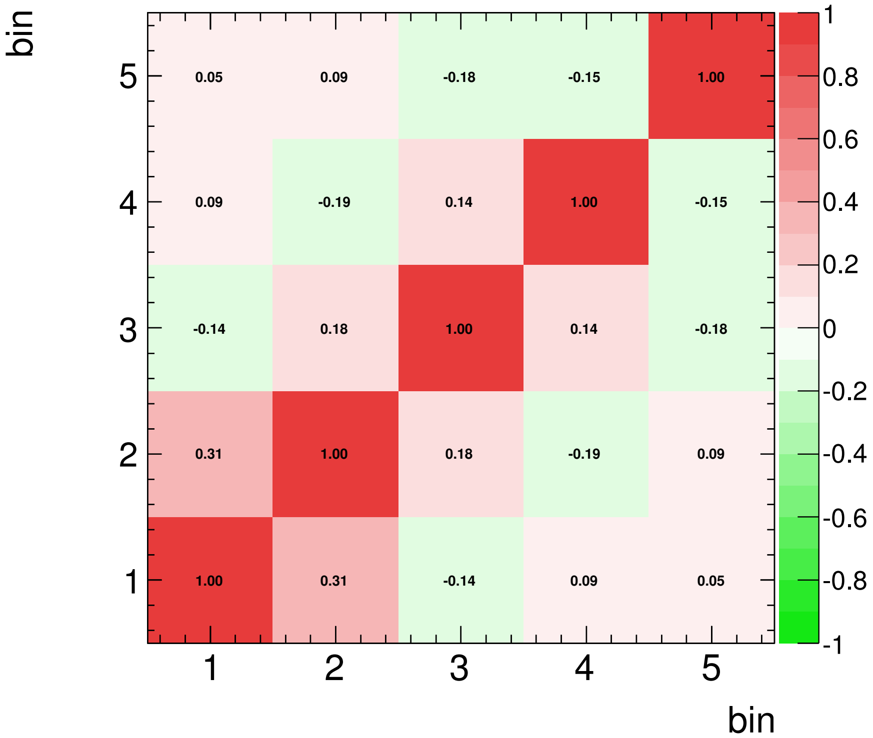

Figure 7.3.: Correlation coefficients of the combined data in the electron and muon channels, including statisical uncertainties only.

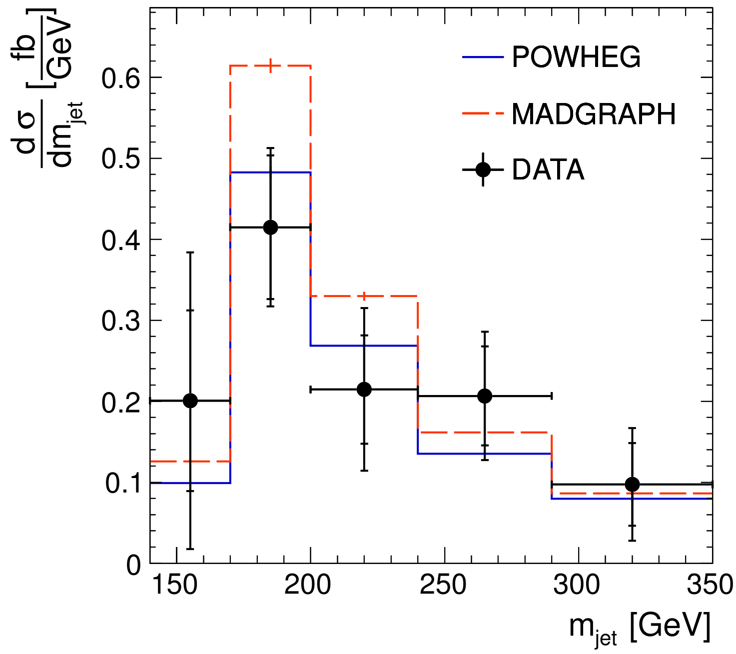

Figure 7.4.: Measured differential cross section for the electron and muon channels as a function of the leading jet mass. The measurement is compared to the particle level distributions in MADRAPH and POWHEG. The unfolding was performed with migrations determined with POWHEG. The inner error bar shows to the statistical uncertainties. Systematic uncertainties are added quadratically to obtain the total uncertainty (outer error bar).

Measurement for jets with $\mathbf{p_T} > 500$ GeV

A second measurement is performed for $p_T > 500$ GeV of the leading jet, unfolded together with events containing jets with lower p_T between 400 and 500 GeV. The aim of this second measurement is to study the influence of a lower p_T sideband region on the systematic uncertainties. Figure 7.4 shows the result including all the systematic uncertainties discussed in section 6.5. For the bin in the peak of the distribution, the systematic uncertainty decreases by a large amount compared to the main result in Fig. 7.1. For the rest of the bins the relative systematic uncertainties increased, which could come from statistical uncertainties of the Monte Carlo samples contributing to the bias tests used to obtain these systematic uncertainties.

All systematic uncertainties are listed in Tab. 7.3 and the cross sections can be found in Tab. 7.4.

bin	1	2	3	4	5
input	± 48.6 %	± 19.2 %	± 27.8 %	± 27.7 %	± 48.2 %
background subtraction	± 13.8 %	± 4.3 %	± 5.4 %	± 4.3 %	± 8.5 %
stat. on the migr. matrix	± 23.1 %	± 8.4 %	± 13.0 %	± 9.5 %	± 19.1 %
total statisical error	± 55.6 %	± 21.4 %	± 31.2 %	± 29.6 %	± 52.5 %
background cross setion	± 6.1 %	± 2.5 %	± 5.5 %	± 1.6 %	± 1.0 %
JER	+ 7.5 %	+ 2.9 %	+ 0.0 %	+ 3.3 %	+ 7.1 %
	- 0.0 %	- 0.0 %	- 13.4 %	- 0.0 %	- 0.0 %
JEC	+ 2.3 %	+ 6.1 %	+ 1.8 %	+ 1.6 %	+ 17.0 %
	- 0.0 %	- 2.7 %	- 21.8 %	- 4.3 %	- 8.3 %
pile up	+ 10.8 %	+ 1.9 %	+ 0.0 %	+ 10.0 %	+ 9.5 %
	- 0.6 %	- 0.0 %	- 17.9 %	- 4.3 %	- 0.0 %
b-tagging scale factors	± 5.0 %	± 1.3 %	± 15.4 %	± 2.9 %	± 7.9 %
MADGRAPH-POWHEG	± 10.1 %	± 1.1 %	± 2.6 %	± 11.4 %	± 9.9 %
scale and mass	± 70.8 %	± 7.4 %	± 20.4 %	± 19.6 %	± 43.8 %
Luminosity	± 2.6 %	± 2.6 %	± 2.6 %	± 2.6 %	± 2.6 %
Trigger	± 1.0 %	± 1.0 %	± 1.0 %	± 1.0 %	± 1.0 %

Table 7.3.: Detailed list of systematic uncertainties and their contribution to the total uncertainty for each measurement bin ($p_T > 500\,\text{GeV}$).

bin	1	2	3	4	5
Range [GeV]	140-170	170-200	200-240	240-290	290-350
Cross section [fb]	6.0	12.4	8.6	10.3	5.8
statistical uncertainty	55.6 %	21.4 %	31.2 %	29.6 %	52.5 %
systematic uncertainty	11.7 %	6.3 %	28.3 %	9.3 %	18.4 %
model uncertainty	71.5 %	7.5 %	20.5 %	22.7 %	44.9 %
total uncertainty	91.3 %	23.6 %	46.8 %	38.4 %	71.5 %

Table 7.4.: List of measured cross sections and systematic uncertainties for each bin ($p_T > 500\,\text{GeV}$).

Determination of the Monte Carlo Mass

One of the main motivations of this measurement is a complimentary determination of the top quark mass, independent of the definition of the top quark mass used in the MC generator. However, theoretical calculations that are needed to compare to the unfolded jet mass spectrum for a determination of the top quark mass, do not exist yet.

Even if calculations would exist, it would still be interesting to determine the Monte Carlo top quark mass $m_{\mathrm{top}}^{\mathrm{MC}}$ and compare the result to existing measurements. In addition, this determination of $m_{\mathrm{top}}^{\mathrm{MC}}$ provides a good estimate for the experimental precision which can be obtained with this measurement.

The determination $m_{\mathrm{top}}^{\mathrm{MC}}$ is done by a one dimensional template fit using simulated distributions generated with different values of $m_{\mathrm{top}}^{\mathrm{MC}}$. These templates are obtained using MADGRAPH for masses of $m_{\mathrm{top}}^{\mathrm{MC}} = 166.5\,\mathrm{GeV}$, $m_{\mathrm{top}}^{\mathrm{MC}} = 169.5\,\mathrm{GeV}$, $m_{\mathrm{top}}^{\mathrm{MC}} = 171.5\,\mathrm{GeV}$, $m_{\mathrm{top}}^{\mathrm{MC}} = 172.5\,\mathrm{GeV}$, $m_{\mathrm{top}}^{\mathrm{MC}} = 173.5\,\mathrm{GeV}$, $m_{\mathrm{top}}^{\mathrm{MC}} = 175.5\,\mathrm{GeV}$ and $m_{\mathrm{top}}^{\mathrm{MC}} = 178.5\,\mathrm{GeV}$.

For every template, the value χ^2 is calculated according to

$$\chi^2 = (\vec{m}_{\mathrm{data}} - \vec{m}_{\mathrm{MC}})^{\mathrm{T}} \mathbf{C}^{-1} (\vec{m}_{\mathrm{data}} - \vec{m}_{\mathrm{MC}})$$

where $\vec{m}$ is a vector of the measured (data) or simulated (MC) normalized cross sections and $\mathbf{C}$ is the covariance matrix of the unfolded data. For a determination of $m_{\mathrm{top}}^{\mathrm{MC}}$ the χ^2 distribution should have a well defined minimum at the best value of $m_{\mathrm{top}}^{\mathrm{MC}}$. Figure 7.5 shows the χ^2 values for all mass points as a function of $m_{\mathrm{top}}^{\mathrm{MC}}$. It shows low χ^2 values for the samples with low values of $m_{\mathrm{top}}^{\mathrm{MC}}$ increasing with higher masses. The smallest value of χ^2 is observed for a mass of $m_{\mathrm{top}}^{\mathrm{MC}} = 169.5\,\mathrm{GeV}$. However, it is not clear if this distribution really shows a minimum, because the only lower mass point with $m_{\mathrm{top}}^{\mathrm{MC}} = 166.5\,\mathrm{GeV}$ has just a slightly higher χ^2 value. More mass point would be needed to see if there is a well defined minimum.

A possible improvement of this method could be obtained by the introduction of correlations from systematic uncertainties to the covariance matrix. In this thesis systematic uncertainties are just added fully uncorrelated to the covariance matrix. Additional correlations could improve the sensitivity of this method.

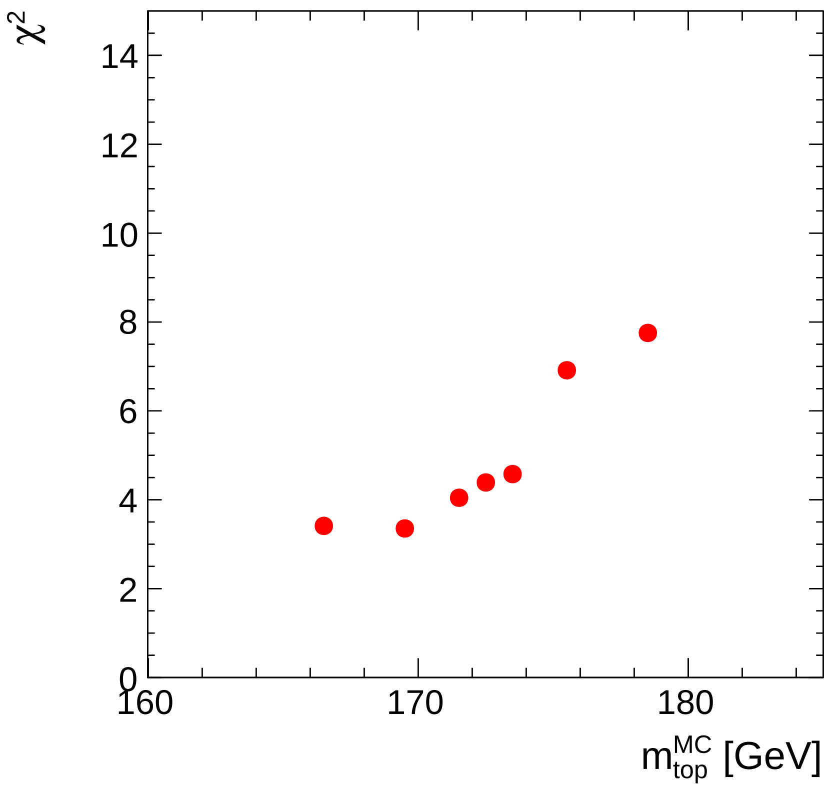

Figure 7.5.: χ^2 values calculated for different values of $m_{\text{top}}^{\text{MC}}$.

8. Summary

In this thesis a measurement has been performed of the invariant mass distribution of large jets containing a complete hadronic top quark decay. The measurement has been done in the electron and muon+jets decay channel of top anti-top quark decays.

As a first step a selection of jets containing a fully merged hadronic top quark decay has been developed on particle level. The minimum transverse momentum of the jets has been chosen to be greater than 400 GeV. To obtain jets including all decay products of a fully hadronic top quark decay, the jets are defined using the Cambridge/Aachen algorithm with a distance parameter of $R = 1.2$. For this selection no requirements on the substructure of the jets have been applied, in order not to bias the measurement. It has been shown that with the selection developed, the spectrum is dominated by jets from fully-merged top quark decays.

For the reconstruction level a selection of top anti-top quark decays in the the electron+jets and muon+jets decay channel was applied. Similar cuts as on particle level have been used to obtain a similar phase space definition on reconstruction level as on particle level. With these selections a regularized unfolding was set up and tested. For the regularization a size regularization acting on the content of each bin, has been chosen, because it showed the best performance. To evaluate the regularization strength, a scan over the average global correlation coefficient was used. Migrations have been evaluated using events simulated with POWHEG interfaced to PYTHIA.

A division of the phase space into two regions, of high and low transverse momentum of the leading jet, has been studied. The high p_T-region of this divided phase space shows less model dependence when taking both regions simultaneously into account in the unfolding.

The measurement is done using a dataset with an integrated luminosity of 19.7 fb^{-1} of proton-proton collisions with a center of mass energy of $\sqrt{s} = 8$ TeV, collected by the CMS detector at the LHC in the years 2010 to 2012. Two distributions are measured, one jet mass distribution for jets with $p_T > 400$ GeV and one for jets with $p_T > 500$ GeV, in which jets with 400 GeV $< p_T <$ 500 GeV are taken into account in the migration matrix of the unfolding. The muon+jets and the electron+jets channels have been unfolded separately. The output has been combined into one single measurement of the differential cross section as a function of the jet mass. Several systematic uncertainties have been included in the final results. The measurement of the jet mass for $p_T > 400$ GeV shows large uncertainties due to model dependence. These uncertainties are reduced in the measurement for $p_T > 500\,\mathrm{GeV}$, but the statistical uncertainties are larger. The measurement is compared to POWHEG and MADGRAPH. It is compatible with both distributions, but better described by POWHEG than by MADGRAPH.

As a last step a fit to the Monte Carlo top quark mass has been studied, that does not provide a reliable result yet.

9. Outlook

This thesis presents a first step towards measuring the jet mass from collimated top quark decays at the LHC. To reach the best possible precision with the 8 TeV dataset collected by the CMS experiment, several improvements are possible.

The unfolding with a lower p_T sideband region showed a possible reduction of systematic uncertainties due to scale variations and the choice of the Monte Carlo generator for p_T greater 500 GeV. This principle can be used for jets with $p_T > 400$ GeV using a sideband region with 300 GeV $< p_T <$ 400 GeV in the future. The systematic uncertainties included due to model dependence are overestimated due to statistical fluctuations. To reduce the influence of statistical fluctuations on the model dependent systematic uncertainties, a Kolmogorov–Smirnov test could be used, to decide, if a bias between the true and the unfolded distribution is a statistical fluctuation or a systematic effect, that has to be included in the result.

The uncertainty due to the hadronization model used has not been taken into account, since it is expected to be much smaller than the model dependence in the present measurement. Once the model dependence is reduced, this uncertainty should be studied as well. This can be done by using a $t\bar{t}$-sample simulated with POWHEG+HERWIG and comparing it to the default POWHEG+PYTHIA sample.

The influence of pile up could be reduced by using new pile up removing techniques like Pile-Up-Per-Particle-Identification (PUPPI) [62].

A measurement of the jet mass in the future with center of mass energies of 13-14 TeV at the LHC can be performed with higher statistics. The top anti-top production cross section increases with higher center of mass energy and the integrated luminosity in the next LHC runs is expected to be much higher than the one of the 8 TeV run. This leads to lower statistical uncertainties, the possibility to study systematic effects in more detail, and a measurement for even higher values of jet p_T, which is preferred by the theory.

A. Control Distributions in the muon channel

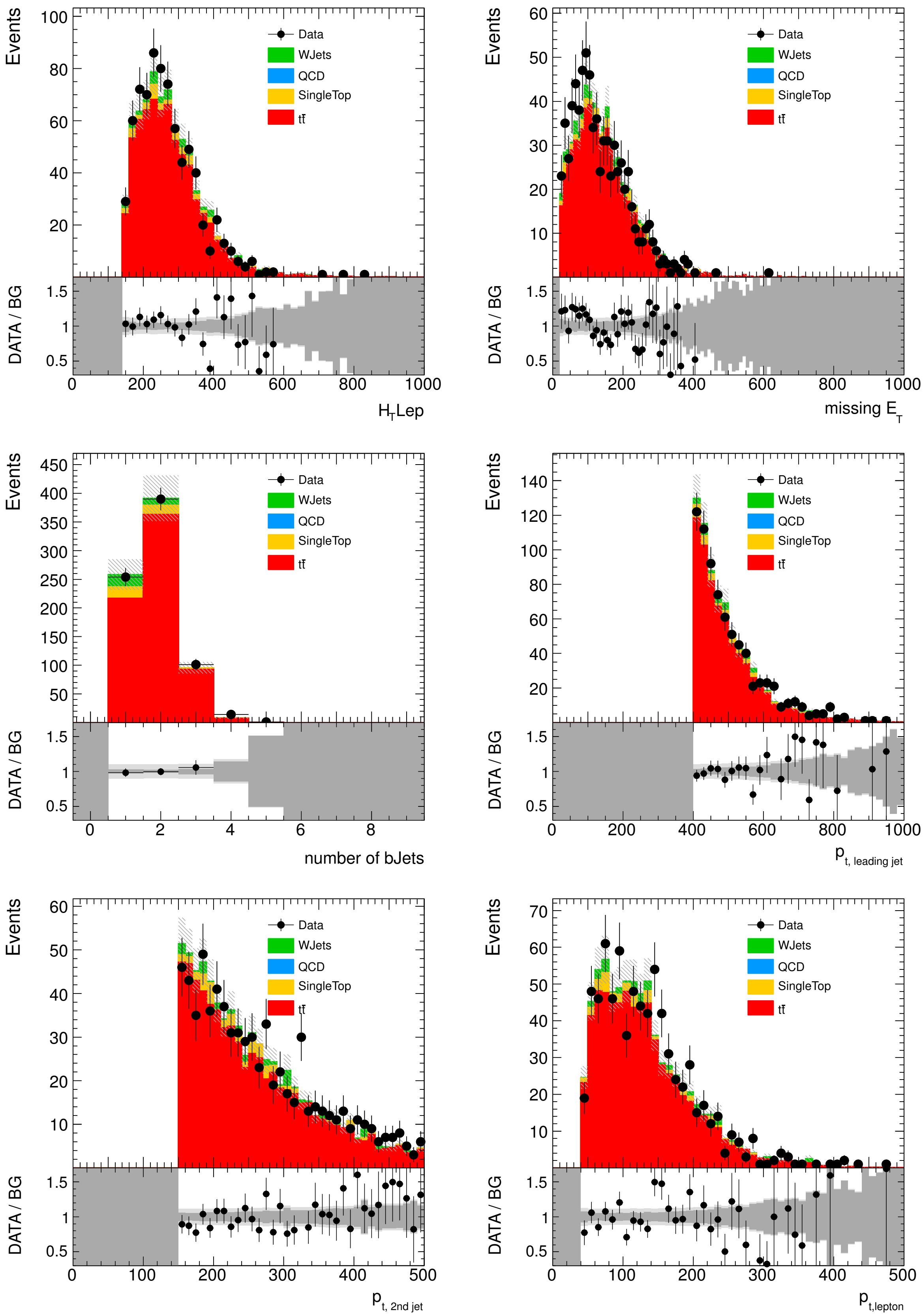

Figure A.1.: Control distributions after the full selection in the muon plus jets channel.

B. Mass dependance of the unfolding

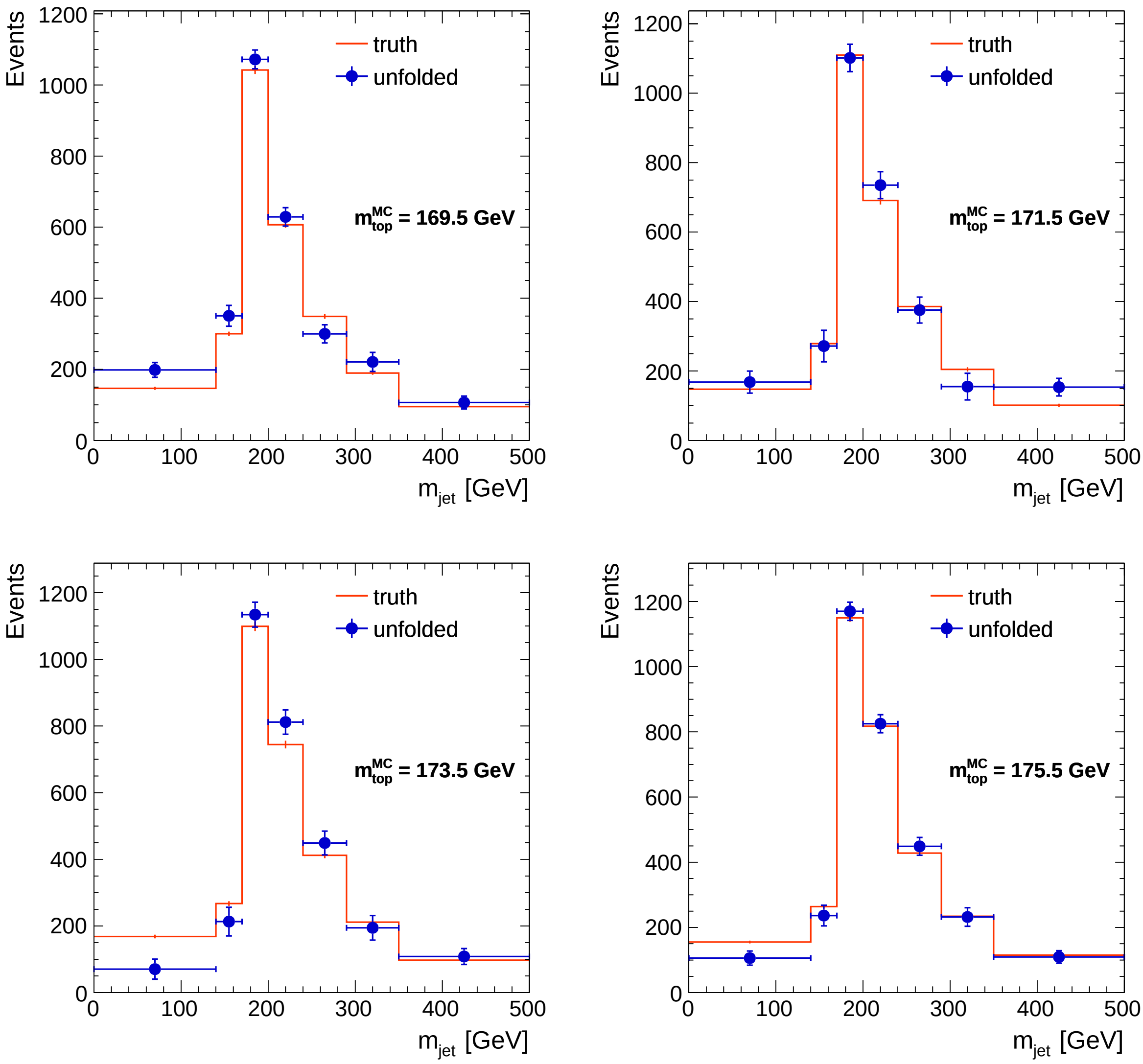

Figure B.1.: Jet mass distributions in events simulated with MADGRAPH with different values of $m_{\text{top}}^{\text{MC}}$ unfolded with migrations determined with the central MADGRAPH sample with $m_{\text{top}}^{\text{MC}} = 172.5\,\text{GeV}$. The unfolded distributions are shown together with the particle level distributions simulated with the same values of $m_{\text{top}}^{\text{MC}}$.

C. Unfolding tests for the divided phase space

C.1. Divided

This section shows the unfolding tests for the phase space divided into two p_T regions with $400\,\mathrm{GeV} < p_T < 500\,\mathrm{GeV}$ and $p_T > 500\,\mathrm{GeV}$. For events with a leading jet $p_T > 500\,\mathrm{GeV}$ the jet mass is added by a constant value of 500 Gev to unfold both p_T regions at the same time.

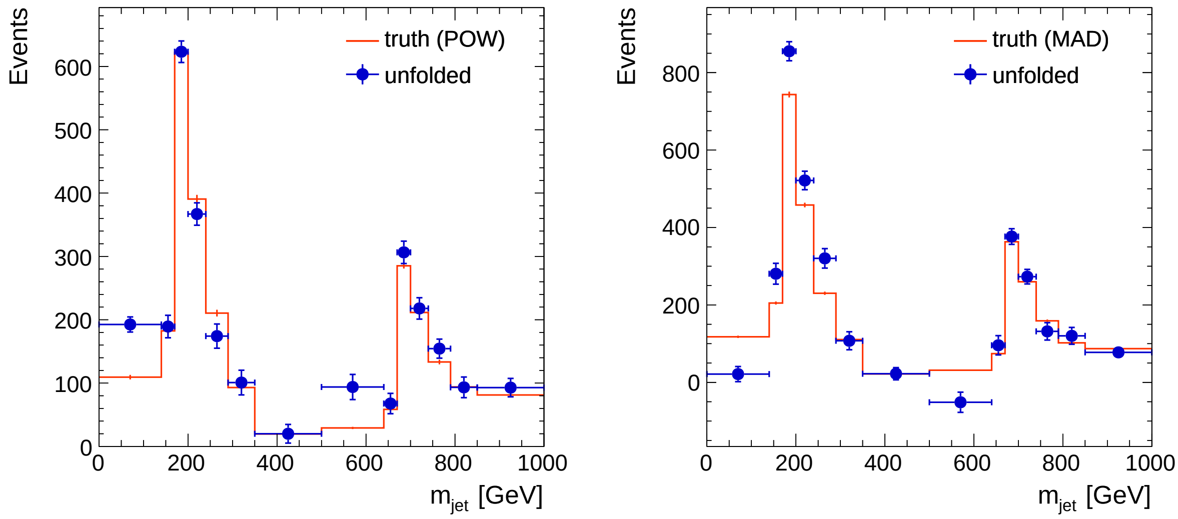

Figure C.1.: Unfolded jet mass distributions in events simulated with MADGRAPH with migrations of POWHEG in comparison with the particle level in MADGRAPH on the right and the other way around on the left.

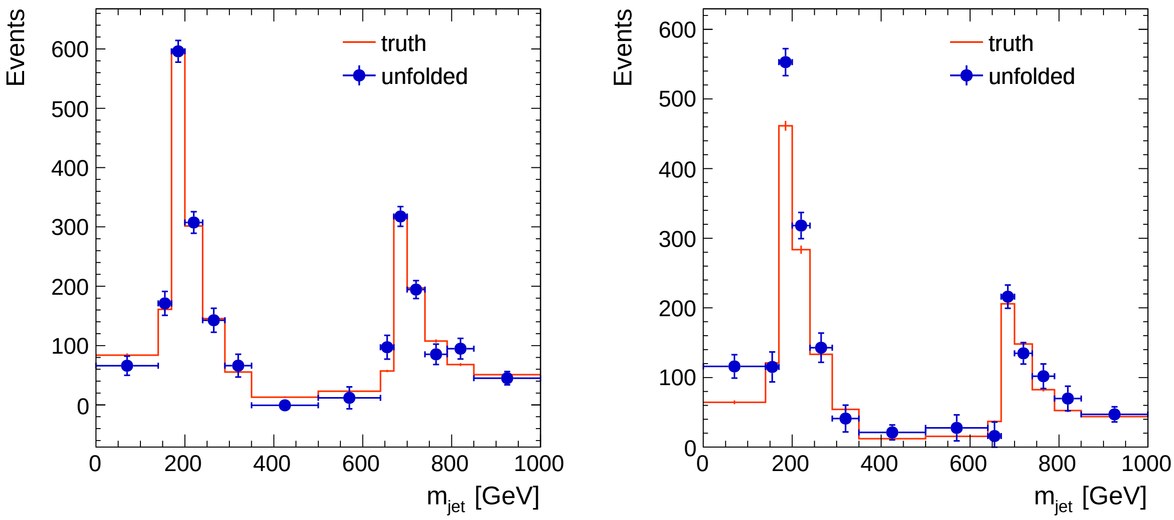

Figure C.2.: Jet mass distributions in events simulated with POWHEG with factorization and renormalization scales varied by factors of 2 (left) and 0.5 (right) unfolded with migrations determined with the default POWHEG sample for $m_{t\bar{t}} >$ 700 GeV. The unfolded distributions are shown together with the particle level distributions simulated with the same generator.

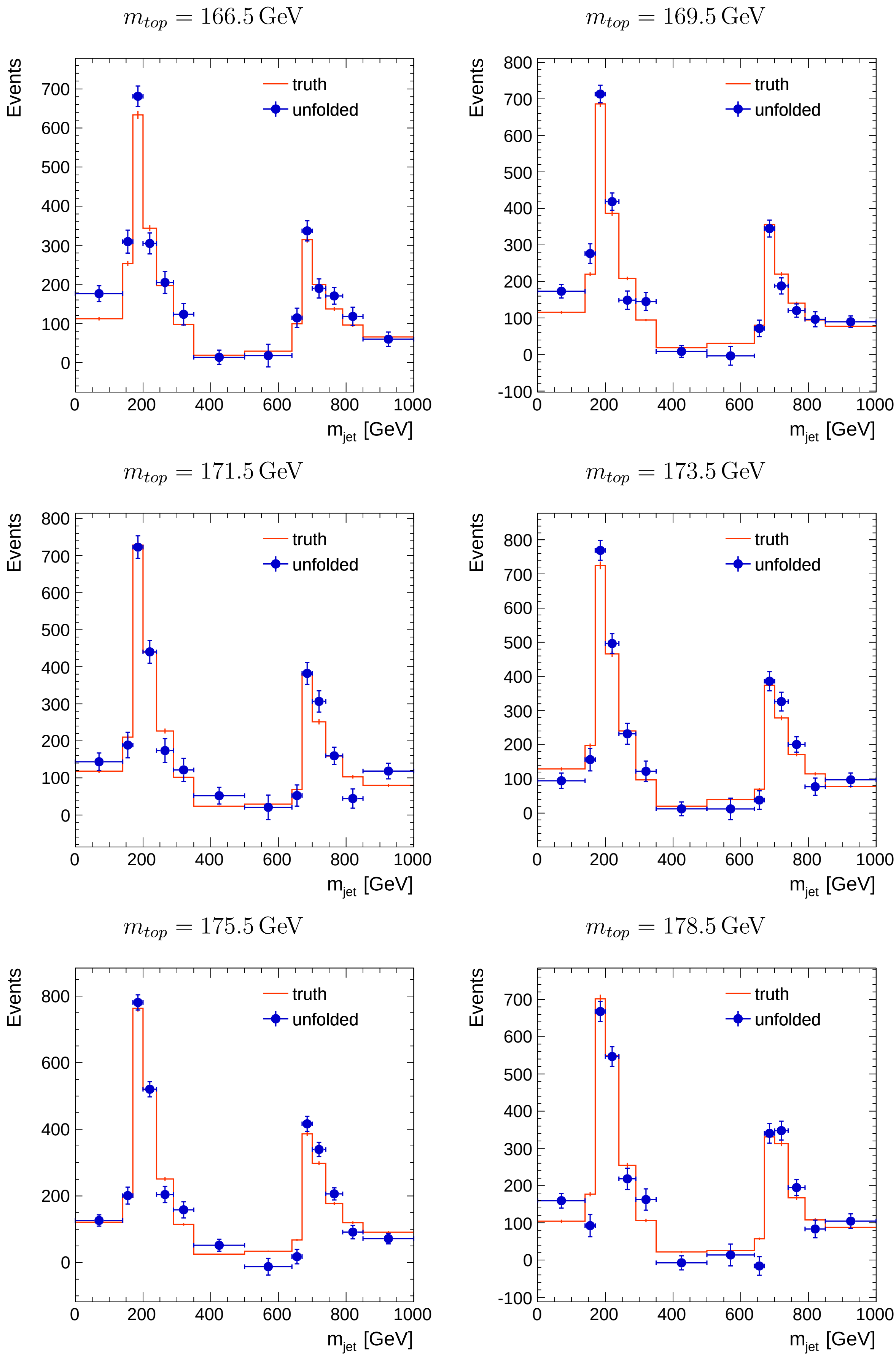

Figure C.3.: Jet mass distributions in events simulated with MADGRAPH with different values of $m_{\mathrm{top}}^{\mathrm{MC}}$ unfolded with migrations determined with the central MADGRAPH sample with $m_{\mathrm{top}}^{\mathrm{MC}} = 172.5\,\mathrm{GeV}$. The unfolded distributions are shown together with the particle level distributions simulated with the same values of $m_{\mathrm{top}}^{\mathrm{MC}}$.

C.2. Combined

This section shows the same test as section C.1 with both p_T regions combined after the unfolding to one phase space with a leading jet $p_T > 400\,\mathrm{GeV}$.

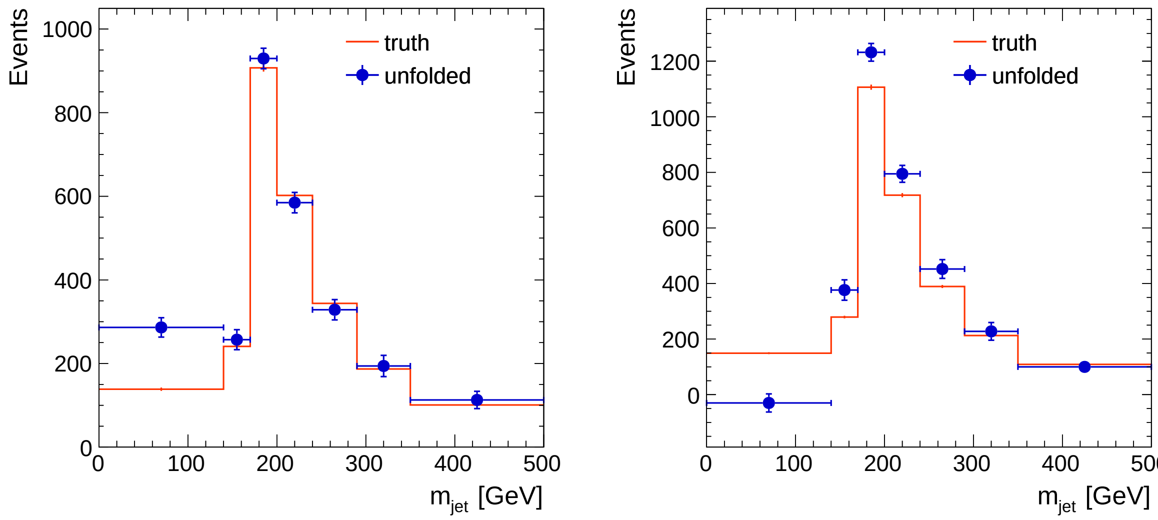

Figure C.4.: Unfolded jet mass distributions in events simulated with MADGRAPH with migrations of POWHEG in comparison with the particle level in MADGRAPH on the right and the other way around on the left.

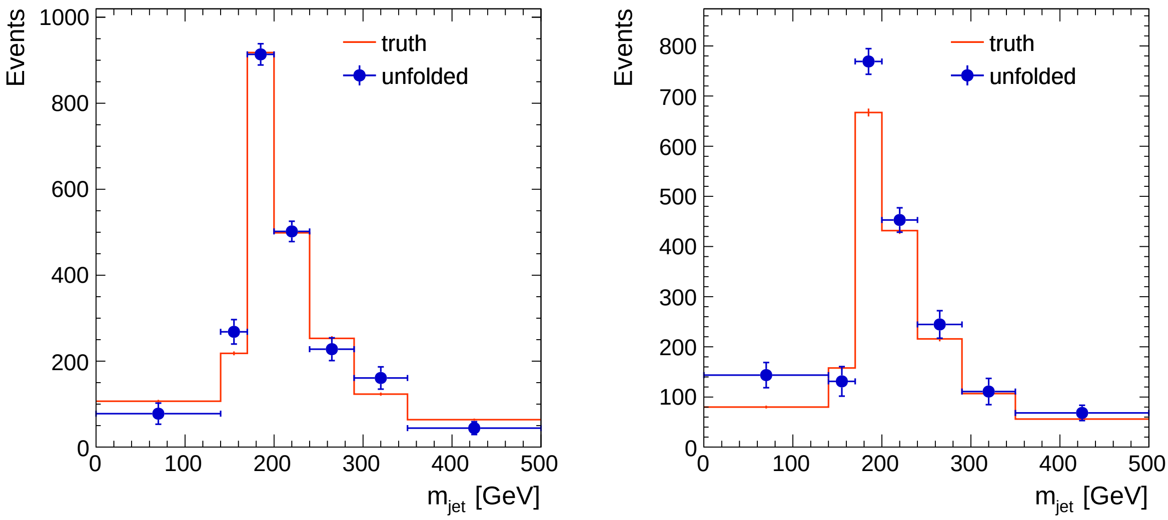

Figure C.5.: Jet mass distributions in events simulated with POWHEG with factorization and renormalization scales varied by factors of 2 (left) and 0.5 (right) unfolded with migrations determined with the default POWHEG sample for $m_{t\bar{t}} > 700\,\mathrm{GeV}$. The unfolded distributions are shown together with the particle level distributions simulated with the same generator.

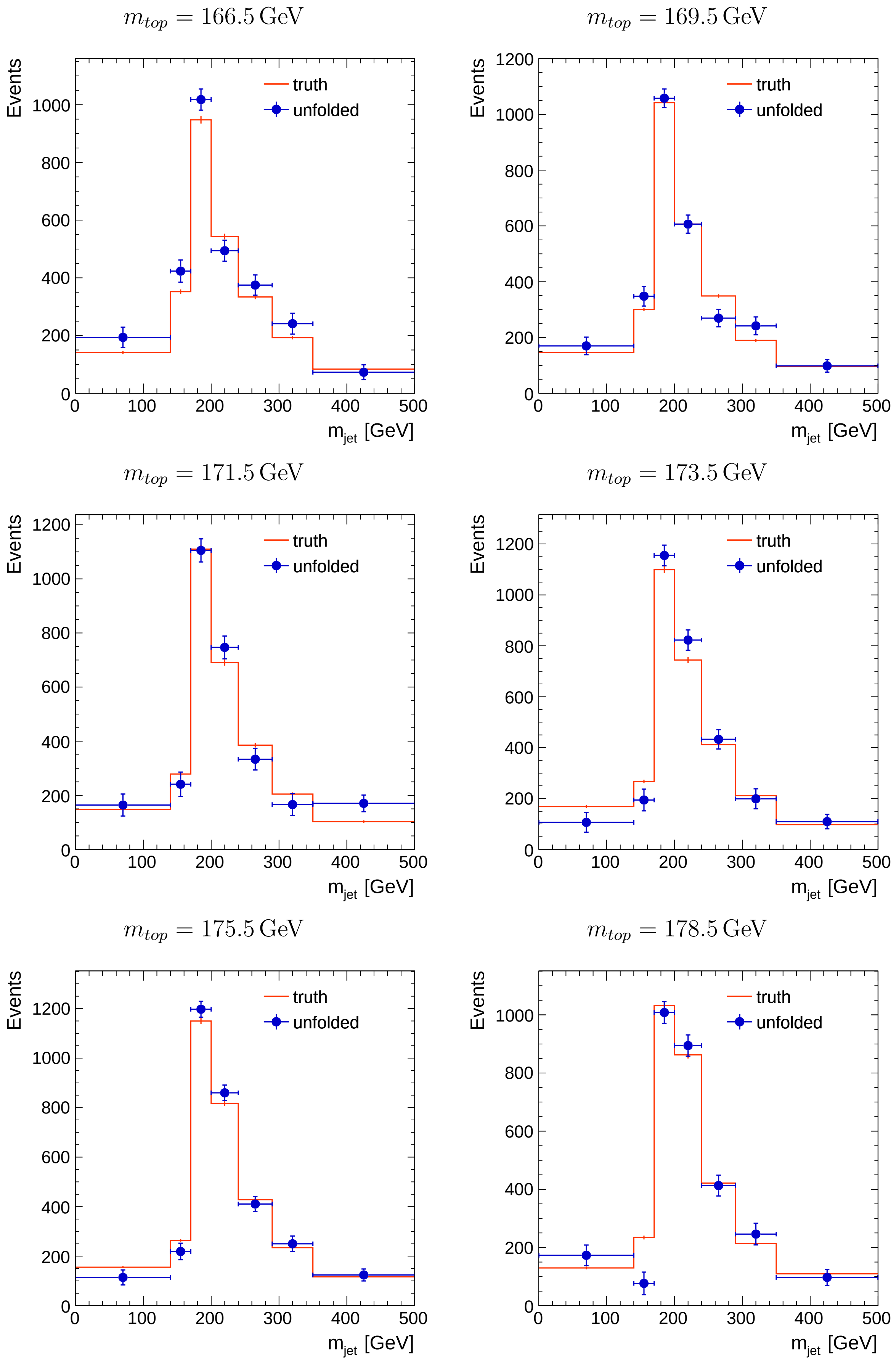

Figure C.6.: Jet mass distributions in events simulated with MADGRAPH with different values of $m_{\text{top}}^{\text{MC}}$ unfolded with migrations determined with the central MADGRAPH sample with $m_{\text{top}}^{\text{MC}} = 172.5\,\text{GeV}$. The unfolded distributions are shown together with the particle level distributions simulated with the same values of $m_{\text{top}}^{\text{MC}}$.

D. Unfolding Data for the Divided Phase Space

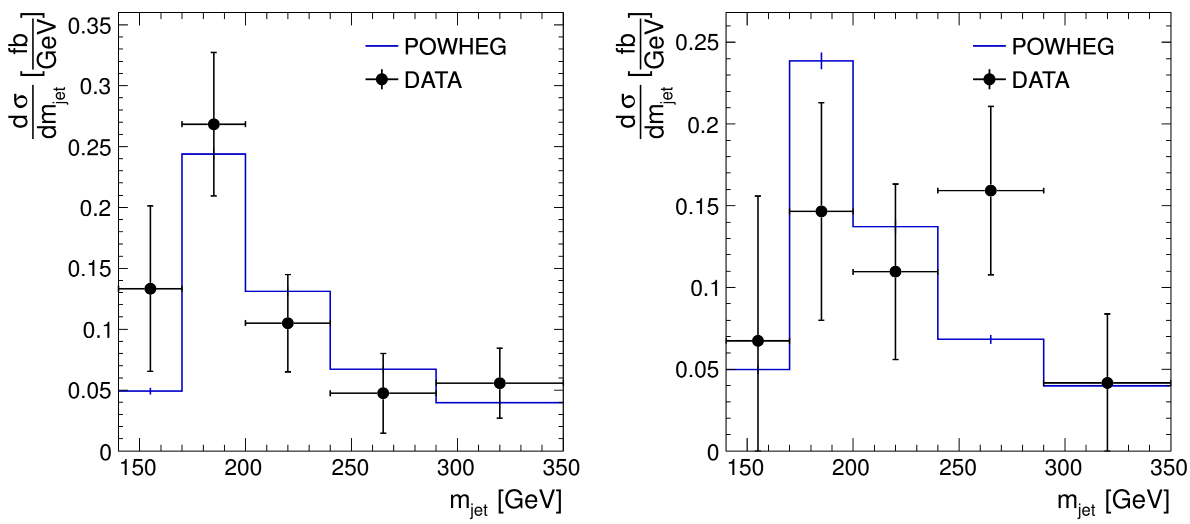

Figure D.1.: Measured differential cross section as a function of the leading jet mass in data, unfolded with migrations obtained from POWHEG, and compared to the particle level distribution in POWHEG. The electron channel is shown on the left and the muon channel on the right. The error bars include statistical errors only.

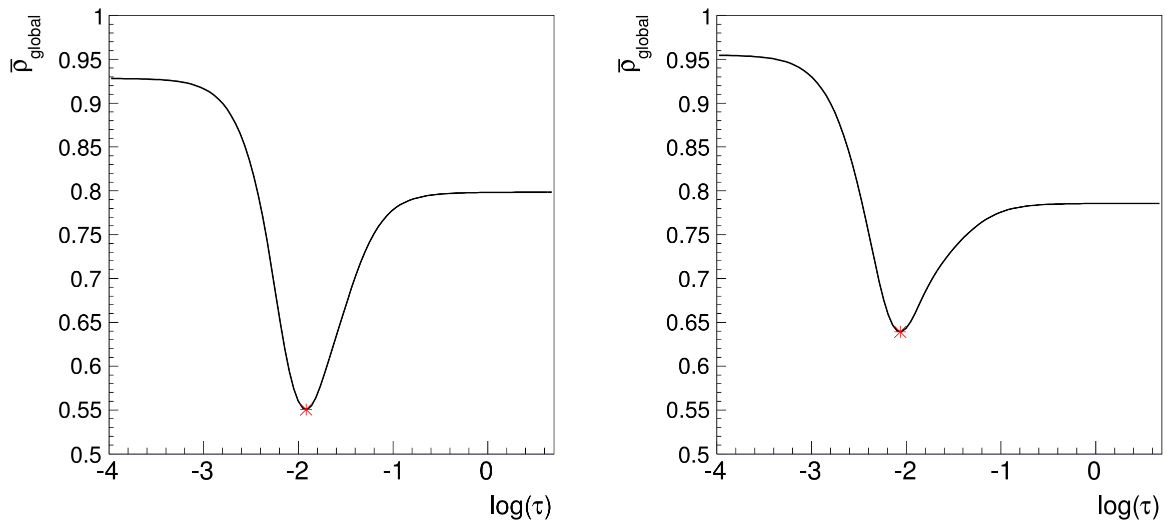

Figure D.2.: Scans over the average global correlation coefficient for the unfolding of the data in the electron channel on the left and the muon channel on the right. Systematic uncertainties are included in the calculation of global correlation coefficients. The red marker shows the tau value that has been chosen.

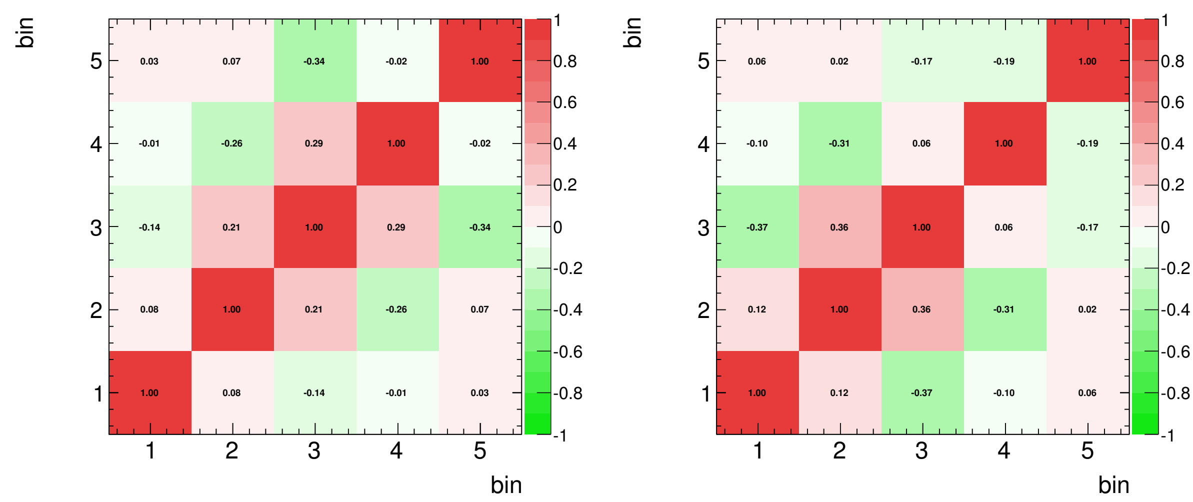

Figure D.3.: Correlation coefficients of the output in the electron channel on the left and the muon channel on the right, including statistical uncertainties only.

Bibliography

[1] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for top-quark partners with charge 5/3 in the same-sign dilepton final state," *Phys. Rev. Lett.* **112** no. 17, (2014) 171801, arXiv:1312.2391 [hep-ex].

[2] **CMS** Collaboration, "Search for W prime to tbbar in the all-hadronic final state," Tech. Rep. CMS-PAS-B2G-12-009, CERN, Geneva, 2014. https://cds.cern.ch/record/1751504. CMS-PAS-B2G-12-009.

[3] **CMS** Collaboration, V. Khachatryan *et al.*, "Search for vector-like T quarks decaying to top quarks and Higgs bosons in the all-hadronic channel using jet substructure," arXiv:1503.01952 [hep-ex].

[4] **CMS** Collaboration, S. Chatrchyan *et al.*, "Searches for new physics using the $t\bar{t}$ invariant mass distribution in pp collisions at $\sqrt{s}$=8 TeV," *Phys. Rev. Lett.* **111** no. 21, (2013) 211804, arXiv:1309.2030 [hep-ex]. CMS-B2G-13-001.

[5] **CMS** Collaboration, "Boosted Top Jet Tagging at CMS,". CMS-PAS-JME-13-007.

[6] The ATLAS and The CDF and The CMS and D0 Collaborations, "First combination of Tevatron and LHC measurements of the top-quark mass," 1403.4427.

[7] A. Jain, I. Scimemi, and I. W. Stewart, "Two-loop Jet-Function and Jet-Mass for Top Quarks," *Phys. Rev. D* **778** (Jan., 2008) 094008, 0801.0743.

[8] S. Schmitt, "TUnfold: an algorithm for correcting migration effects in high energy physics," 1205.6201.

[9] D. Griffiths, *Introduction to Elementary Particles.* John Wiley & Sons, New York, USA, 1987.

[10] **CMS** Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett. B* **716** (July, 2012) 30, 1207.7235. CMS-HIG-12-028.

[11] **ATLAS** Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett. B* **716** (July, 2012) 1–29, 1207.7214.

[12] https://www.stfc.ac.uk.

[13] K.A. Olive et al., "Particle Data Group," *Chin. Phys. C* **38** (2014) 090001.

[14] **Gargamelle Neutrino** Collaboration, F. Hasert *et al.*, "Observation of Neutrino Like Interactions without Muon or Electron in the Gargamelle Neutrino Experiment," *Nucl. Phys.* **B73** (1974) 1–22.

[15] S. Abachi, "Observation of the Top Quark," *Phys. Rev. Lett.* **74** (1995) 2632–2637, `hep-ex/9503003`.

[16] **CDF** Collaboration, "Observation of Top Quark Production in Pbar-P Collisions," *Phys. Rev. Lett.* **74** (1995) 2626–2631, `hep-ex/9503002`.

[17] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha\frac{4}{S})$," *Phys. Rev. Lett.* **110** (2013) 252004, `arXiv:1303.6254 [hep-ph]`.

[18] G. Degrassi, S. D. Vita, J. Elias-Miró, J. R. Espinosa, *et al.*, "Higgs mass and vacuum stability in the Standard Model at NNLO," `1205.6497`.

[19] M. Baak, M. Goebel, J. Haller, A. Hoecker, *et al.*, "The Electroweak Fit of the Standard Model after the Discovery of a New Boson at the LHC," *Eur. Phys. J. C* **72** (Sept., 2012) 2205, `1209.2716`.

[20] **CMS** Collaboration, "Combination of ATLAS and CMS top quark pair cross section measurements in the emu final state using proton-proton collisions at 8 TeV,". CMS-PAS-TOP-14-016.

[21] **D0** Collaboration, M. Abazov, V. B. Abbott, S. Acharya, B. M. Adams, *et al.*, "Precision Measurement of the Top Quark Mass in Lepton + Jets Final States," *Phys. Rev. Lett.* **113** (Jul, 2014) 032002. `http://link.aps.org/doi/10.1103/PhysRevLett.113.032002`.

[22] P. Skands and D. Wicke, "Non-perturbative QCD Effects and the Top Mass at the Tevatron," *Eur. Phys. J. C* **52** (2007) 133–140, `hep-ph/0703081`.

[23] A. H. Hoang and I. W. Stewart, "Top Mass Measurements from Jets and the Tevatron Top-Quark Mass," *Nucl. Phys. Proc. Suppl.* **185** (Aug., 2008) 220–226, `0808.0222`.

[24] S. Fleming, A. H. Hoang, S. Mantry, and I. W. Stewart, "Jets from Massive Unstable Particles: Top-Mass Determination," *Phys. Rev. D* **77** (2008) 074010, `hep-ph/0703207`.

[25] S. Fleming, A. H. Hoang, S. Mantry, and I. W. Stewart, "Top Jets in the Peak Region: Factorization Analysis with NLL Resummation," *Phys. Rev. D* **77** (Nov., 2007) 114003, 0711.2079.

[26] R. Placakyte, for the H1 Collaboration, and for the ZEUS Collaboration, "Parton Distribution Functions," 1111.5452.

[27] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5 : Going Beyond," 1106.0522.

[28] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *JHEP* **1006** (Feb., 2010) 043, 1002.2581.

[29] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP* **0605** (2006) 026, hep-ph/0603175.

[30] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, *et al.*, "HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)," *JHEP* **0101** (2001) 010, hep-ph/0011363.

[31] J. Allison, K. Amako, J. Apostolakis, H. Araujo, *et al.*, "Geant4 developments and applications," *Nuclear Science, IEEE Transactions on* **53** no. 1, (Feb., 2006) 270 –278.

[32] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, *et al.*, "Geant4—a simulation toolkit," *NIM A* **506** no. 3, (2003) 250 – 303.

[33] https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[34] P. B. Lyndon Evans, "LHC Machine," *Journal of Instrumentation* **3** (2008) S0801.

[35] **CMS** Collaboration, "The CMS experiment at the CERN LHC," *Journal of Instrumentation* **3** (2008) S08003.

[36] **CMS** Collaboration, G. L. Bayatian, S. Chatrchyan, G. Hmayakyan, A. M. Sirunyan, *et al.*, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software.* Technical Design Report CMS. CERN, Geneva, 2006.

[37] **CMS** Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET,". CMS-PAS-PFT-09-001.

[38] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better Jet Clustering Algorithms," *JHEP* **9708** (1997) 001, hep-ph/9707323.

[39] S. D. Ellis and D. E. Soper, "Successive Combination Jet Algorithm For Hadron Collisions," *Phys. Rev. D* **48** (1993) 3160–3166, hep-ph/9305266.

[40] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-k_t jet clustering algorithm,” *JHEP* **0804** (Feb., 2008) 063, 0802.1189.

[41] **CMS** Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS,” *JINST* **6** (July, 2011) 11002, 1107.4277.

[42] **CMS** Collaboration, “Pileup Removal Algorithms,” tech. rep., CERN, Geneva, 2014. CMS-PAS-JME-14-001.

[43] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, “Jet substructure as a new Higgs search channel at the LHC,” *Phys. Rev. Lett.* **100** (Feb., 2008) 242001, 0802.2470.

[44] D. Krohn, J. Thaler, and L.-T. Wang, “Jet Trimming,” *JHEP* **1002** (Dec., 2009) 084, 0912.1342.

[45] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, “Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches,” *Phys. Rev. D* **81** (Dec., 2009) 094023, 0912.0033.

[46] T. Plehn and M. Spannowsky, “Top Tagging,” 1112.4441.

[47] **CMS** Collaboration, “Missing transverse energy performance of the CMS detector,” *JINST* **6** (June, 2011) 09001, 1106.5048. CMS-JME-10-009.

[48] **CMS** Collaboration, S. Chatrchyan *et al.*, “Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV,” *JINST* **7** (2012) P10002, arXiv:1206.4071 [physics.ins-det]. CMS-MUO-10-004.

[49] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, *et al.*, “Electron reconstruction in CMS,” *Eur. Phys. J.* **C49** (2007) 1099–1116.

[50] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *ASME. J. Fluids Eng.* **82(1)** (1960) 35–45.

[51] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, “Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC,”.

[52] **CMS** Collaboration, “https://twiki.cern.ch/twiki/bin/viewauth/CMS/MultivariateElectronIdentification,”.

[53] **CMS** Collaboration, “Identification of b-quark jets with the CMS experiment,” *JINST* **8** (Nov., 2012) P04013, 1211.4462. CMS-BTV-12-001.

[54] Frank Tackmann, private communication.

[55] **CMS** Collaboration, "Search for ttbar resonances in semileptonic final state,". CMS-PAS-B2G-12-006.

[56] **CMS** Collaboration, "Measurement of the differential top-quark pair production cross section in the dilepton channel in pp collisions at p s = 8 TeV,". CMS-PAS-TOP-12-028.

[57] **CMS** Collaboration, "Measurement of differential top-quark pair production cross sections in the lepton+jets channel in pp collisions at 8 TeV,". CMS-PAS-TOP-12-027.

[58] **CMS** Collaboration, "Identification techniques for highly boosted W bosons that decay into hadrons," *JHEP* **12** (Oct., 2014) 017, `1410.4227`. CMS-JME-13-006.

[59] **CMS** Collaboration, "Measurement of the production cross section for a W boson and two b jets in pp collisions at sqrt(s) = 7 TeV," *Phys. Lett. B* **735** (Dec., 2013) 204, `1312.6608`. CMS-SMP-12-026.

[60] **CMS** Collaboration, S. Chatrchyan *et al.*, "Observation of the associated production of a single top quark and a W boson in pp collisions at $\sqrt{s}$ =8 TeV," *Phys.Rev.Lett.* **112** no. 23, (2014) 231802, `arXiv:1401.2942 [hep-ex]`.

[61] **CMS** Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update,". CMS-PAS-LUM-13-001.

[62] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup Per Particle Identification," *JHEP* **1410** (2014) 59, `arXiv:1407.6013 [hep-ph]`.

Danksagung

Zu aller erst möchte ich mich bei Prof. Dr. Johannes Haller bedanken, für die Möglichkeit an diesem Thema arbeiten zu dürfen, und die kontinuierlich Rücksprache während der gesamten Zeit dieser Arbeit. Als nächstes möchte ich Prof. Dr. Peter Schleper danken, dass er sich dazu bereit erklärt hat, als Zweitgutachter für diese Arbeit zur Verfügung zu stehen. Weiterhin gebührt Dr. Roman Kogler großer Dank für die intensive Rücksprache bei allen Problemen dieser Arbeit in technischer und physikalischer Hinsicht. Außerdem danke ich Dr. Frank Tackmann für die Rücksprache in theoretischen Fragen. Abschließend danke ich der gesamten Arbeitsgruppe für ein angenemes und produktives Arbeitsumfeld und viele interessante Gespräche.

Erklärung

Hiermit bestätige ich, dass die vorliegende Arbeit von mir selbstständig verfasst wurde und ich keine anderen als die angegebenen Hilfsmittel - insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen - benutzt habe und die Arbeit von mir nicht einem anderen Prüfungsverfahren eingereicht wurde. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Ich bin damit einverstanden, dass die Masterarbeit veröffentlicht wird.

Ort, Datum Unterschrift