Precise measurement of the top-quark mass at the CMS experiment using the ideogram method

Dissertation zur Erlangung des Doktorgrades an der Fakultät für Mathematik, Informatik und Naturwissenschaften Fachbereich Physik der Universität Hamburg

vorgelegt von

Markus Seidel

Hamburg, 2015

Gutachter der Dissertation:	Prof. Dr. Peter Schleper
	Prof. Dr. Wolfgang Wagner
	Prof. Dr. Ian Brock
Mitglieder der Prüfungskommission:	Prof. Dr. Ludwig Mathey (Vorsitz)
	Prof. Dr. Sven-Olaf Moch
	Dr. Andreas Meyer
Datum der Disputation:	23. Juni 2015

Abstract

The mass of the top quark is measured using a sample of $t\bar{t}$ candidate events with one electron or muon and at least four jets in the final state, collected by CMS in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ at the LHC. The candidate events are selected from data corresponding to an integrated luminosity of $19.7\,\text{fb}^{-1}$. For each event the top-quark mass is reconstructed from a kinematic fit of the decay products to a $t\bar{t}$ hypothesis. In order to minimize the uncertainties from jet energy corrections, the top-quark mass is determined simultaneously with a jet energy scale factor (JSF), constrained by the known mass of the W boson decaying to quark-antiquark pairs. A joint likelihood fit taking into account multiple interpretations per event – the ideogram method – is used. From the simultaneous fit, a top-quark mass of $172.15 \pm 0.19\,(\text{stat.}+\text{JSF}) \pm 0.61\,(\text{syst.})$ GeV is obtained. Using an additional constraint from the determination of the jet energy scale in γ/Z+jet events yields $m_t = 172.38 \pm 0.16\,(\text{stat.}+\text{JSF}) \pm 0.49\,(\text{syst.})$ GeV. The results are discussed in the context of different event generator implementations. Possible kinematic biases are studied by performing the measurement in different regions of the phase space.

Zusammenfassung

Die Masse des Topquarks wird in Ereignissen mit einem Myon oder einem Elektron und mindestens vier Jets gemessen, die vom CMS-Experiment in Proton-Proton-Kollisionen bei $\sqrt{s} = 8\,\text{TeV}$ am LHC aufgezeichnet wurden und einer integrierten Luminosität von $19.7\,\text{fb}^{-1}$ entsprechen. Für jedes Ereignis wird die Topquarkmasse durch einen kinematischen Fit der Zerfallsprodukte rekonstruiert. Um die Unsicherheit durch die Jetenergiekorrekturen zu minimieren, wird die Topquarkmasse gleichzeitig mit einem Jetenergie-Skalenfaktor (JSF) bestimmt, der aus der bekannten Masse des W-Bosons, das in Quark-Antiquark-Paare zerfällt, gewonnen wird. Die Ideogrammethode berücksichtigt dabei mehrere mögliche Ereignisinterpretationen. Aus dem simultanen Fit wird eine Topquarkmasse von $172.15 \pm 0.19\,(\text{stat.}+\text{JSF})$ $\pm 0.61\,(\text{syst.})$ GeV bestimmt. Wird der JSF zusätzlich entsprechend der Jetenergiekalibration aus γ/Z+Jet-Ereignissen eingeschränkt, ergibt sich eine gemessene Topquarkmasse von $172.38 \pm 0.16\,(\text{stat.}+\text{JSF}) \pm 0.49\,(\text{syst.})$ GeV. Diese Ergebnisse werden im Kontext verschiedener Simulationen diskutiert. Die Abhängigkeit des Ergebnisses vom betrachteten Phasenraum wird bestimmt.

Contents

1. Introduction

Particle physics probes the fundamental building blocks of matter and the interactions between them, which combine to form the substances of our daily life. The top quark was discovered at the Fermilab Tevatron in the year 1995 [1, 2], and is the heaviest known fundamental particle to date. Due to its large mass, the production of top quarks requires particle collisions at very high energies. The CERN Large Hadron Collider (LHC) [3] achieves a higher energy and collision rate compared to the Tevatron, thus making it a top-quark "factory". The large number of top-quark events recorded by the LHC experiments therefore allows for an unprecedented precision in the determination of the top-quark properties and interactions.

The top-quark mass (m_t) being measured at the Tevatron and the LHC is an essential parameter of the Standard Model of particle physics (SM). Since the discovery of a Higgs-Boson at the LHC in 2012 [4, 5], the precise knowledge of the top-quark mass allows for probing the internal consistency of the Standard Model [6–8]. In addition, the instability scale of the electroweak vacuum was calculated to higher precision than before [9, 10], commonly interpreted as the energy scale at which Physics beyond the Standard Model (BSM) is needed for a consistent description of nature. So far, no BSM phenomena have been directly observed in high-energy collider physics.

Different methods are employed for the measurement of the top-quark mass. They all commonly require an advanced understanding of the experiment used for detecting the decay products of the unstable top quark. Every possible decay mode of the top quark provides its own challenges for top-quark reconstruction, and constitutes a benchmark for detector performance and calibration by itself.

The measurement presented in this thesis is performed in the "lepton+jets" channel, shown in Figure 1.1, where the two top quarks decay via $t \to bW$ and one W boson decays into a quark-antiquark pair while the other one decays into a charged lepton and the corresponding neutrino. This final state therefore consists of a lepton, at least four jets (two bottom jets and two light jets from the top-quark decays),

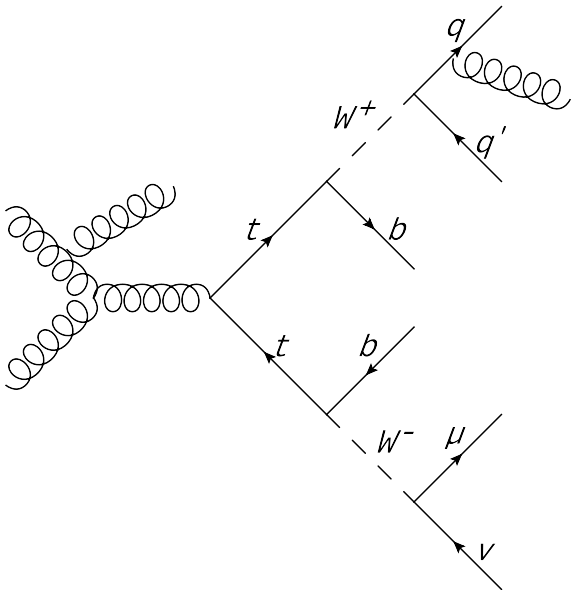

Figure 1.1.: Top-quark pair production and decay in the muon+jets channel.

and an undetected neutrino. The lepton+jets channel combines good rejection of background processes with the ability to fully reconstruct the event kinematics.

The analysis employs a kinematic fit of the decay products to a $t\bar{t}$ hypothesis that enhances the fraction of correctly reconstructed top quarks. For each event, two-dimensional likelihood functions are used to estimate simultaneously the top-quark mass and a jet energy scale factor (JSF) from the reconstructed masses of both the top quark and the W boson in a 2D approach. The top-quark mass is also extracted with fixing the JSF to unity (1D approach), and with a hybrid approach that combines all available information and gains from anti-correlated systematic uncertainties between the 2D and 1D approaches. As the measurement presented in this thesis is not limited by statistical precision, there is the unique possibility to access its phase-space dependence.

In this thesis, the measurement using the lepton+jets channel at a center-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$ is described, where a preliminary version is already publicly available [11]. In addition, the methods presented here have been successfully applied in the following analyses: the measurement of the top-quark mass at $\sqrt{s} = 7\,\text{TeV}$ [12, 13], its phase-space dependence [14], and the CMS measurements in the all-jets channel [15–17]. The analysis setup was also used as benchmark for improving the energy resolution of bottom jets using a regression technique [18].

This thesis is structured as follows: Chapter 2 gives an overview of the Standard Model describing the fundamental particles and interactions, the role of the top quark in this theory, and its phenomenology. The calculation of physical observables is performed by event generators described in Chapter 3. The experimental setup is depicted in Chapter 4. Chapter 5 is devoted to the event reconstruction. Chapter 6 describes the selection of a pure sample of $t\bar{t}$ events and the kinematic fit. Using the selected events, the measurement of the top-quark mass is performed with the ideogram method portrayed in Chapter 7. Systematic uncertainties are discussed in Chapter 8. The result is presented in Chapter 9, in which also the compatibility with other measurements is discussed. Chapter 10 is devoted to the study of top-quark reconstruction at generator level using different setups and to the tuning of the b-fragmentation function. The phase-space dependence of the measurement is investigated in Chapter 11. A summary and an outlook for future measurements of the top-quark mass are given in Chapter 12.

2. The top quark in the Standard Model

2.1. The Standard Model of particle physics

The Standard Model of particle physics (SM) describes the interactions of fundamental particles and has been experimentally verified up to the TeV energy scale, which corresponds to probing distances of the order of 10^{-19} m, four orders of magnitude below the size of a proton.

Units

As usual in high-energy physics, energies are given in the unit of electron volts, where 1 eV $= 1.6 \times 10^{-19}$ J is the kinetic energy an electron gains when it is accelerated through a potential difference of 1 V. Using natural units, the speed of light c and reduced Planck constant $\hbar$ are equal to 1. Momenta and masses are then also given in eV, according to the relativistic energy-momentum relationship $E^2 = p^2 + m^2$. The relation between energy and length is determined by $1 = \hbar c \approx 200$ MeV fm, so that lengths have a unit of inverse energy.

In energy-momentum space, a particle is characterized by its momenta in the three spatial dimensions and either its mass or total energy. If the z–axis is chosen to be along the beam axis in a collider experiment, it is convenient to define a particle with its total energy, the momentum component transverse to the beam axis $p_\mathrm{T} = \sqrt{p_x^2 + p_y^2}$, and its direction in $\eta - \phi$ space, where ϕ is the azimuthal angle around the beam and the pseudo-rapidity η is $-\ln\tan(\theta/2)$ with the polar angle θ measured from the beam axis. The difference in the direction of flight of two particles is given as their distance in $\eta - \phi$ space $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

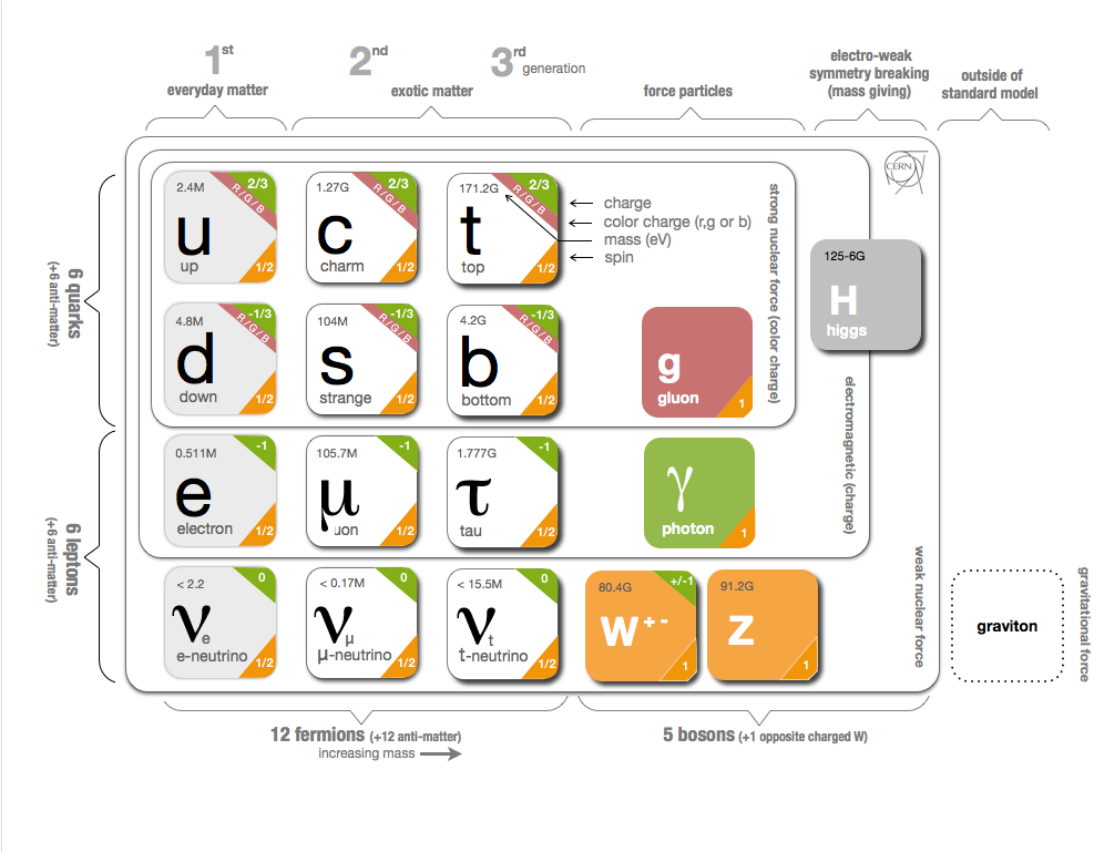

Figure 2.1.: The particles of the Standard Model. Graphic by [21].

Basic structure of the Standard Model

The fundamental matter particles consist of fermions, while forces are described by bosons, as shown in Figure 2.1. Distinct combinations of color-charged quarks form nucleons (protons and neutrons), other hadrons and atomic nuclei, tied together by the strong force that is mediated by gluons. Quarks and charged leptons like the electron carry electromagnetic charge that allows interaction with photons. Uncharged leptons (neutrinos) only interact via the weak force that is responsible for nuclear decay and fusion processes. The weak force is the only force that allows transitions between the fermions through the decay of heavier to lighter quarks or leptons. The Brout-Englert-Higgs mechanism of electroweak symmetry breaking introduces an additional Higgs boson, and gives mass to the W and Z bosons of the weak force, as opposed to the massless gluons and the photon. Modern introductory texts of varying depth can be found in the literature, the following presentation is based on [19, 20].

The SM is a renormalizable quantum field theory defined by the gauge group

$$SU(3)_C \times SU(2)_L \times U(1)_Y .$$

$SU(3)_C$ describes Quantum Chromodynamics (QCD), the strong interaction between quarks and gluons, where C stands for the "color" charge. $SU(2)_L \times U(1)_Y$ is the gauge group of electroweak interactions, the unification of the weak force and electromagnetism. The L stands for "left-handed", indicating the chirality of weak interactions, and Y for "hypercharge" in order to distinguish the group from the $U(1)_{EM}$ of Quantum Electrodynamics (QED).

Gauge theories

Gauge theories are based on the concept of invariance under specific symmetry transformations. Noether's theorem states that continuous symmetries of a theory give rise to conserved quantities. The simplest example for a symmetry and an associated conservation law is that the equations of motion for a free particle do not depend on its space-time coordinates (translational invariance), as there are no interactions of any kind. Then the conserved quantities are energy and momentum of the particle.

Continuous symmetry transformations are described in the framework of group theory by Lie groups. One important Lie group is the Lorentz group that is generated by rotations and boosts in 4-dimensional space-time. Lorentz transformations Λ obey $\eta_{\mu\nu}\Lambda^\mu_\rho\Lambda^\nu_\lambda = \eta_{\rho\lambda}$, where $\eta_{\mu\nu} = \left(\begin{smallmatrix} 1 & \\ & -\mathbf{1} \end{smallmatrix}\right)$ is the metric tensor of special relativity and $\mathbf{1}$ is understood to be the 3×3 identity matrix. This defines Lorentz transformations to preserve the dot product of four-vectors: for a four-vector $v^\mu = (t, x, y, z)^T$ the dot-product with itself is $\eta_{\mu\nu}v^\mu v^\nu = t^2 - x^2 - y^2 - z^2$, giving the same value for an observer of the physical system in any moving or rotated reference frame. Another important four-vector is $w^\mu = (E, p_x, p_y, p_z)^T$, where the dot-product with itself, $E^2 - \vec{p}^2 = m^2$, is exactly the squared invariant mass of the particle. The underlying algebra of the Lorentz group $SO(1,3)$ reveals that it contains two copies of the group $SU(2)$ that also describes quantum mechanical spin. Therefore, the quantum fields describing the particles have well-defined transformation properties, depending on their spin. There are scalar fields for spin-0 particles that are invariant under Lorentz transformations, as well as spinor fields (spin-1/2) and vector fields (spin-1) with more complex transformation properties.

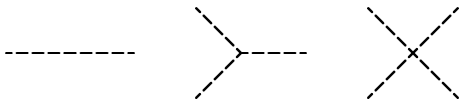

Figure 2.2.: Feynman diagrams of a self-interacting scalar field, showing particle propagation and two possible types of interaction.

The classical dynamics of each particle field are given by the Euler-Lagrange equation

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi \right)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0,$$

where the Lagrangian density $\mathcal{L} = \mathcal{T} - \mathcal{V}$ is the difference of kinetic and potential energy density. The Lagrangian is required to be invariant under Lorentz transformations. For free particles, $\mathcal{L}_{free}$ contains a kinetic term like $\partial^\mu \phi \partial_\mu \phi$ (from the requirement of invariance under translations and boosts – Poincaré invariance) and a mass term quadratic in the field, $m^2\phi^2$ for the example of a scalar field. The Euler-Lagrange equation then recovers the Klein-Gordon equation for scalar bosons, the Dirac equation for fermions, and the Proca equation for vector bosons from the respective Lagrangian.

Interactions can be included in the theory by adding terms like $\lambda_3\phi^3$ or $\lambda_4\phi^4$ to the Lagrangian, where λ denotes the coupling constant. Note that the mass can be interpreted as the coupling strength of a particle with itself when propagating from one point to another. These terms can be represented as Feynman diagrams, joining multiple particles in an interaction vertex, as shown in Figure 2.2. In principle one could add further interaction terms, involving also other particles, "by hand".

A consistent interacting theory is achieved by requiring local gauge invariance of the free fermion field under transformations of a certain Lie group. Then the force is generated by the specifying group, as an interaction term with a gauge field is required to maintain local gauge invariance. Furthermore, a kinetic term for the (massless) gauge field needs to be added. The transformations of the Lie group can be written as $e^{i\alpha_i X_i}$, where the α_i are continuous parameters (for example angles for rotation groups) and the X_i are matrices called the generators of the group. The bosonic gauge fields can be expressed in terms of linear combinations of the generators of the Lie group. The eigenvectors of diagonal ("Cartan") generators correspond to particle states, and the eigenvalues to the charges. The non-Cartan generators are combined such that they shift particles from one state to another, like gluons or W bosons.

Quantization

Canonical quantization of a field theory is done by promoting the (scalar) field ϕ and its momentum $\Pi = \frac{\partial \mathcal{L}}{\partial \dot{\phi}}$ to operators obeying the commutation relation

$$[\phi(x), \Pi(x')] = i\delta^{(4)}(x - x').$$

This is the field version of Heisenberg's uncertainty principle, stating that it is impossible to determine both the field's value and momentum with arbitrary precision. Expansion of the fields in plane waves (that are solutions to the equations of motion) leads to the creation and annihilation operators $a_p^\dagger$ and a_p, where p is the particle momentum. These allow for the construction of the Fock space of multiparticle states by acting repeatedly on the vacuum state $|0\rangle$. Complex fields give rise to particles and antiparticles with opposite electromagnetic charge. For the quantization of fermion fields Ψ, an anticommutation relation is imposed, so that the resulting multiparticle state is antisymmetric under the exchange of two particles. The scattering amplitude for a process is then given by the vacuum expectation value of the time-ordered product of the involved particle fields (one for each in/outgoing particle) and a term containing the interaction Hamiltonian $\mathcal{H}_I$,

$$\begin{aligned} & \langle p_1 \dots p_n \,|\, iT \,|\, k_1 \dots k_m \rangle \\ \propto & \int d^4x_1 \dots d^4x_{m+n} e^{i(p_1x_1 + \dots + p_nx_n - k_1x_{n+1} - \dots - k_mx_{m+n})} \\ & \times \left\langle 0 \,\middle|\, T\left\{\phi(x_1) \dots \phi(x_{m+n}) \exp\left[-i\int d^4x \mathcal{H}_I\right]\right\} \,\middle|\, 0 \right\rangle, \end{aligned}$$

where k denote the incoming and p the outgoing momenta, and $\mathcal{H}_I$ contains the same interaction term as the Lagrangian. Using Wick's theorem, this expression can be evaluated in orders of the coupling λ, and systematic study of the result leads to a set of Feynman rules for the theory. These rules can be seen as intermediate steps in the calculation of the amplitude that is graphically represented as a Feynman diagram.

The path integral quantization treats particles as being in a superposition of all possible states. The amplitude for a particle at q_1, t_1 going to q_2, t_2 is then given by

$$\langle q_2, t_2 \,|\, q_1, t_1 \rangle = \int \prod_{i=1}^{\infty} dq_i e^{i\int_{t_1}^{t_2} dtL} = \int \mathcal{D}q e^{iS},$$

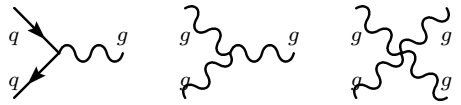

Figure 2.3.: Interaction terms of QCD: Quark/antiquark annihilation/creation or gluon radiation (left, depending on direction of time), gluon→gluon splitting (center), and four-gluon interaction (right).

where $\mathcal{D}q$ represents the summation over all possible paths of the particle. The most probable path is then the one that minimizes the action S, which recovers the principle of least action from classical mechanics. For small couplings λ, the expression can be evaluated order-by-order in perturbation theory, resulting in the same Feynman rules as the canonical quantization. In addition, the amplitudes of the path integral formulation can be evaluated non-perturbatively using a lattice of space-time points.

Quantum Chromodynamics

The underlying group of QCD is $SU(3)_C$ in the fundamental representation of eight 3×3 Gell-Mann matrices. There are three "color" charges denoted "red", "blue" and "green" that are acted on by the gluons corresponding to the generators. As the gluons are color-charged they can interact with themselves, the basic QCD vertices are shown in Figure 2.3. The gluon self-interaction leads to an effective coupling constant α_s that is small for high energies (asymptotic freedom) and large for low energies, corresponding to distances of the order of fm, as shown in Figure 2.4. As a result, the potential energy between spatially separated color charges becomes so large that new quark-antiquark pairs are created and form color-neutral hadrons with the original quarks (confinement).

Electroweak interactions and symmetry breaking

The electroweak force is described by the gauge group $SU(2)_L \times U(1)_Y$ with four initially massless vector boson gauge fields. In addition there is a complex scalar Higgs field with a "wine bottle" potential $V_H = \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2$, where the vacuum expectation value is ~ 246 GeV. Below the energy scale of electroweak unification, the $SU(2)_L \times U(1)_Y$ symmetry of the Higgs field is broken, and the effective low-energy theory contains linear combinations of the original vector fields, of which three gain masses (the $W^{+/-}$ and Z^0 bosons) while one remains massless (the pho-

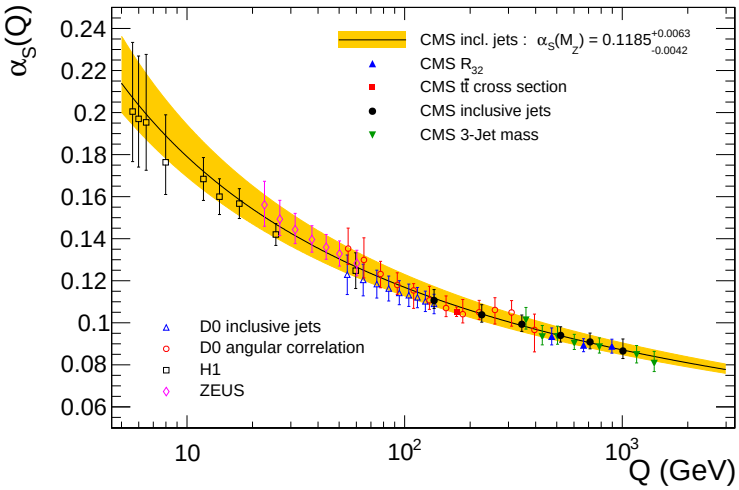

Figure 2.4.: The value of the strong coupling α_s depending on the energy scale. The running of α_s has been verified over three orders of magnitude. Plot from [22].

ton). Their interactions are shown in Figure 2.5. In addition, a massive scalar Higgs boson emerges that couples to all particles, giving them a gauge invariant mass term. The Higgs self-coupling is similar to that of the scalar field in Figure 2.2.

Weak decays of quarks are mediated by W bosons. The branching ratios are given by the squared elements $|V_{ij}|^2$ of the CKM matrix [23,24] that describes the mixing of the mass eigenstates q and the electroweak eigenstates q':

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix} = \begin{bmatrix} 0.9743 & 0.2253 & 0.0035 \\ 0.2252 & 0.9735 & 0.0410 \\ 0.0086 & 0.0403 & 0.9992 \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}.$$

The CKM matrix can be parametrized by three mixing angles and one CP-violating complex phase as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{13}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{13}} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{23}c_{13} \end{bmatrix}$$

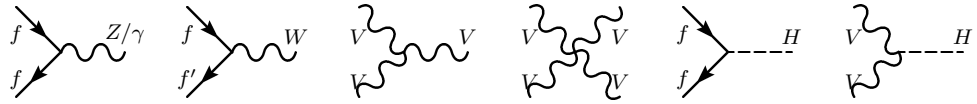

Figure 2.5.: Interaction terms of the electroweak and Higgs sector: Annihilation/creation of fermion/antifermion pairs or radiation of a Z-boson or photon off fermions (left, depending on direction of time), flavor-changing charged currents of $W^{\pm}$ bosons, self-interactions of the heavy vector bosons $W^{\pm}$ and Z (where charge conservation is understood), Higgs-fermion coupling, and Higgs–W/Z coupling.

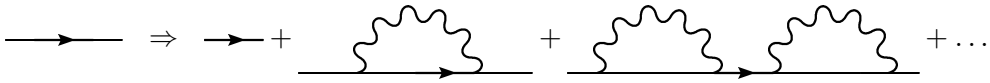

Figure 2.6.: Examples for loop contributions to the fermion propagator.

where s_{ij} and c_{ij} denote the sine and cosine of the mixing angles $\theta_{12} = 13.04 \pm 0.05°$, $\theta_{13} = 0.201 \pm 0.011°$, $\theta_{23} = 2.38 \pm 0.06°$, and the value of the complex phase is $\delta_{13} = 1.20 \pm 0.08$ rad.

As the CKM matrix differs from the identity matrix, the mass and electroweak eigenstates are not identical, and the charged currents of the weak interaction couple to quarks from different generations.

2.2. Top-quark mass and electroweak fits

Renormalization and mass schemes

When the perturbative expansion for the scattering amplitude is taken above the leading order (LO), the Feynman graphs contain closed particle loops. Evaluation of these loops results in divergences that can be canceled by a renormalization procedure for couplings, fields and masses. The quantities then become dependent on an energy scale Λ used to regulate divergences at either high or low energy scales, dependent on the theory. A theory is renormalizable when the Lagrangian contains only couplings that are dimensionless or have a dimension of positive mass. The Fermi theory of weak interactions ($G_F \propto 1/m_W^2$) or gravity ($G \propto 1/m_{Pl}^2$) are not renormalizable but still useful effective field theories for energies below their characteristic mass scale.

The need for renormalization of particle masses arises from loop contributions to the particle propagator, shown in Figure 2.6, that change the (fermion) propagator in the following way:

$$\frac{i}{\not{p} - m_0} \quad \Rightarrow \quad \frac{i}{\not{p} - \underbrace{m_0(\Lambda)}_{\text{'bare' mass}} - \underbrace{\delta m_0(\Lambda)}_{\text{divergent}} - \underbrace{\Sigma'(m_0(\Lambda))}_{\text{finite}}} := \frac{i}{\not{p} - m^{pole}}$$

There are divergent and finite terms in the denominator of the propagator. Now the particle mass in the pole mass scheme can be redefined as absorbing all additional terms. This corresponds to the mass of a free particle measured from a large distance, or at $E \to 0$. Thus, the pole mass scheme is convenient for particles like leptons that are subject only to electroweak (and gravitational) interactions, where the potential approaches zero at infinite distance and the coupling becomes small.

On the other side, the pole mass of the strongly-interacting quarks carries a renormalon ambiguity of the order of $\Lambda_{QCD} \approx 200$ MeV [25]. Below this energy, the strong coupling α_s becomes numerically large and perturbation theory is not applicable anymore. It is therefore more convenient to define a short-distance mass like the one in the $\overline{\text{MS}}$ scheme. This scheme absorbs only the divergent terms into the mass definition, so that the renormalized mass $\bar{m}$ depends on the scale Λ.

The difference between the top-quark pole and $\overline{\text{MS}}$ mass when taking into account QCD corrections up to $\mathcal{O}(\alpha_s^3)$ is calculated to be [26]

$$\left\{\bar{m}_t - m_t^{pole}\right\}_{QCD} = m_t^{pole}\left[-\frac{4}{3}\frac{\alpha_s}{\pi} - 9.125\left(\frac{\alpha_s}{\pi}\right)^2 - 80.405\left(\frac{\alpha_s}{\pi}\right)^3\right],$$

where α_s and $\bar{m}_t$ are evaluated at the scale m_t^{pole}. For a pole mass of 173.5 GeV, this results in a difference of -10.38 GeV. However, the inclusion of electroweak corrections at $\mathcal{O}(\alpha)$ and $\mathcal{O}(\alpha\alpha_s)$ yields

$$\left\{\bar{m}_t - m_t^{pole}\right\}_{SM} = \left\{\bar{m}_t - m_t^{pole}\right\}_{QCD} + m_t^{pole}\left[0.0664 - 0.00115 \times \left(\frac{m_H}{1\text{ GeV}} - 125\right)\right],$$

where the largest contribution stems from "tadpole" diagrams [27]. With a top-quark pole mass $m_t^{pole} = 173.5$ GeV and $m_H = 125$ GeV the electroweak corrections have a numerical value of $+11.52$ GeV. Thus, the QCD corrections are approximately compensated and the numerical difference between the pole and $\overline{\text{MS}}$ mass shrinks to the order of 1 GeV, so that $m_t^{pole} = 173.50$ GeV corresponds to $\bar{m}_t\left(m_t^{pole}\right) = 174.64$ GeV.

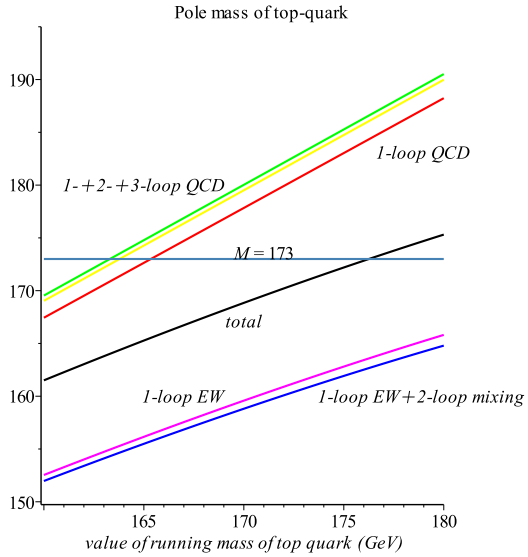

Figure 2.7.: Relation between the top-quark mass in pole and $\overline{\text{MS}}$ schemes, including either QCD or electroweak corrections, or both (=total). Plot from [28].

A generalization is the MSR scheme, where another scale R is associated with the scheme, so that $m_t^{pole} = m_t^{MSR}(R,\mu) + \delta m_t(R,\mu)$ [29]. For $R = m_t$, the $\overline{\text{MS}}$ scheme is reproduced, while $R = 0$ leads to the pole mass scheme with renormalon ambiguity.

Experimental determinations of the top-quark mass (like the one presented in this thesis) are based on fits to mass-sensitive distributions that extract the input mass of the MC event simulation setup (see Chapter 3). It is now argued that the mass implemented in the simulation, m_t^{MC}, corresponds to the one in the MSR scheme, where R is on the order of the scale above which a perturbative parton shower description is used: $m_t^{MC} = m_t^{MSR}(R = 1\text{ GeV}) \pm \mathcal{O}(1\text{ GeV})$ [30]. Future analysis comparing simulation and hadron level QCD predictions are expected to reduce the uncertainty, and will provide a well-defined conversion. Another analysis MC→MSR($R = 3^{+6}_{-2}$ GeV)→$\overline{\text{MS}}$→pole mass conversion [31] arrives at a numerical difference of $0.05^{+0.32}_{-0.62} \pm 0.50$ GeV with uncertainties from the choice of the scale R and unknown higher order corrections. Therefore, the result of the measurement presented in this thesis is expected to be close to m_t^{pole}.

Electroweak fit and vacuum stability

The value of the top-quark mass influences the values of other observables or parameters of the Standard Model. The first indication for its large mass was the

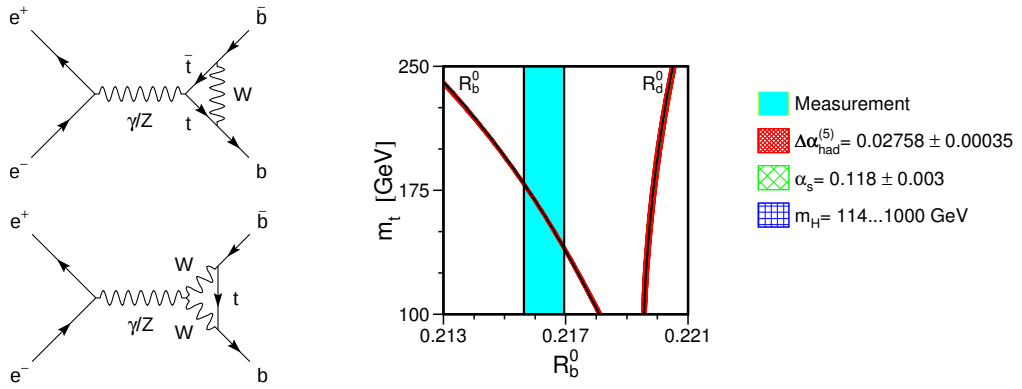

Figure 2.8.: Next-to-leading order vertex corrections containing top quarks in γ/Z boson decays (left), and sensitivity of the branching ratio $\gamma/Z \to b\bar{b}$ on the top-quark mass (right). Plot from [33] .

observation of $B^0 - \bar{B}^0$ mixing in Υ (4S) decays in 1987 [32]. The mixing is mediated via the weak interaction and top-quark loops, and the unexpectedly high rate set a lower bound of $m_t > 50$ GeV. A particularly sensitive observable is the branching ratio of Z bosons into bottom quarks. The branching ratio is changed by the presence of top quarks in vertex corrections at next-to-leading order (NLO) and strongly depends on the top-quark mass, shown in Figure 2.8. Using global fits of electroweak precision observables, the correct range of the top-quark mass was known even before it was measured directly. A historical overview of mass determinations is shown in Figure 2.9.

Quantum corrections from top quarks enter the propagators of the W, Z and Higgs bosons and modify their mass (Figure 2.10). Recently, a new scalar boson was discovered at the LHC [4, 5]. Under the assumption that the discovery corresponds to a SM Higgs boson, all electroweak parameters of the Standard Model have been measured. As the top-Higgs coupling is large, this allows for a more precise indirect determination of the top-quark mass, which is in good agreement with direct measurements, and an improved consistency check of the Standard Model [8], both shown in Figure 2.11. At the current level of precision, there is no sign of failure in the electroweak sector.

After the Higgs discovery, the determination of the top-quark mass gained additional interest as the top quark modifies the Higgs quartic coupling λ such that it could attain negative values at very large energy scales (Figure 2.12, left). This would lead to a second minimum of the Higgs potential, rendering the electroweak vacuum potentially unstable if there is a lower-energy state into that it could de-

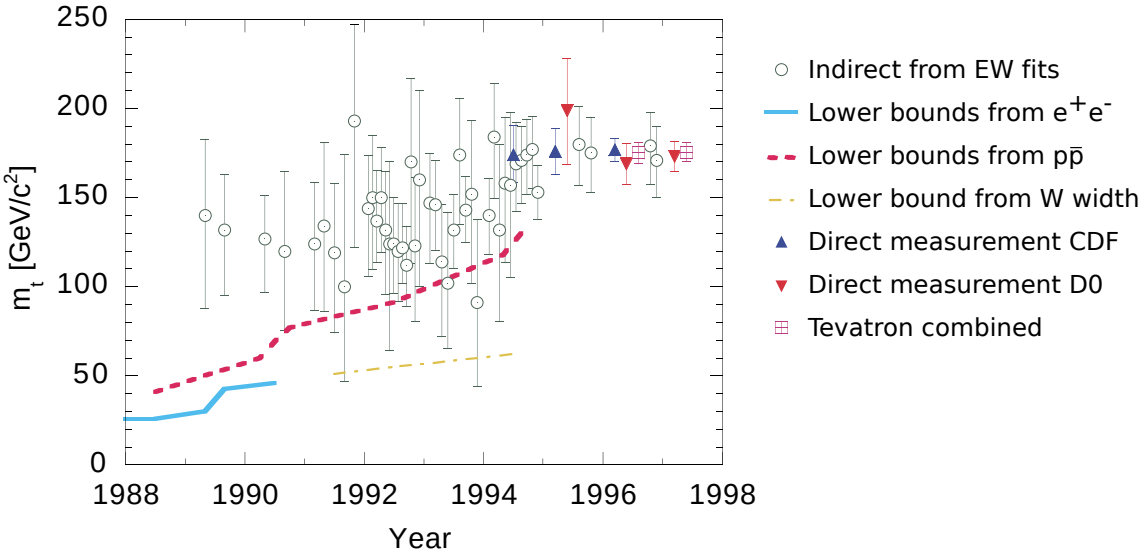

Figure 2.9.: Indirect determinations, lower bounds and direct measurements of the top-quark mass from 1988 to 1998. Plot from [34].

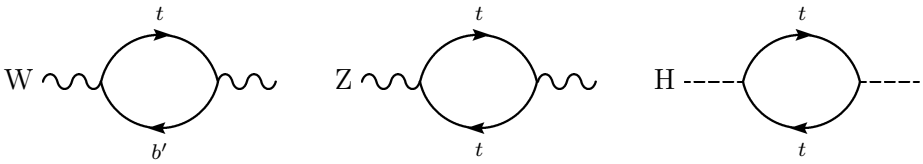

Figure 2.10.: Top-quark loops in the propagators of the W, Z and Higgs bosons.

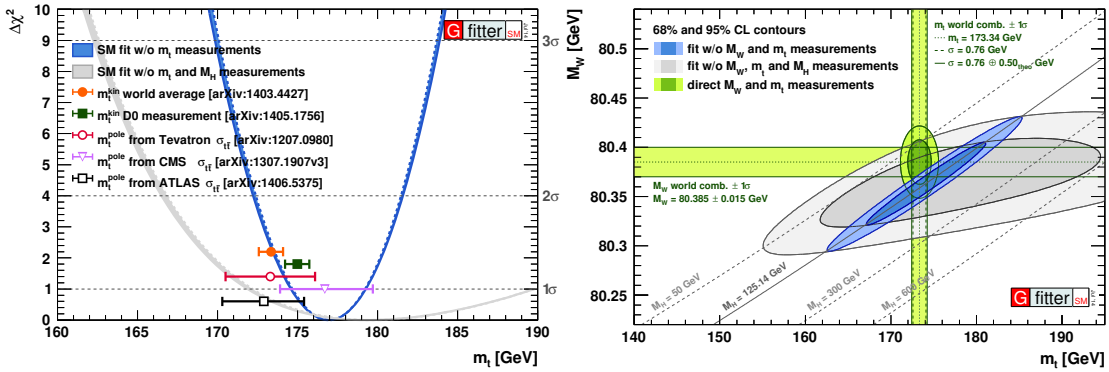

Figure 2.11.: Recent indirect determination of the top-quark mass with and without inclusion of the Higgs result compared to direct measurements (left), and the electroweak fit in the $m_t - m_W$ plane (right). Plots from [8].

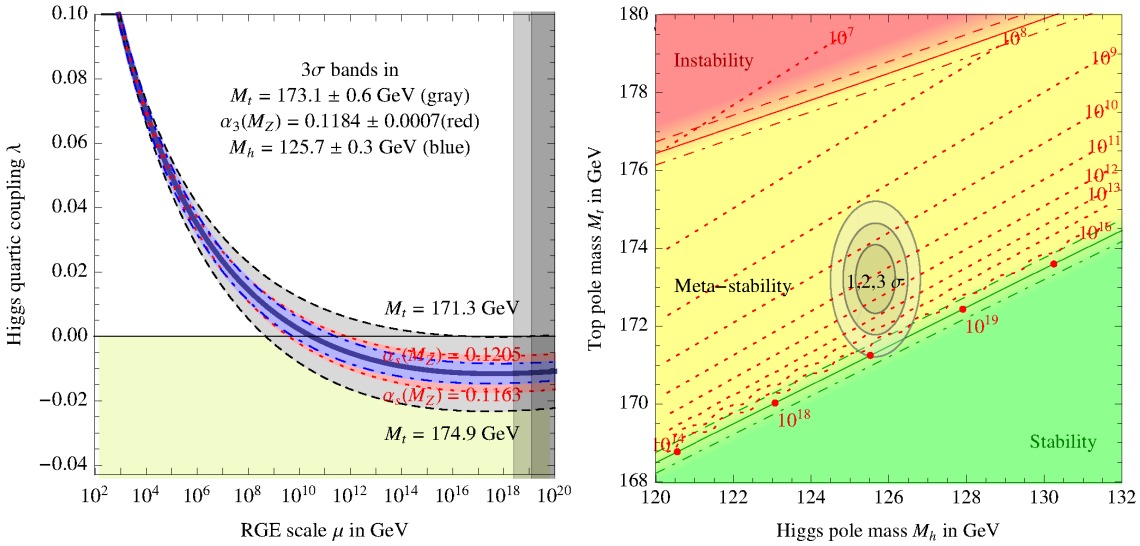

Figure 2.12.: The Higgs quartic coupling as function of the energy scale (left), and the SM phase diagram in the $m_t - m_H$ plane (right). The red-dotted lines show the instability scale in GeV. Plots from [10].

cay [10, 35, 36]. The current measured values of top-quark and Higgs-boson masses suggest an instability scale Λ_I on the order of 10^{11} GeV (Figure 2.12, right) with an uncertainty spanning several orders of magnitude. This magnitude of the instability scale means that our universe is meta-stable, i.e., the half-life of the vacuum exceeds the current age of the universe. Imposing that λ is positive at all scales, Λ_I is the scale where the Standard Model breaks down and a new theory of particle physics is required, containing the SM as a low-energy effective theory. This conclusion would be even more evident if m_t were larger than 178 GeV, where the life-time of the SM vacuum becomes shorter than the current age of the universe.

The peculiar relationship between the masses of the top quark and the Higgs boson with regard to vacuum stability and the hierarchy problem might point to a missing piece in our understanding of particle masses. Therefore, the precise measurement of both masses is an important aspect of the LHC program.

2.3. Top-quark phenomenology

Top-quark production

At hadron colliders, top quarks can be produced either in $t\bar{t}$ pairs or singly. The LO pair production modes are induced by either two gluons or a quark-antiquark pair,

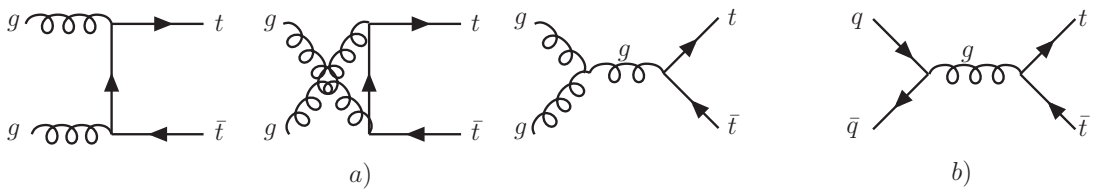

Figure 2.13.: The LO $t\bar{t}$ production processes at hadron colliders:
(a) gluon-gluon fusion, (b) quark-antiquark annihilation

shown in Figure 2.13. The inclusive cross section for $t\bar{t}$ production at a proton-proton collider with the center-of-mass energy $\sqrt{s}$ is given by

$$\sigma^{t\bar{t}}_{h_1 h_2}(s, m_t) = \sum_{i,j} \int_0^1 dx_1 dx_2 f_i(x_1, \mu_F)\, f_j(x_2, \mu_F)\, \hat{\sigma}_{ij},$$

where $\hat{\sigma}_{ij}(\hat{s} = x_1 x_2 s, m_t, \alpha_s(\mu_R), \mu_R, \mu_F)$ is the cross section for the partonic process $ij \to t\bar{t}$ (with $i, j = g, q, \bar{q}$) convolved with the parton distribution functions (PDFs) $f_i(x, \mu_F)$ of the proton. At leading order, the parton distribution functions describe the probability for finding a parton i with momentum fraction x inside the proton, at a factorization scale μ_F. Like the renormalization scale μ_R, the factorization scale is set to a typical scale associated with the process. A suitable choice is $\mu_F = \mu_R = m_t$. The dependence of the cross section on these artificial scales is a remnant of fixed-order perturbation theory, and commonly the theory uncertainty is evaluated by variation of the scales by factors of ½ and 2. The uncertainty decreases with the inclusion of higher order corrections, where the state-of-the-art is next-to-next-to-leading order (NNLO) [37]. At the LHC, being a proton-proton collider, antiquarks can enter the process only as sea quarks, so that 90% of $t\bar{t}$ production are due to $gg \to t\bar{t}$, and 10% originate from $q\bar{q} \to t\bar{t}$. The measurement of the $t\bar{t}$ production cross sections is an integral part of the physics program of the colliders operating at the TeV scale. So far, the rate of $t\bar{t}$ events agrees with the expectations, shown in Figure 2.14 for an assumed top-quark mass of 172.5 GeV. In turn, the cross-section measurement can be used to extract the top-quark mass in the $\overline{\text{MS}}$ or pole scheme with a precision of about 3 GeV [38].

The single-resonant production of top quarks occurs via weak interactions. The LO diagrams are shown in Figure 2.15. The t and s channels have a similar structure with different virtuality of the mediating W boson. In the tW channel, the top quark is produced in association with a W boson, requiring a bottom quark in the initial

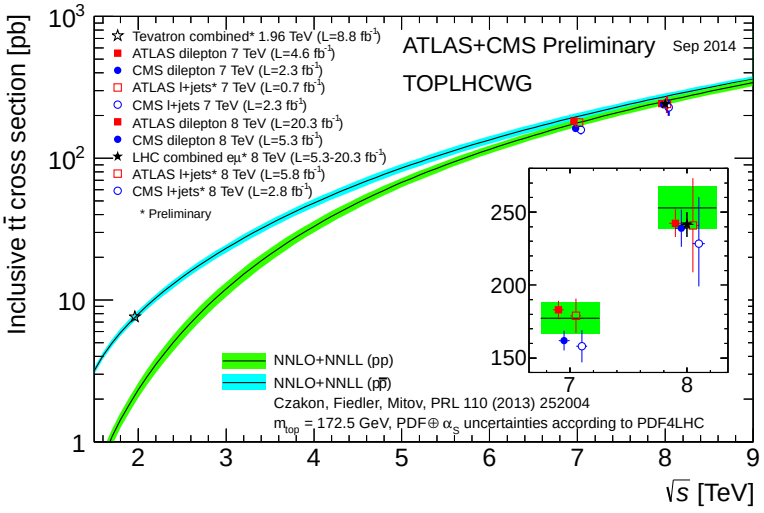

Figure 2.14.: The top quark pair production cross section as a function of the center-of-mass energy $\sqrt{s}$ for pp and $p\bar{p}$ collisions. Plot from Top LHC working group. Plot from [39].

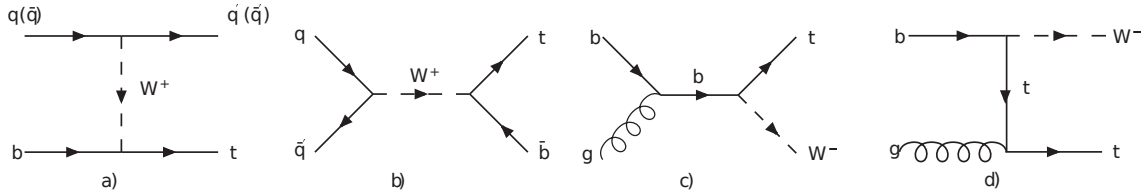

Figure 2.15.: The LO single top production processes at hadron colliders: (a) t channel, (b) s channel, (c,d) tW channel

state. As bottom quarks need to come in $b\bar{b}$ pairs inside the proton, the tW process interferes with $t\bar{t}$ production at NLO. The measured cross sections show no deviation from the SM expectation, shown in Figure 2.16.

Top-quark decays

The large value of the top-quark mass distinguishes it from other quarks and results in a rich phenomenology of the top-quark decays [40].

As the top-quark mass exceeds that of the W boson, it decays weakly via $t \to b'W^+$ with a lifetime of $\tau_t \approx 5 \times 10^{-25}$ s, where b' can be any down-type quark. The lifetime of the top quark is expected to be an order of magnitude smaller than the hadronization time $\tau_{had} \approx 1/\Lambda_{QCD} \approx 3 \times 10^{24}$ s which would be required to form

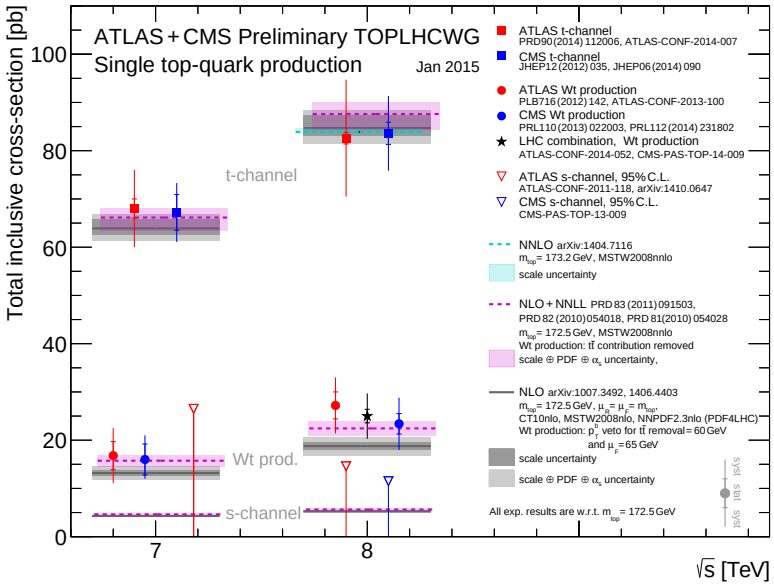

Figure 2.16.: Single-top production cross section at two center-of-mass energies $\sqrt{s}$ for pp collisions at the LHC. Plot from Top LHC working group. Plot from [39].

bound states (top mesons or baryons) with other quarks. This means that the top quark can be studied in the same state as it was produced, like a "bare" quark.

The inverse of the lifetime is the width of the invariant mass distribution of the top-quark decay products $\Gamma_t = 1/\tau_t$. The decay width was determined indirectly from the CKM matrix element V_{tb} and the single-top production cross-section in the t-channel to be $\Gamma_t = 1.36^{+0.14}_{-0.11}$ GeV [41] which is in excellent agreement with the NLO prediction of $\Gamma_t = 1.33$ GeV for $m_t = 172.5$ [42]. A larger value from the experimental determination would be a sign of new top decay modes, possibly into particles beyond the Standard Model. Mass measurements based on the reconstruction of the invariant mass are typically not sensitive to the exact value of the decay width, as the detector resolution is one order of magnitude larger.

The relative probabilities of the decay modes are given by the squares of the corresponding elements of the CKM matrix:

$$\begin{aligned} \mathrm{BR}\left(t \to dW^+\right) &\approx 1 \times 10^{-4}, \\ \mathrm{BR}\left(t \to sW^+\right) &\approx 1.6 \times 10^{-3}, \\ \mathrm{BR}\left(t \to bW^+\right) &\approx 0.998. \end{aligned}$$

Table 2.1.: Decay modes of top-quark pairs, depending on the approximate W-boson branching ratios: dilepton ($1/9$), *lepton+jets* ($4/9$), and **all-jets** ($4/9$) channel. Decay modes containing τ leptons pose additional experimental challenges so that they are rarely used in the analysis of top-quark properties.

BR $(t\bar{t} \to b\bar{b} + X)$	$e^+\nu_e$	$\mu^+\nu_\mu$	$\tau^+\nu_\tau$	$q\bar{q}'$
$e^-\bar{\nu}_e$	1/81	1/81	(1/81)	*6/81*
$\mu^-\bar{\nu}_\mu$	1/81	1/81	(1/81)	*6/81*
$\tau^-\bar{\nu}_\tau$	(1/81)	(1/81)	(1/81)	*(6/81)*
$q\bar{q}'$	*6/81*	*6/81*	*(6/81)*	**36/81**

The subsequent decay of the W boson can occur in the following decay modes:

$$\begin{aligned} \mathrm{BR}\left(W^+ \to \ell^+\nu_\ell\right) &= 0.108,\ \ell = e, \mu, \tau, \\ \mathrm{BR}\left(W^+ \to q\bar{q}'\right) &= 0.337 \times |V_{qq'}|^2, \end{aligned}$$

where the hadronic decays are enhanced with a color factor of 3.

Top-quark pair events can thus be classified into three event categories: "dilepton", "lepton+jets" and "all-jets", as shown in Table 2.1. The dilepton events allow the selection of a clean sample of top quarks after suppression of the Drell-Yan background, achieved by a veto on invariant dilepton masses in a window around the Z-boson mass. The event and mass reconstruction is difficult due to the presence of two hard neutrinos, and the branching ratio is small, especially as usually only electrons and muons are selected in the analysis. The lepton+jets channel (usually electron+jets and muon+jets) features a large branching fraction, a clean selection after b-jet identification, and constrained kinematics, which makes it a good candidate for experimental analysis. The all-jets channel has the largest branching ratio, no "loss" by decays into τ leptons and overconstrained kinematics since no neutrino from W decay is present. The challenge lies in reduction of the non-resonant multijet background. For the measurement presented in this thesis, $t\bar{t}$ candidate events in the lepton+jets decay mode are selected.

As the top quark does not hadronize, its spin information is imprinted on the decay products, predicting angular correlations. These spin correlations can be measured using the angles between the decay products of the two W bosons [43, 44]. A supersymmetric spin-0 partner of the top quark, the "stop" would decay isotropically. Thus, the measurement of top spin correlations can exclude a light

stop with $m_{\tilde{t}} \approx m_t$ [45]. Currently, a stop with $m_t \leq m_{\tilde{t}} \leq 191$ GeV is excluded with 95% confidence level [44].

3. Event simulation

Cross-section integrals need to be carried out over a high-dimensional phase space. For their calculation, random-number based Monte-Carlo (MC) methods are chosen that can converge for arbitrary dimensions statistically with $1/\sqrt{n}$, where n is the number of randomly sampled phase-space points. Each sampled point corresponds to a possible event with a weight proportional to the function value, i.e., the matrix element and a phase-space factor. Using these generated events, MC methods deliver predictions for any observable at a particle collider. These predictions are vital for the estimation of unobservable parameters as the top-quark mass, and the estimation of signal acceptance and selection efficiencies. The precision of the prediction depends on the level of sophistication of the employed methods. The core of event simulation is the matrix-element calculation of the process under interest. For obtaining observable hadrons and possibly passing them on to a detector simulation, parton shower and hadronization models are needed. All these stages are implemented in publicly available codes [46–58]. This summary is based on the contents of [46, 59].

3.1. Parton density functions

The processes under investigation are induced by the interaction between high-energetic (anti-)proton beams. The protons consist not only of the three valence quarks (up, up, down) that define the quantum numbers of the proton, but a large "sea" of gluons, quarks and antiquarks from quantum fluctuations inside the proton contributes as well. In a leading order picture, the parton density function (PDF) can be seen as the probability to find a certain parton with the momentum fraction x. As shown in Figure 3.1, at low momentum transfer $Q^2 = 10$ GeV^2 (corresponding to a distance of the order of 0.1 fm) the valence quarks carry a large portion of the momentum. When the proton is probed with higher momentum transfer, smaller quantum fluctuations can be resolved, leading to an increased contribution of gluons

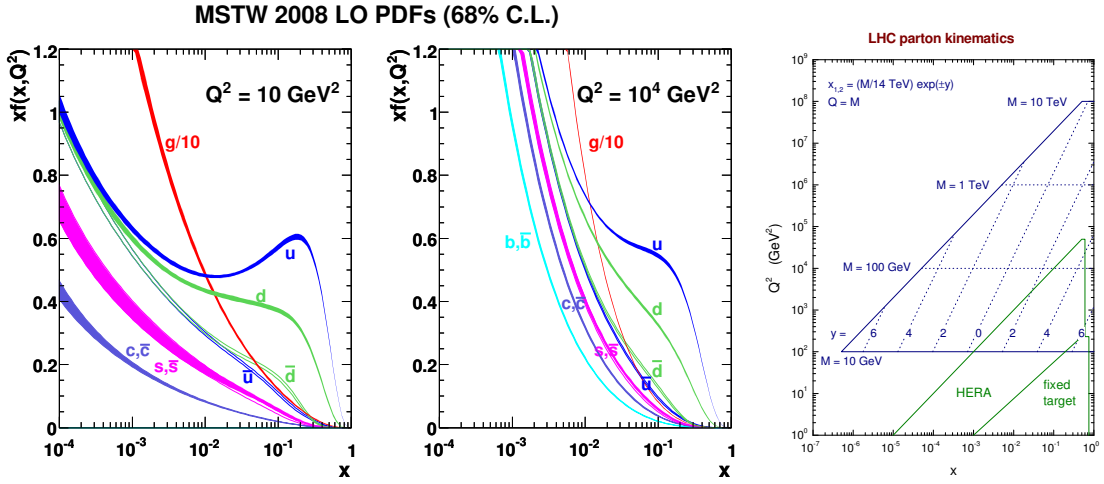

Figure 3.1.: The MSTW2008LO parton distribution function [63] for different Q^2 (left, center). The gluon density is scaled by a factor $1/10$. The right figure shows the x, Q^2 values accessible at fixed-target experiments, HERA and the LHC.

at high energies. The PDFs are usually evaluated at a factorization scale μ_F which is chosen to be of the order of the characteristic scale of the hard process, which could e.g. be the mass of a resonance under investigation. This ensures that the PDFs contain those fluctuations that are relevant for the initial state of the hard process. The DGLAP equations [60–62] are used for extrapolation between measurements at different Q^2, and is based on evolution kernels that contain the matrix elements for gluon radiation and splitting. The DGLAP equations are essential for predicting the PDF at higher Q^2, as needed for the LHC (see Figure 3.1, right), from the data collected at lower Q^2. At the LHC energy of $\sqrt{s} = 8$ TeV, a minimal momentum fraction of $x = \sqrt{x_1 x_2} \geq 2m_t/\sqrt{s} \approx 0.04$ is needed to produce top-quark pairs. There are multiple collaborations that provide PDF sets based on different parametrization and experimental data. Using different PDF sets usually has only a small impact on the kinematics of the final state at the LHC.

3.2. Hard matrix element

The differential cross section for a final state F (like $t\bar{t}$) in perturbation theory is given by

$$\mathrm{d}\sigma_F = \sum_{k=0}^{\text{legs}} \mathrm{d}\Phi_{F+k} \left| \sum_{\ell=0}^{\text{loops}} \mathcal{M}_{F+k}^{(\ell)} \right|^2 ,$$

where Φ_{F+k} denotes a specific momentum configuration, and $\mathcal{M}_{F+k}^{(\ell)}$ is the matrix element for the final state F plus k additional outgoing partons ("legs") and ℓ loops in the corresponding diagram. The matrix element with $k = \ell = 0$ corresponds to the tree level, usually implemented in general-purpose generators like Pythia [56, 57] and Herwig [54, 55].

For $k > 0$, the matrix element contains an accurate description of additional quarks and gluons (or photons) in the final state. With current methods, about $k \approx 5$ additional partons are calculable. As the cross sections suffer from infrared divergences from soft and collinear partons, the calculation requires cuts on angles or transverse momenta. The highest-multiplicity matrix elements are implemented in Alpgen with $k = 6$ additional partons in $t\bar{t}$ production [50]. MadGraph [64] and Sherpa [51] contain methods for the automated calculation of matrix elements for user-defined processes, where in practice $k = 3$ or $k = 4$ is possible for $t\bar{t}$ production.

NLO matrix elements are typically defined as $k + \ell = 1$, containing one loop and one real emission. The KLN theorem [65, 66] states that due to unitarity the IR divergences of the loop amplitudes cancel against those from the soft and collinear real emissions for constant $k + \ell$ as the singularity structure is similar. The NLO cross section can thus be rewritten with process-independent functions (*Catany-Seymour* dipole [67] or *antenna* [68] functions) that contain the singularities and are subtracted from the real emission and added to the loop integral of the calculation.

The newest versions of MadGraph [47] and Sherpa [69] contain (interfaces to) loop amplitude generators [70–72] so that matrix elements with multiple emissions can be evaluated at the 1-loop level. For more than 1 additional leg, the infrared divergences are not fully canceled so that they need to be regulated by cuts.

3.3. Parton shower

The dipole or antenna functions are extracted from the ratio of the squared amplitudes $|\mathcal{M}_{F+1}|^2 / |\mathcal{M}_F|^2$ of any reference process and include the divergent terms

proportional to $1/s_{ij} = 1/(2E_iE_j(1-\cos\theta_{ij}))$ for splitting of an initial parton into nearly on-shell massless partons i and j. For similar color structures, the dipole/antenna functions can be used in a parton shower to approximate the higher-leg matrix elements by recursion. The parton shower is characterized by its evolution variable which can be a dipole mass or the transverse momentum with respect to the dipole (ARIADNE [73], SHERPA [74], VINCIA [75]), angle between initial parton and emission (HERWIG 6/++) or p_{T} with regard to the emitting parton (PYTHIA 6/8). The behavior close to the collinear divergences is similar for all ordering variables but there are differences in the other regions of the emission phase space. The parton shower generates all real emissions until the hadronization scale Q_0, where perturbation theory cannot be applied anymore. Imposing unitarity yields an approximation of the all-order loop corrections as well, so that the integral is finite over the whole phase space. Such parton showers have *leading-logarithm* (LL) accuracy.

The probability of *no* emission of an event between two scales $Q_1 > Q_2$ is given by the *Sudakov factor*

$$\Delta(\Phi_F, Q_1, Q_2) = \exp\left[-\sum_r \int_{Q_2}^{Q_1} \frac{\mathrm{d}\Phi_{F+1}^r}{\mathrm{d}\Phi_F} S_r(\Phi_{F+1})\right],$$

where the radiation functions S_r can either be the DGLAP splitting kernels or dipole/antenna functions:

$$\begin{aligned} \text{DGLAP:} \quad & q \to qg,\ q \to gq,\ g \to gg,\ g \to q\bar{q}, \\ \text{dipole/antenna:} \quad & q\bar{q} \to qg\bar{q},\ qg \to qgg,\ gg \to ggg,\ qg \to qq'\bar{q}',\ gg \to gq\bar{q}. \end{aligned}$$

Technically, the parton showers are implemented as iterative Markov chains. Starting with the scale Q_1 of the hard process, the radiation scale Q_2 is found by solving the Sudakov factor for a random number between 0 and 1, and an emission is generated according to the radiation functions. This process is iterated, starting with Q_2 and evolving down to the cutoff scale Q_0.

For the initial state radiation, the incoming parton momenta are fixed by the hard process, so that *backwards evolution* is necessary to generate the initial-state shower. The Sudakov factor then contains the ratio of the PDFs at the different scales, so that the backwards evolution is suppressed when no matching "mother" partons are available in the PDF.

3.4. Matching matrix element and parton shower

For matching multi-leg LO matrix elements with parton showers, CKKW(-L) [76,77] or MLM [78] matching is employed. These methods ensure that parton shower emissions from $\mathcal{M}_{F+0}$ do not overlap in phase space with the partons contained in the 1-additional-leg matrix element $\mathcal{M}_{F+1}$, and so forth for higher k, effectively removing the double-counting of emissions. Above some matching scale (that should not be more than one order of magnitude lower than the scale of the hard process), the shower is suppressed so that emissions are described by the matrix element. Sudakov factors ensure smooth matching between both phase-space regions and add the LL approximation of the loop corrections. As an additional unweighting is needed, such procedures typically have an efficiency of around 30% with respect to the number of events generated at matrix-element level.

There are two methods used for matching a NLO ($k + \ell = 1$) matrix element with a parton shower. In the MC@NLO method [49, 79], the radiation functions of the parton shower enter as subtraction terms in the 1-leg and 1-loop matrix elements. Applying the same parton shower as used in the ME subtraction then gives the correct distributions with NLO+LL accuracy. This approach yields negative weights for phase-space points where the shower approximation is larger than the exact matrix element, degrading the statistical accuracy of a generated sample by $20 - 60\%$.

In the POWHEG method [80] the matrix element contains the hardest emission and positive virtual corrections so that no negative-weight events occur. This approach yields NLO+LL accuracy given a p_{T}-ordered shower like PYTHIA is used to evolve the event. An angular-ordered shower like HERWIG might generate a harder emission at lower angle, unless a *truncated shower* is implemented for the specific process. Currently, there is no such implementation for top-quark pair production.

Both methods have been generalized to the matching of multi-leg one-loop matrix elements with parton showers [69, 81, 82].

3.5. Hadronization

After evolution of the parton shower down to the cutoff scale in the order of 1 GeV, hadrons are formed via Lund string [83] or cluster fragmentation [84, 85] models.

In the Lund string model implemented in PYTHIA [56, 83], a color-connected quark-antiquark pair $q_0\bar{q}_0$ from the parton shower spans a color string with a tension

$\kappa \approx 1$ GeV/fm. Gluons are treated as "kinks" in the color strings. The potential energy between the color charges is lowered by production of a new pair $q_1\bar{q}_1$ out of the QCD vacuum with a probability proportional to $\exp(-\pi m^2/\kappa)$, implying a suppression of heavier quark flavors. Then z, the fraction of $(E + p_z)$ taken by the hadron $q_0\bar{q}_1$, is distributed according to the fragmentation function $f(z)$, where the z–axis is aligned in direction of the color string. The momentum transverse to the string direction is smeared by a Gaussian distribution with $\sigma \approx 0.3$ GeV. For light quark flavors, the symmetric Lund fragmentation function is given by

$$f(z) \propto \frac{1}{z}(1-z)^a \exp\left(\frac{-bm_\perp^2}{z}\right),$$

where the tunable Lund parameters have default values of $a = 0.3$ and $b = 0.58\,\mathrm{GeV}^{-2}$, and $m_\perp^2 = E^2 - p_z^2$ is the transverse mass of the new hadron. For heavy quark flavors (charm and bottom), the fragmentation function is modified [86] to

$$f(z) \propto \frac{1}{z^{1+r\cdot bm_\perp^2}}(1-z)^a \exp\left(\frac{-bm_\perp^2}{z}\right),$$

introducing an additional parameter r which typically has a value between 0 and 1, and can be set for charm (r_c) and bottom (r_b) fragmentation separately.

The cluster model, with different implementations in HERWIG and SHERPA, starts with forcing $g \to q\bar{q}$ branchings. After that, color singlet *clusters* are formed. If the cluster mass is above some threshold, it decays to two lighter clusters, until all clusters are decayed to a pair of hadrons.

Both models include hadron decay tables that contain the available experimental information [87]. They also need to include predictions of well-motivated unobserved hadron states and the rates of unobserved decay modes.

Hadronization models cannot be derived from first principles of perturbative QCD and typically have a large number of parameters that have been tuned to e.g. e^+e^- data. Similar tunes may lead to different predictions in LHC events [88], where the detector response depends on the momentum distributions and types of stable particles.

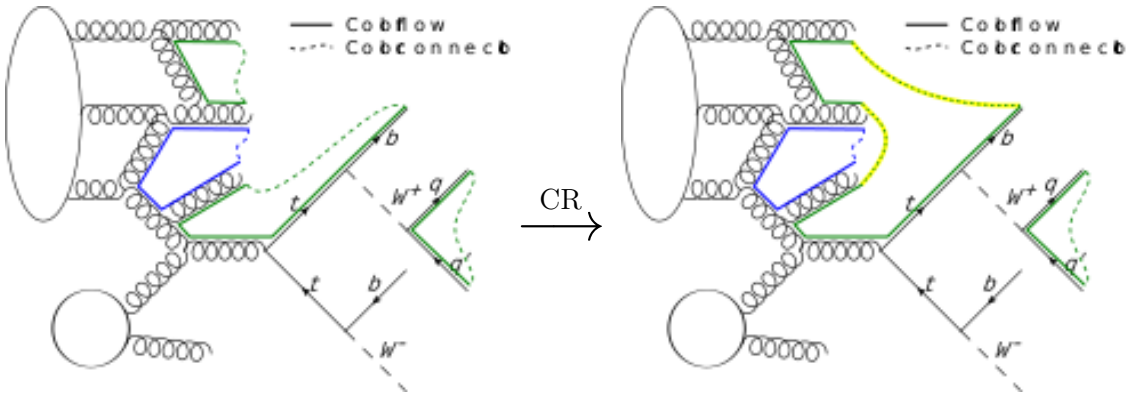

Figure 3.2.: An example for a color reconnection that is likely to alter the kinematics of the bottom jet in a top-quark pair event. For simplicity of the figure, the color connections are attached to large-scale partons. In the real model, their ends are sitting at low-scale branchings further down the parton shower evolution.

3.6. Underlying event

A hard scattering process is accompanied by additional parton interactions of the protons, called the underlying event (UE). The UE is enhanced by the fact that there is a large geometrical overlap between the colliding protons when a hard scattering occurs. The additional interactions are mainly due to the process $gg \to gg$. The available models are tuned to collision data, c.f. [89, 90].

3.7. Color reconnection

Color reconnection models allow non-perturbative changes in the color configuration of the event before the hadronization stage, typically reducing the total potential energy between QCD color charges [91]. This empirical mechanism improves the description of the mean transverse momentum of charged particles vs. their multiplicity in minimum-bias events. Still, the current models are not able to give a consistent tune to data in all p_T ranges. A new set of color reconnection models was recently implemented in Pythia 8 [92]. An example for color reconnection is illustrated in Figure 3.2.

4. Experimental setup

4.1. Large Hadron Collider

The Large Hadron Collider (LHC) [93] at the European Organization for Nuclear Research (CERN) is a storage ring for counter-rotating proton or lead-ion beams with a design energy of 7 TeV per beam. With a circumference of 26.7 km it is the largest collider of the CERN accelerator complex, shown schematically in Figure 4.1. The LHC tunnel lies in a depth of $45 - 170$ m below the earth's surface. It is divided into eight straight sections and eight arcs. One octant is composed out of a straight section and the neighbored half arcs. The four experiments ATLAS [94] (multi-purpose detector), ALICE [95] (optimized for heavy-ion collisions), CMS [96] (multi-purpose), and LHCb [97] (B physics experiment) are located at the centers of the octants 1, 2, 5, and 8, where the beams are brought to collision.

The protons are produced by a duoplasmatron source that ionizes hydrogen gas with an electron beam. The protons have a kinetic energy of about 100 keV after leaving the source. They enter the Linac2 linear accelerator that uses radio-frequency (RF) cavities to increase the proton energy to 50 MeV. The Proton Synchrotron Booster (PSB) then accelerates the protons to 1.4 GeV, from where they are injected into the Proton Synchrotron (PS). The PS accelerates the protons to 25 GeV and splits the beam into smaller bunches of 4 ns length with a spacing of 25 or 50 ns as required for LHC operation. The Super Proton Synchrotron (SPS) increases the proton beam energy to 450 GeV before the beams are finally injected into the LHC for further acceleration, storage and collisions.

The number of events per second for a certain process is given by $\mathrm{d}N/\mathrm{d}t = \sigma \times L$, where σ is the cross section of the process of interest, and L denotes the *instantaneous luminosity*. The LHC was operated at $\sqrt{s} = 7$ TeV in the years 2010 and 2011, and at $\sqrt{s} = 8$ TeV in 2012, yielding large cross sections for processes involving heavy particles like the top quark. For protons in this energy regime, synchrotron radiation is still negligible, so that the maximum energy is limited by the bending power of

the dipole magnets. The 1232 super-conducting LHC dipole magnets were designed for a maximum field strength of 8.33 T, allowing a maximum center-of-mass energy of $\sqrt{s} = 14$ TeV. The eight RF cavities for each beam are operated at 400.8 MHz, providing an energy gain of 485 keV per turn and proton, and bunch lengths of about 10 cm.

The instantaneous luminosity L is given by

$$L = \frac{\gamma f_{rev} k_b n_1 n_2}{4\pi\epsilon^*\beta^*} F\left(\theta_c, \sigma_z, \sigma^*\right).$$

The enumerator contains the relativistic γ factor, the revolution frequency f_{rev}, the number of bunches k_b, and n_i which is the number of particles per bunch of the colliding beams. The denominator specifies the beam profile at the interaction point (IP), commonly given as the product of emittance ϵ^* and the beta-function β^*. The emittance is a measure of the beam phase space. A low ϵ^* means that the protons in a bunch are concentrated in a small area and are nearly collinear. The beta-function at the interaction point is the distance where the beam is twice as large compared to the IP, so a low β^* corresponds to a good focusing of the beam. F is a geometric reduction factor due to the crossing angle θ_c between the beams, and also depends on the bunch profile in longitudinal (σ_z) and transverse (σ^*) direction. A crossing angle chosen too small would lead to unwanted beam-beam interactions before the nominal IP.

In proton-proton mode, the LHC is designed for a peak instantaneous luminosity of $L = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. During the operation in 2012, a peak luminosity of $7.7 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ was achieved with 1368 bunches at 50 ns spacing, each containing 1.6×10^{11} protons, an emittance of approximately 2.4 µm, $\beta^* = 0.6$ m, and $F \approx 0.8$. The evolution of the daily peak luminosity delivered to the CMS experiment is shown in Figure 4.2. An integrated luminosity $\mathcal{L}$ corresponding to $6.1\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 7\,\mathrm{TeV}$ and $23.3\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 8\,\mathrm{TeV}$ was delivered to the CMS experiment.

4.2. CMS detector

The Compact Muon Solenoid (CMS) experiment [96] at the LHC was built as a multi-purpose particle detector to study proton-proton, proton-lead, or lead-lead collisions at the TeV scale. It has large dimensions of 22 m in length and 15 m in diameter, and a weight of 14 000 metric tons. The CMS detector has an "onion"

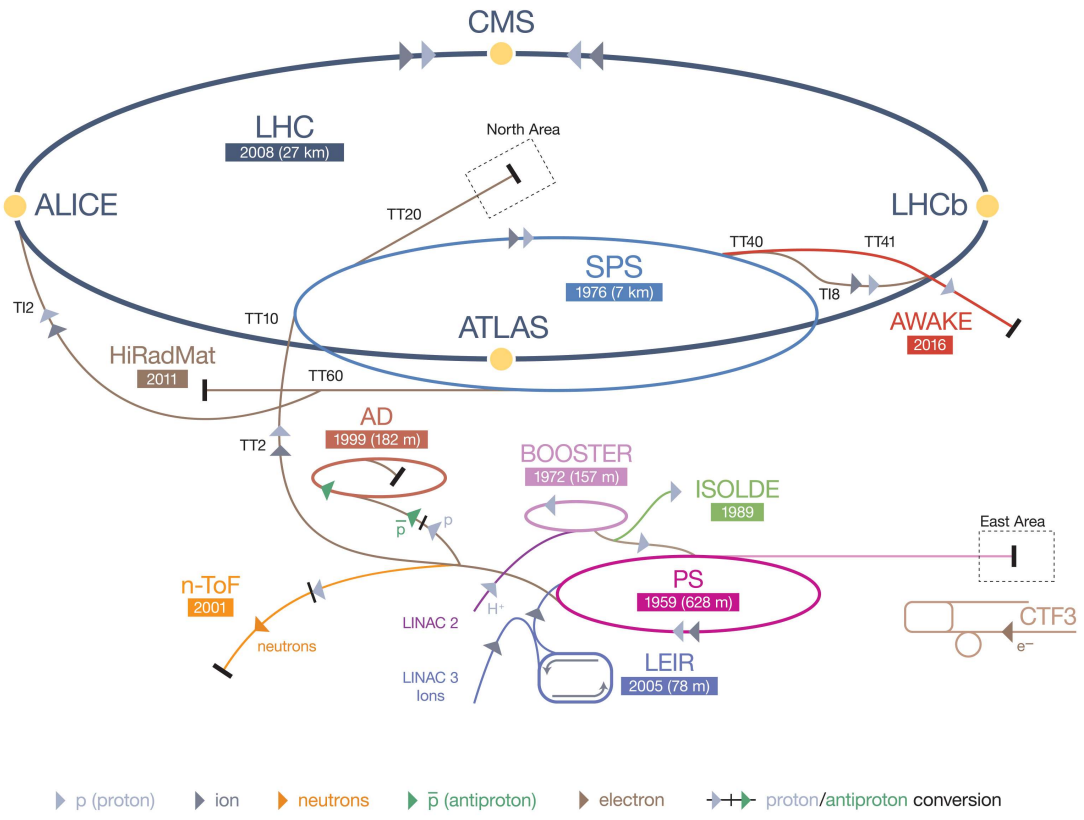

Figure 4.1.: CERN accelerator complex. Source: [98].

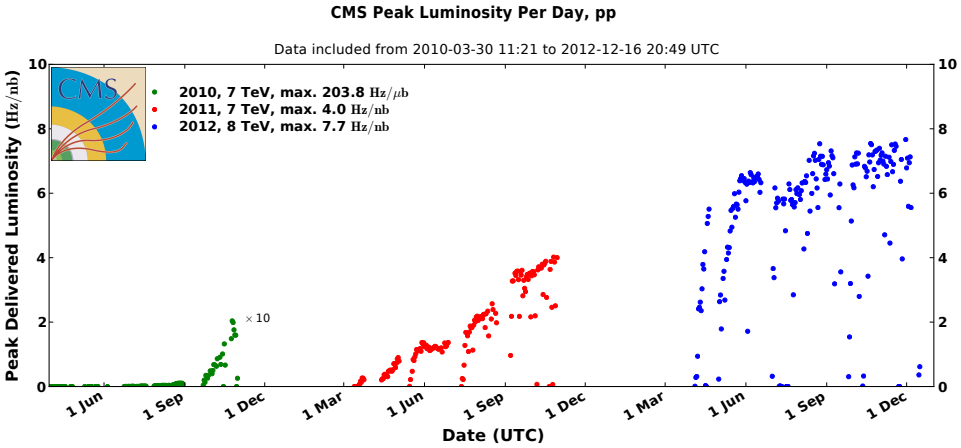

Figure 4.2.: Peak luminosity delivered to CMS in proton-proton collisions. Source: CMS collaboration [99].

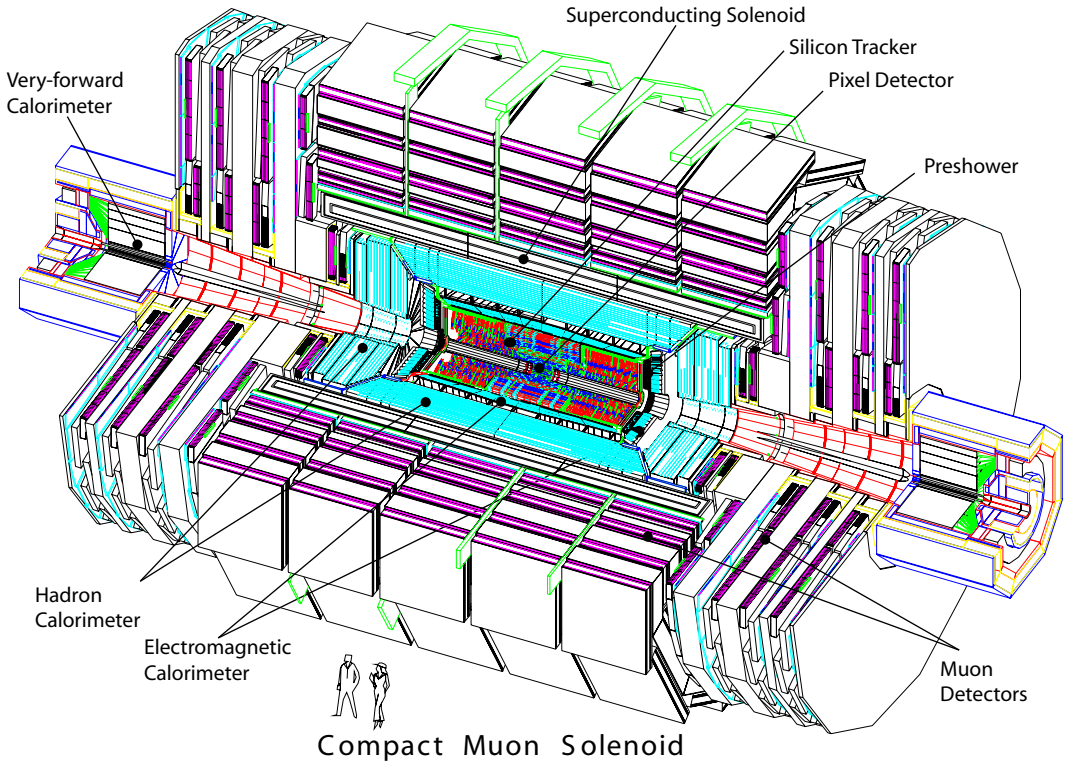

Figure 4.3.: An overview of the CMS detector. Source: CMS collaboration [100].

structure, where each layer is specialized on detecting different types of particles. It covers nearly the full solid angle of 4π around the nominal interaction point.

The central feature of the CMS apparatus is a superconducting solenoid. Contained within the field volume are a silicon pixel and strip tracker, a crystal electromagnetic calorimeter (ECAL) and a brass/scintillator hadron calorimeter (HCAL). Muons are measured in gas-ionization detectors embedded in the steel return yoke. An overview is given in Figure 4.3.

Coordinate system

CMS uses a right-handed coordinate system, with the origin at the nominal interaction point, the x axis pointing to the center of the LHC ring, the y axis pointing up (perpendicular to the plane of the LHC ring), and the z axis along the counterclockwise-beam direction. The polar angle, θ, is measured from the positive z axis and the azimuthal angle, ϕ, is measured in the x-y plane. The pseudorapidity $\eta = -\ln\tan(\theta/2)$ is regularly used instead of θ.

Solenoid magnet

The name-giving solenoid magnet is 12.9 m long and has an internal diameter of 5.9 m. It provides a magnetic field of 3.8 T and thus bending power for the precise measurement of charged particle momenta in the tracking system. It is built of 50 km of superconducting cable with a cross section of 64×22 mm^2 in 2168 windings. The magnet is operated with a current of 19.5 kA, storing a total energy of 2.7 GJ. The magnetic return flux in the iron yoke has a field strength of 2 T, bending the path of muons into the opposite direction.

Tracker

The CMS tracker consists of silicon pixel and silicon strip detector modules, covering the pseudorapidity range $|\eta| < 2.5$. It is designed to provide high momentum resolution and reconstruction efficiency for charged tracks, as well as the identification of primary and secondary vertices. A schematic overview of the tracker system is shown in Figure 4.4, and its performance at $\sqrt{s} = 7\,\text{TeV}$ is described in [101].

The pixel tracker is located closest to the beam pipe, consisting of three barrel layers at radii of 4.4 cm, 7.3 cm and 10.2 cm, and four endcap disks located at distances of ± 34.5 cm and ± 46.5 cm from the nominal interaction point. It is organized in 1440 modules with in total 66 million hybrid pixel sensors. The pixel size is 100×150 µm, providing a resolution of approximately 10 µm in transverse and $20 - 40$ µm in longitudinal track impact parameters. After alignment using cosmic muon data, the position of each pixel is precisely known. This results in a precise three-dimensional determination of the hit coordinates.

The silicon strip tracker built around the pixel detector is 5.8 m long and has an outer diameter of 1.1 m. It comprises 9.6 million silicon microstrip channels that are ordered in four layers of the inner and six layers of the outer barrel tracker (TIB/TOB) parallel to the beam line, as well as in three inner disks (TID) and nine endcap disks (TEC) perpendicular to the beam. The strip detectors have a resolution of $23 - 52$ µm perpendicular to the strip direction, giving two-dimensional hit information. The first two TIB and TOB layers, the first two TID and TEC rings, and the fifth TEC ring are equipped with additional strips that are slightly tilted (100 mrad) and provide "stereo" position information in strip direction with a resolution of $230 - 530$ µm.

The material budget of the inner tracking system needs to be minimized so that track directions are not altered by multiple scattering. The depth reaches from

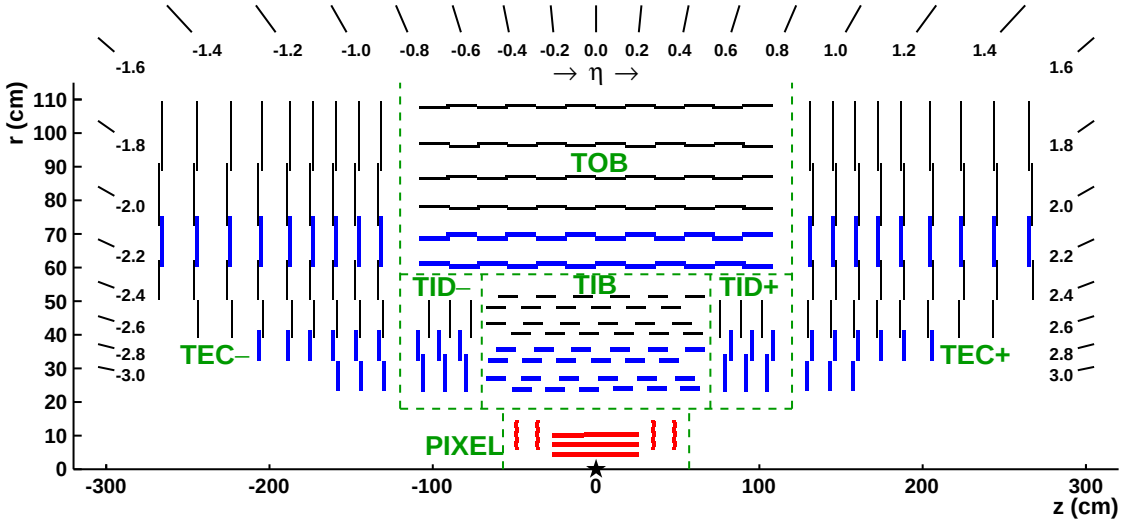

Figure 4.4.: The CMS inner tracker system. The pixel tracker is marked in red, stereo modules in blue. Source: CMS collaboration [101].

0.4 radiation lengths in the barrel to two radiation lengths in the transition region between barrel and endcaps.

Calorimeters

The electromagnetic calorimeter (ECAL) is designed to provide excellent photon and electron energy resolution. It is composed by 61 200 scintillating lead tungstate ($PbWO_4$) crystals in the barrel region ($|\eta| < 1.5$) and 7 324 crystals in each endcap, covering $1.5 < |\eta| < 3.0$. The ECAL in the barrel (EB) has an internal diameter of 1.29 m. The barrel crystals have a length of 23 cm and are directed towards the nominal interaction point with a tilt of 3°. They have a front area of 22×22 mm^2 and thus cover 0.0174 in $\Delta\phi$ and in $\Delta\eta$ for $\eta \approx 0$. The endcap calorimeters (EE) are located 3.14 m away from the IP at each side of the detector, where the $28.6 \times 28.6 \times 220$ mm^3 crystals are arranged on an $x - y$ grid in a circular form.

The lead tungstate crystals have a short radiation length of $X_0 = 0.89$ cm, so that the length of the barrel crystals corresponds to 25.8 radiation lengths. Their scintillation light is detected by silicon avalanche photodiodes (APDs) in the barrel, and vacuum phototriodes (VPTs) in the endcaps. Measurements of the ECAL performance at $\sqrt{s} = 7$ TeV can be found in [102]. The relative energy resolution

measured from an electron test beam is

$$\frac{\sigma}{E} = \underbrace{\frac{2.8\%}{\sqrt{E\,(\mathrm{GeV})}}}_{\text{stochastic}} \oplus \underbrace{\frac{12\%}{E\,(\mathrm{GeV})}}_{\text{noise}} \oplus \underbrace{0.3\%}_{\text{constant}} ,$$

yielding 0.4% for a 100 GeV electron. During irradiation, the transparency of the ECAL crystals decreases, resulting in a lower energy response. The crystals recover during low-luminosity runs and technical stops. Their transparency is continuously monitored using laser and LED light and the response is corrected accordingly.

Preshower detectors (ES) based on lead absorbers and silicon strip detectors are installed in front of the endcaps to distinguish between prompt photons and π^0 mesons.

The hadronic calorimeter (HCAL) is built to provide good resolution for dijet events and the determination of the missing transverse momentum. As it is contained within the solenoid, maximization of absorber material is crucial to catch the complete hadronic showers. Therefore, the HCAL consists of alternating layers of brass as absorber (thickness of 5 cm) and plastic scintillator as active material (0.37 cm). The scintillation light is converted by wavelength-shifting fibres and channeled to hybrid photodiodes (HPDs) via clear fibres.

The barrel HCAL (HB) contains 15 brass plates and covers a pseudorapidity region of $|\eta| < 1.4$. It is divided into 2304 segments ("towers"), each covering 0.087 in $\Delta\phi$ and $\Delta\eta$. Within $|\eta| < 1.26$, additional scintillator layers outside the magnet coil, called the hadron outer calorimeter (HO), increase the thickness of the calorimetry system to about 10 hadronic interaction lengths.

The endcap hadronic calorimeters (HE) cover the region $1.3 < |\eta| < 3.0$ and are composed out of 2304 towers. Depending on the tower positions, the segmentation in ϕ is 5° (small $|\eta|$) or 10° (large $|\eta|$), respectively, and $0.087-0.35$ in η, with coarser segmentation at high $|\eta|$.

Forward HCALs (HF) are installed at $3.0 < |\eta| < 5.0$, in a distance of 11.2 m from the interaction point. Steel is used as an absorber and has a total depth of 1.65 m. Quartz fibres are inserted as active material. The particles emit Cerenkov light that is detected by HPDs. In total, the two HF modules contain 900 towers, having a segmentation in η of $0.1 - 0.3$ and $10° - 20°$ in ϕ. Each HF tower has two readout channels at different depths (8.2 and 9.5 hadronic interaction lengths) allowing for a separation of the electromagnetic and hadronic shower components.

Muon system

The muon system is designed to provide a high muon identification efficiency and charge determination up to momenta of 1 TeV. As shown in Figure 4.5 (left), it consists of four cylindrical muon barrel (MB) stations, and four muon endcap (ME) disks at each side of the detector.

The MB stations are located at distances of 4.0, 4.9, 5.9, and 7.0 m from the beam axis. They contain 250 aluminum drift tube chambers (DTs) covering the pseudorapidity range $|\eta| < 1.2$. Each DT station provides a resolution of 100 µm in position and 1 mrad in direction. A DT is accompanied by $1-2$ resistive plate chambers (RPCs) operated in avalanche mode. The RPCs have a coarse position resolution but provide fast response and a good time resolution to assign a muon to the correct bunch crossing.

The RPCs are also used in the endcaps so that they cover a region of $|\eta| < 1.6$, with the possibility of an extension to $|\eta| < 2.1$ in the future. The spatial resolution in the endcaps is provided by 486 cathode strip chambers (CSCs) that span a region of $0.9 < |\eta| < 2.5$. The CSCs are optimized for the large neutron background and high muon rate in the forward region, and have a spatial resolution of $100-200$ µm and an angular resolution of 10 mrad.

In total, the muon system covers an active area of 25 000 m^2 and has 1 million readout channels. The time resolution for all muon subdetector systems is around 3 ns, providing excellent trigger capabilities. In combination with the inner tracker, the muon system improves the momentum resolution in the barrel for muons with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$, as shown in Figure 4.5 (right). A detailed study of the muon performance in collisions at $\sqrt{s} = 7\,\mathrm{TeV}$ is available in [103].

Trigger

At LHC design luminosity, 1 billion interactions are expected per second, with bunch crossings every 25 or 50 ns. As only 1000 events per second can be written to permanent data storage, a two-tier trigger system is employed to select the most interesting collision events for use in physics analyses.

The Level-1 (L1) trigger is built of custom hardware processors using information from the calorimeters and the muon system. Using reduced granularity and resolution, the L1 trigger scans the events for high-energetic photons, electrons, muons and jets, and calculates the total (missing) transverse energy of an event. The L1 decision is faster than 1 µs. For a total time of 3.2 µs, the whole detector data is

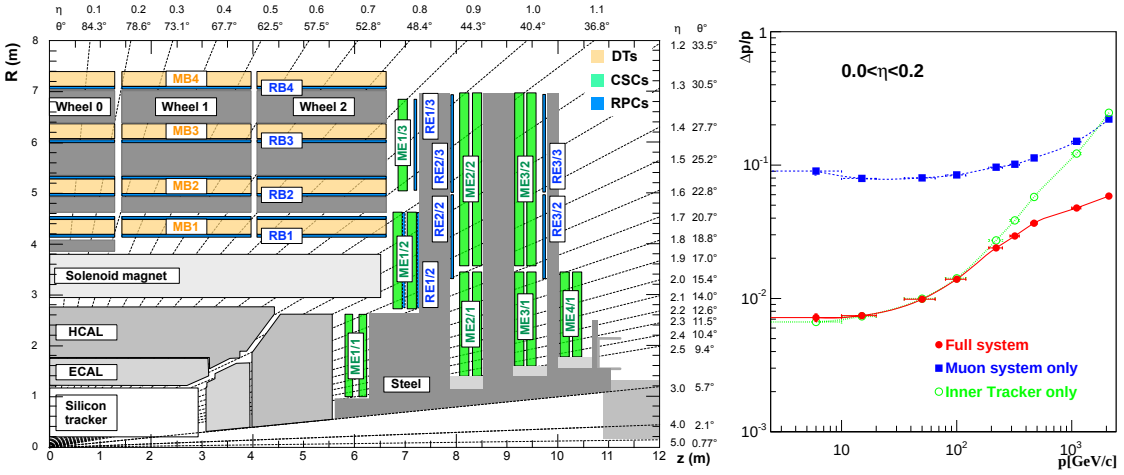

Figure 4.5.: The CMS muon system (left) and momentum resolution for muons (right). Source: CMS collaboration [100, 103].

stored in a buffer, before it is discarded or passed on to further processing. The L1 trigger is designed for reducing the data rate to $16 - 100$ kHz.

The high-level trigger (HLT) reduces the data rate further using a farm of standard processors. The HLT software reconstructs events using data with higher granularity than the L1 trigger and algorithms similar to those used in the offline reconstruction. The total rate of events passing the HLT in the 2012 run was 920 Hz, where prescales were applied to keep the rate approximately constant as the instantaneous luminosity increased during the run. Data corresponding to a rate of 550 Hz were fully reconstructed and available for analysis within 48 hours. The data from the remaining rate of 370 Hz was "parked" for reconstruction after data taking [104, 105].

5. Event reconstruction

In order to analyze the data, physics objects are reconstructed from the track candidates and calorimeter clusters recorded by the detector using the particle-flow algorithm. Lepton candidates and jets are required to be compatible with a common primary interaction vertex.

5.1. Primary vertex identification

Given the high instantaneous luminosity in the 2012 data-taking, on average 21 interactions are expected per bunch crossing. The single interaction leading to the triggered event needs to be identified, so that particles stemming from the additional pileup events can be excluded from the event interpretation.

The interaction vertices are reconstructed in CMS as follows [101]: Charged particle tracks are selected that stem from the interaction region, having an impact parameter significance below 5σ. The impact parameter is the distance between the track and the beam spot at closest approach. Each track needs to be fitted from at least five hits in the inner tracking system, with at least two hits in the pixel tracker. The normalized χ^2 of the track fit is required to be below 20.

The *adaptive vertex fitter* [101, 106] is used to cluster the tracks to vertex candidates. The algorithm tries to find as many vertices as possible without splitting true vertices. It assigns a weight (or probability) w between 0 and 1 to each connection between a track and a vertex candidate, based on their positions on the z-axis and the corresponding uncertainties. A minimal weight of 0.5 is required for each connection, and at least two tracks of a candidate vertex must be incompatible ($w < 0.5$) with other vertices. As figure of merit for a fitted vertex, the number of degrees of freedom is defined as

$$n_{\mathrm{dof}} = -3 + 2 \sum_{i=1}^{\mathrm{tracks}} w_i,$$

so that a large number of compatible tracks results in a high value.

Using CMS data at $\sqrt{s} = 7\,\text{TeV}$, the primary-vertex resolution was found to be below 50 µm in the x/y– and z–directions for vertices with more than ten associated tracks [101].

The mean number of fitted primary vertex candidates per event in 2012 data is 14.55, corresponding to a vertex identification efficiency of about 70%, including the loss from vertices with neutral particles only. For each event, the vertex with the highest $\sum p_{\text{T}}^2$ of the associated tracks is regarded as the primary interaction vertex. It is required to be within 24 cm in longitudinal and 2 cm in transverse direction from the nominal interaction point, and to have $n_{\text{dof}} > 4$.

5.2. Particle flow reconstruction

Events are reconstructed using the particle-flow (PF) algorithm [107,108] that combines the information from all CMS sub-detectors to identify and reconstruct individual objects. As illustrated in Figure 5.1, muons are traversing the whole detector, leaving signals in the inner tracker and the muon system. Electrons leave a track and an energy deposit in the ECAL, while photons are detected only by the ECAL. The momentum of charged hadrons can be accurately measured in the tracker, and their energy is deposited mainly in the calorimeters. The energy of neutral hadrons is measured in the calorimeters only.

Muons are easily identified and their tracks are not considered as candidates for charged hadrons or electrons. The PF algorithm then connects the remaining tracks with the calorimeter clusters. If the energy of connected calorimeter clusters is significantly larger than the corresponding track energies the excess is attributed to an overlapping neutral hadron in the HCAL and to a photon in the ECAL. Calorimeter deposits without an associated track are also identified with neutral particles. The energy deposits of hadrons in the calorimeters are corrected for a different response in the ECAL (compared to electrons and photons) and the non-linear response of the HCAL.

Jets typically consist out of 65% charged hadrons, 25% photons and 10% neutral hadrons. Using the list of particle-flow candidates for clustering (cf. Section 5.4) improves the measurement of jet energy and directions significantly with regard to calorimeter jets, as shown in Figure 5.2. The PF jets are closer to their "true" energy due to the energy calibration of the individual identified particles. The energy and

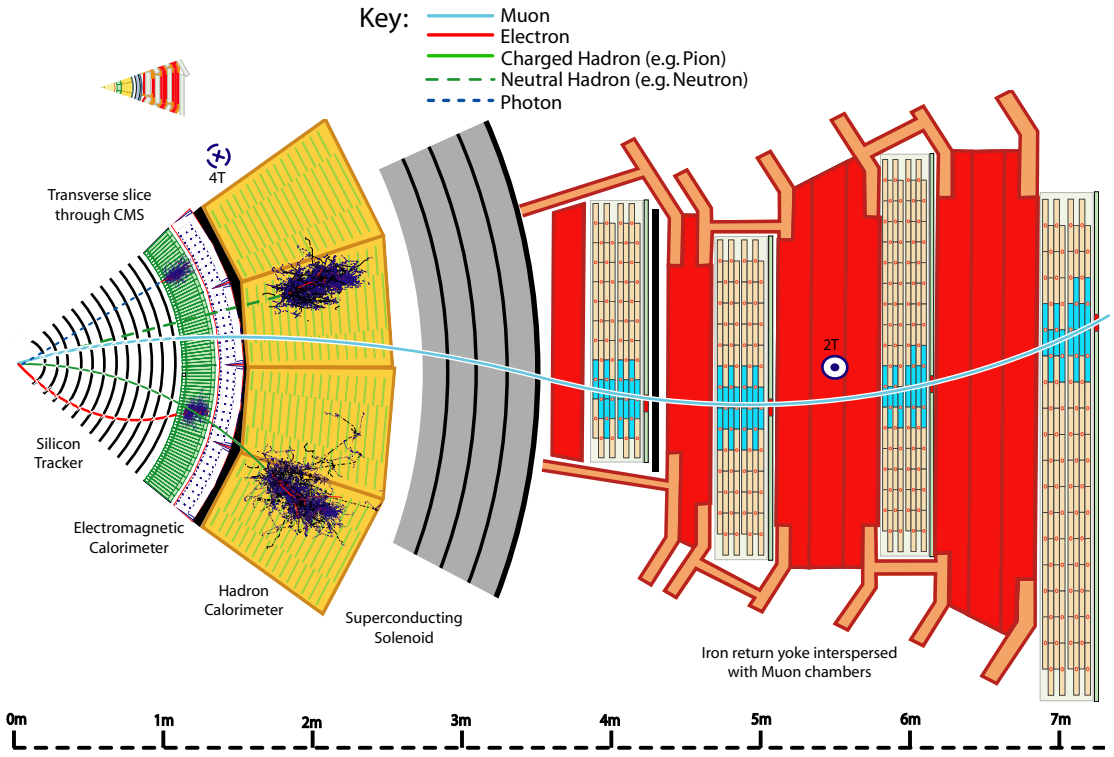

Figure 5.1.: Slice through the CMS detector, illustrating the signatures of different types of particles. Source: CMS collaboration [109].

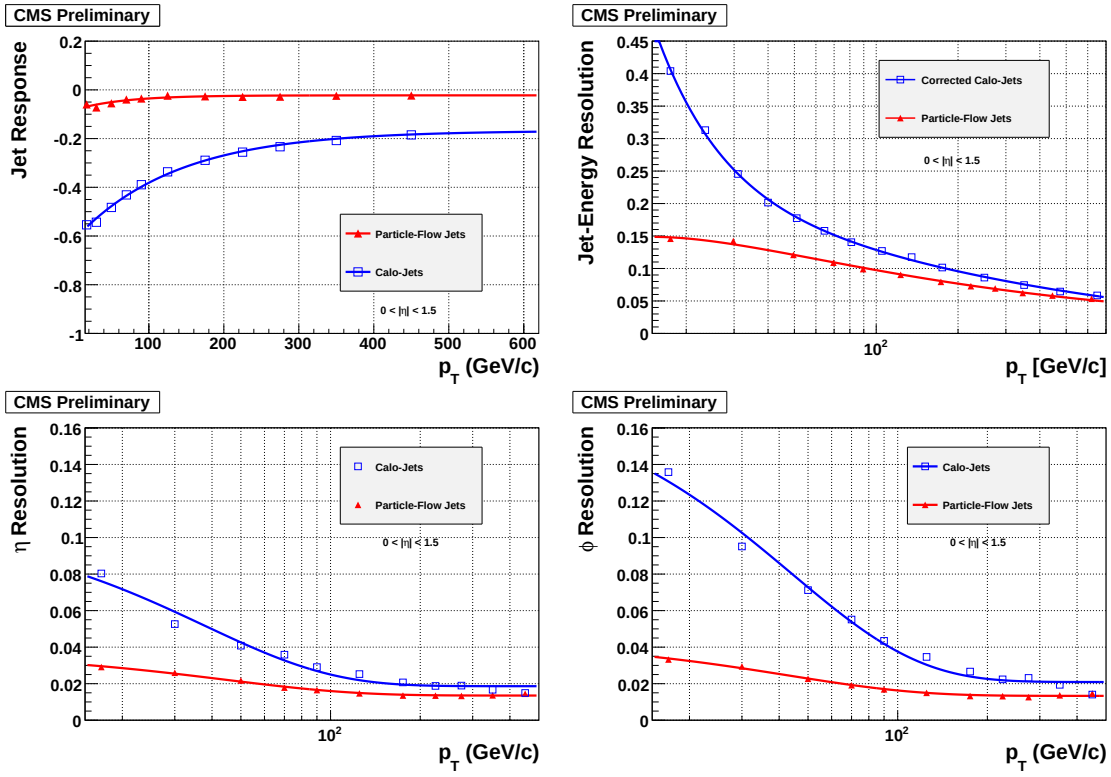

Figure 5.2.: Jet response (upper left), energy resolution (upper right), η resolution (lower left), and ϕ resolution (lower right) for jets using calorimeter clusters only or particle flow candidates as input. No additional pp interactions are included in this preliminary study. From [107].

angular resolution is improved over the whole p_{T} range, demonstrating that possible misassignments of tracks and calorimeter clusters are well under control.

5.3. Lepton reconstruction

In this analysis, exactly one high-energetic lepton is required for the event selection. Due to the special challenges of τ-lepton reconstruction, only muons and electrons are considered as leptons here.

Muons

The performance of the CMS muon reconstruction is described in [103, 110]. Muons are reconstructed in both tracker and muon system (so-called *global* muons) with valid hits in at least 5 tracker layers, one pixel layer, and one of the muon stations. A χ^2/ndf smaller than 10 is required for the global track fit. In order to assure that

the muon is coming from the hard interaction, its impact parameter with respect to the beam spot is required to be $|dB| < 0.02$ cm, and the distance to the primary vertex in z–direction needs to be smaller than 0.5 cm.

The particle-flow isolation I within a radius R is defined as

$$I\left(R\right) = \left(\sum_{ch(R)} p_T + \max \left\{ \sum_{nh(R)} p_T + \sum_{\gamma(R)} p_T - \frac{1}{2} \sum_{chPU(R)} p_T,\, 0 \right\} \right) / p_{\mathrm{T}}^{\ell},$$

where $ch\left(R\right)$ denotes all charged hadrons in a radius R around the lepton direction, nh stands for the neutral hadrons, γ for the photons, and $chPU$ for the charged hadrons stemming from other vertices. This expression reduces the influence of pileup events by assuming that the contribution from neutral pileup particles is approximately half of that from charged pileup. Muons are selected if $I < 0.12$ in a cone of $\Delta R < 0.4$.

The muon momentum scale is calibrated from the J/Ψ and Z boson resonance masses with a precision of 0.2%. The relative muon momentum resolution is in the range of $1.5\% \lesssim {}^{\sigma(p_{\mathrm{T}})}/_{p_{\mathrm{T}}} \lesssim 5\%$ and has been validated in data [103, 110].

Electrons

Electrons are reconstructed from a charged track matched to energy deposited in the ECAL [111]. Due to the tracker material, electrons radiate a significant amount of their energy by bremsstrahlung even before reaching the calorimeters. Therefore, the electron energy is not fully contained in a 5×5 crystal cluster around the largest energy deposit used as seed. In the barrel, the bremsstrahlung yields additional energy deposits in ϕ direction, so a *supercluster* is built by extending an initial window of 5×1 crystals in $\eta \times \phi$ into the ϕ direction. The extension is stopped when a rise in energy deposit is encountered in newly added crystals, or 17 steps (corresponding to 0.3 rad) are reached.

In the endcaps, the supercluster is built from 5×5 clusters around the seed crystal in a range of ± 0.07 in η and ± 0.3 rad in ϕ. The energy deposits from the preshower are added in a range of up to ± 0.45 rad in ϕ and a range of ± 0.15 in pseudorapidity.

The reconstruction of electron tracks can be done with the standard Kalman filter (KF) approach. However, the energy losses due to bremsstrahlung lead to a change in curvature as the electron travels into the outer tracker regions, and the KF track

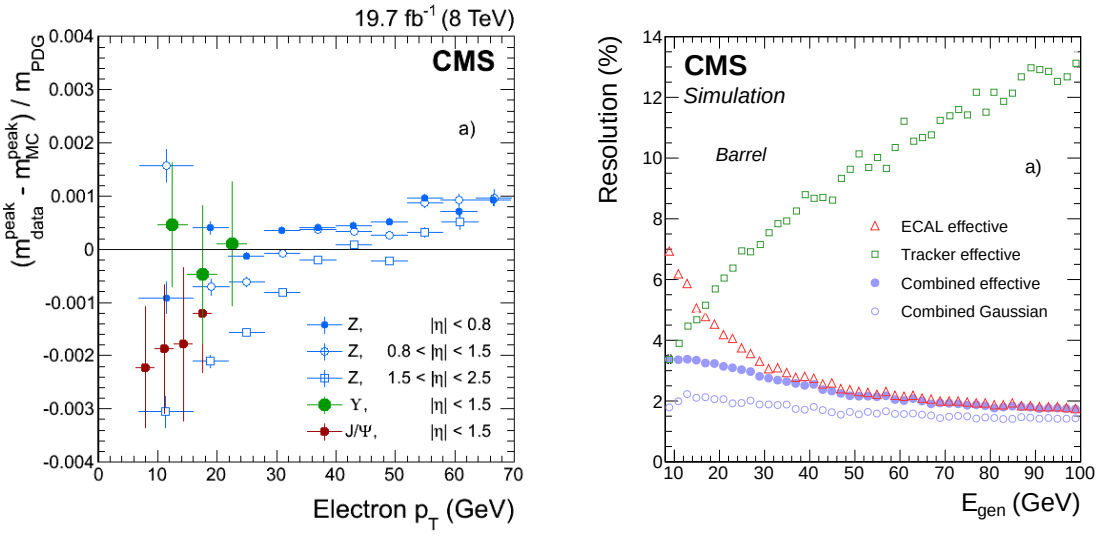

Figure 5.3.: Electron performance for the CMS experiment: relative difference in reconstructed resonance mass as a function of the electron p_{T} (left) and relative resolution as a function of the electron energy (right). Source: CMS collaboration [111].

cannot be fitted very well. The Gaussian sum filter (GSF) is able to take into account the change in curvature and is used to improve poorly fitted KF tracks.

Electrons are then identified by a multi-variate analysis [112] that uses information on the quality of the track-to-cluster matching, the shower shape, additional HCAL deposits, and the difference between the KF- and GSF-fitted track.

Electrons from photon conversions are rejected by looking for missing hits in the inner tracker layers and fitting two electron candidates to a common vertex compatible with a photon. The electrons are excluded from analysis if the pseudorapidity of their supercluster is within the transition region $1.4442 < |\eta| < 1.5660$ between the barrel and endcap calorimeters, where the electron identification is very inefficient.

As for the selected muons, the electron impact parameter with regard to the beam spot needs to be $|dB| < 0.02$ cm, with a distance to the primary vertex in z–direction less than 0.5 cm. A particle-flow isolation $I < 0.1$ within a cone of $\Delta R < 0.3$ is required.

From the decays of the $\Upsilon\,(1S)$, J/Ψ and Z resonances to e^+e^- pairs, the electron energy scale in simulation was found to agree with data within 0.5% in the barrel and endcap regions, as shown in Figure 5.3 (left). The combined energy resolution from tracker and ECAL is better than 4% and agrees with the expectations from simulation shown in Figure 5.3 (right).

5.4. Jet reconstruction

Jets are the signatures of the partonic quarks and gluons, and are composed out of the hadrons formed in the fragmentation process. Their energies need to be corrected for additional particles from pileup and detector effects. The use of b-tag algorithms allows to identify the jets stemming from the bottom quarks of the top-quark decay.

Jet clustering

Jets are formed from the particle-flow candidates using the anti-k_T cluster algorithm [113, 114]. The anti-k_T algorithm belongs to a class of infrared-safe and collinear-safe jet algorithms described in the following. An algorithm is infrared-safe when the resulting jets are unchanged by additional infinitesimally soft radiation. It is collinear-safe when an input particle can be split into two particles with nearly the same direction without changing the result. Both requirements arise from the singularity structure of QCD described in Chapter 3, in order to obtain an observable that is meaningful in perturbation theory and does not depend on the (parton shower) cutoff employed in the calculation.

A class of sequential cluster algorithms can be defined by the two quantities

$$\begin{aligned} d_{ij} &= \min\left(p_{\mathrm{T},i}^{2k}, p_{\mathrm{T},j}^{2k}\right) \frac{\Delta_{ij}^2}{R^2}, \\ d_{iB} &= p_{\mathrm{T},i}^{2k}, \end{aligned}$$

where Δ_{ij} denotes the distance in $\eta - \phi$ space, R is the radius parameter giving the characteristic cluster scale, and k steers the behavior of the algorithm. Both d_{iB} and d_{ij} are calculated for all input objects or pairs of objects, respectively. If the smallest one is a d_{ij} those both objects are clustered together. If a d_{iB} is smallest, the corresponding object is considered to be a jet and excluded from further clustering. The clustering stops when no objects are left.

With $k = 1$ one obtains the Durham k_T algorithm [115], with $k = 0$ the Cambridge-Aachen (CA) algorithm [116, 117], and with $k = -1$ the anti-k_T algorithm. These algorithms result in very different jet shapes, shown in Figure 5.4. At hadron colliders, the k_T and CA algorithms tend to cluster a significant amount of soft contributions from pileup or the underlying event in an irregular shape around the center of the jet. The anti-k_T algorithm results in cone-shaped jets with a maximum size defined by the radius R.

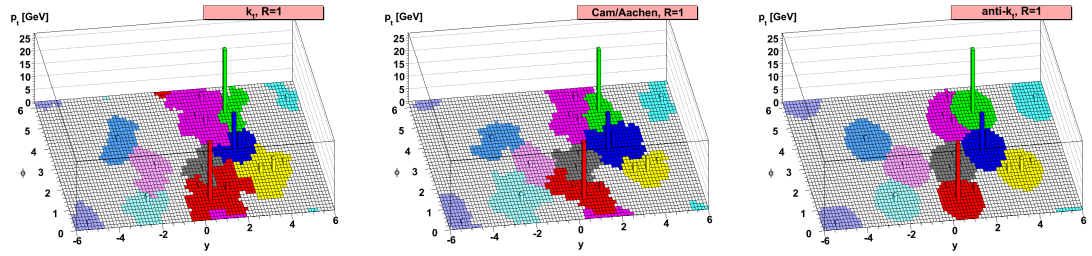

Figure 5.4.: Sensitivity of different jet cluster algorithms to soft radiation, Durham k_T (left), Cambridge-Aachen (center), anti-k_T (right). The effectively covered area in the $\eta - \phi$ plane is shown for the same event with several jets. From [114].

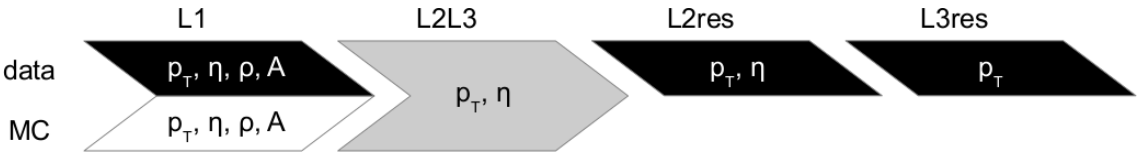

Figure 5.5.: Jet energy correction scheme used in the CMS collaboration. From [119].

The default jet reconstruction at CMS employs the anti-k_T algorithm with $R = 0.5$ clustered from particle-flow candidates. Charged particles are excluded if they are associated to another vertex than the selected primary vertex. Isolated leptons are excluded from the clustering.

In order to suppress fake jets, jet candidates are required to have an energy fraction from charged hadrons larger than 0.0 and smaller than 0.99, an energy fraction from neutral hadrons and photons each smaller than 0.99, and at least 2 jet constituents, where at least one is required to be a charged hadron, an electron, or a muon.

Jet energy corrections

Jet energy corrections (JEC) are employed to bring the energy (or rather the transverse momentum) of reconstructed particle-flow jets into agreement with the energy of the corresponding *GenJets*. The GenJets are defined as the "true" jets at the hadron level accessible in simulated events, excluding neutrinos and pileup contributions. The JEC scheme used in CMS [118] is shown in Figure 5.5.

Additional pileup events are found to add a transverse momentum of about 0.6 GeV per unit area in $\eta - \phi$ space and additional interaction. This amounts to 0.5 GeV on average per interaction for anti-k_T jets with $R = 0.5$, shown in Figure 5.6 (left). The impact of pileup events is reduced by rejecting tracks that stem

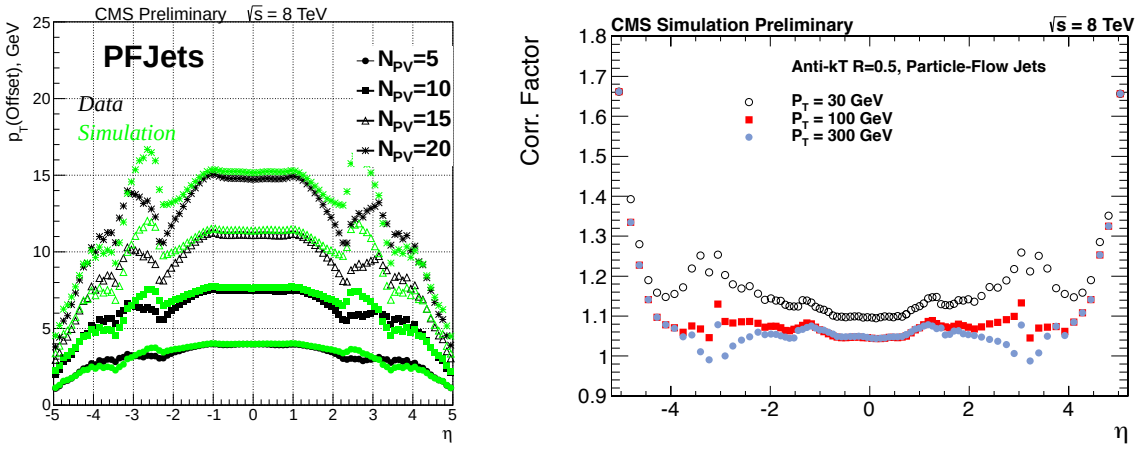

Figure 5.6.: Jet energy corrections at CMS: L1 pileup offset correction in dependence of the number of interaction vertices and the jet η (left), and the L2L3 MC truth correction in dependence of the jet p_{T} and η (right). Source: CMS collaboration [121].

from other vertices than the selected primary one. This charged hadron subtraction (CHS) is applied in this analysis. The *L1 pileup offset* correction adds a correction factor depending on the jet transverse momentum p_{T}, the jet pseudorapidity η, the offset momentum density ρ calculated from k_{T} jets with $R = 0.6$, and the jet area A, using an extension of the *hybrid jet area* method [113, 120]. The correction can be determined in MC simulation by comparing identical events with and without pileup events, and has a size of -15% or -25% for 30 GeV jets in the barrel with and without CHS, respectively. The data-to-simulation scale factor is determined using the *Random Cone* method that extracts the momentum offset density ρ from Zero Bias events [118].

The *L2L3 MC truth* response R is defined as $\langle p_{\mathrm{T}}^{\mathrm{reco}} \rangle \,/\, \langle p_{\mathrm{T}}^{\mathrm{gen}} \rangle$ binned in $p_{\mathrm{T}}^{\mathrm{gen}}$ and η. The reconstructed jet momenta are evaluated in a QCD dijet sample after detector simulation and L1 correction. The corresponding GenJets are found by requiring a distance $\Delta R < 0.25$ between the jet axes. The simulation predicts a p_{T}-dependent correction factor of $5 - 15\%$ needed in the barrel, accounting for the non-linear response in the calorimeter and tracking inefficiencies. The increase of inactive material in the endcap and forwards regions requires a correction factor of up to 70%, as shown in Figure 5.6 (right). The L1 and L2L3 MC truth corrections are applied to both data and simulated events.

Remaining data-MC differences are corrected using the p_T *balance* and the *Missing* E_T *Projection Fraction* (MPF) methods that calibrate the jets relative to reference objects, like leptonically decaying Z bosons, photons or jets in the barrel. The correction factor for jets in collision data is then given by the response ratio $R^\mathrm{sim}/R^\mathrm{data}$ in γ/Z+jet and dijet events. The p_T balance method imposes that the p_T of the probed jet should be equal to the reference p_T in the limit of no additional event activity α and defines the jet response as

$$R_\mathrm{jet}^{p_\mathrm{T}} = \left\langle \frac{p_\mathrm{T}^\mathrm{probe}}{p_\mathrm{T}^\mathrm{ref}} \right\rangle \Bigg|_{\alpha\to 0}, \quad \alpha = \frac{p'_\mathrm{T}}{p_\mathrm{T}^\mathrm{ref}},$$

where p'_T is here defined as the transverse momentum of the first additional jet. The MPF response takes advantage of the fact that any missing transverse energy/momentum in the probed event topologies should stem from a mismeasurement of jet momenta. The MPF response is defined as

$$R_\mathrm{jet}^\mathrm{MPF} = \left\langle 1 + \frac{\vec{E}_\mathrm{T}^\mathrm{miss} \cdot \vec{p}_\mathrm{T}^\mathrm{ref}}{\left(\vec{p}_\mathrm{T}^\mathrm{ref}\right)^2} \right\rangle \Bigg|_{\alpha\to 0}$$

and is found to have improved stability against additional radiation compared to the p_T balance method. The *L2 residual* correction is obtained from dijet events using the MPF method, and corrects for the η-dependence of the jet response, binned in p_T [18, 119]. The absolute scale and its p_T-dependence is then corrected by the *L3 residual* correction, obtained from γ/Z+jet events using the p_T balance and MPF methods in a range of $30 < p_\mathrm{T} < 700\,\mathrm{GeV}$. Extrapolation to even higher jet p_T is done using the *multijet balance* (MJB) method, where the reference object is a system of multiple calibrated lower-p_T jets.

As shown in Figure 5.7 (left), the size of the L2 residual corrections is up to 15% in the transition region between endcap and forward calorimeters but below 3% in the region covered by the tracker. The absolute L3 correction and its extrapolation to $\alpha \to 0$ is shown in Figure 5.7 (right). The `Winter14_V5` residual corrections applied in this analysis were derived as a function of the jet p_T. The p_T-dependence is fitted from the combination of the measurements in the γ+jet, $Z\,(\to \mu^+\mu^+)$+jet, and $Z\,(\to e^+e^+)$+jet channels, taking into account the p_T uncertainties of the respective reference objects.

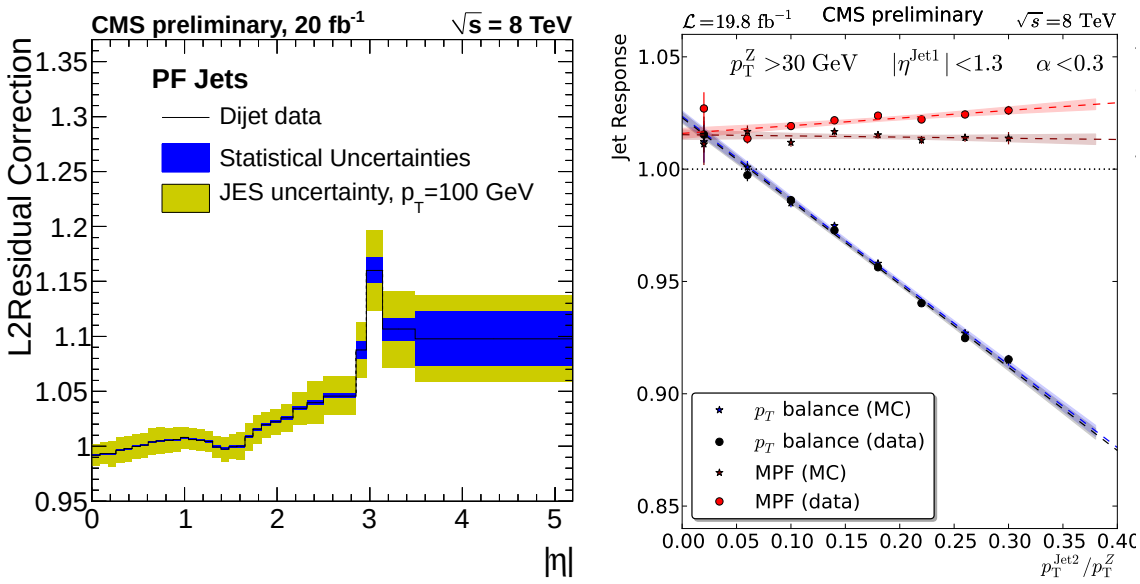

Figure 5.7.: Residual data–MC jet energy corrections at CMS: Relative L2 correction depending on η (left) and the absolute L3 correction with extrapolation to no additional jet activity (right). Source: CMS collaboration [121].

Jet energy uncertainties

For jets with $p_{\mathrm{T}} = 30\,\mathrm{GeV}$, the total JEC uncertainty is about 2% in the barrel and slightly larger at higher η, shown in Figure 5.8 (left). The time stability component is omitted as the analysis is performed on the complete 2012 data. The relatively large flavor uncertainty is given for a generic QCD sample, and will be recalculated for the flavor composition of the sample selected in the analysis. Excluding these two sources, the uncertainty is 1.5% in the barrel. The p_{T}-dependence is shown in Figure 5.8 (right). The uncertainty is lowest, below 0.5%, for jets with transverse momenta $100\,\mathrm{GeV} < p_{\mathrm{T}} < 300\,\mathrm{GeV}$, where most of the γ/Z+jet data reside. At low p_{T}, the dominant uncertainty comes from the difference between the pileup p_{T} offset density ρ in simulated events and as obtained from the Random Cone method in data.

The flavor uncertainty is determined for each jet flavor separately from a comparison of jets simulated with Pythia 6 Z2* and Herwig++ EE-3C, resulting in Figure 5.9. The flavor uncertainties are given relative to the constraints from data so that the uncertainty is lowest for the γ+jet, Z+jet and QCD (dijet) flavor mixtures at those p_{T}/η where the respective measurements of the absolute jet energy scale are most precise. Thus, shifting all jet momenta in a simulated sam-

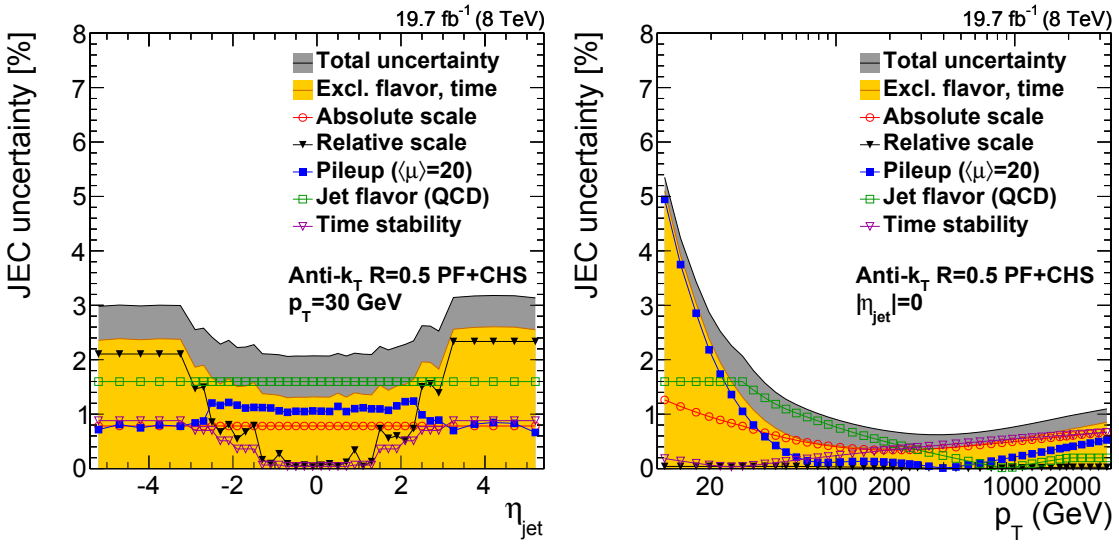

Figure 5.8.: Summary of the uncertainties on the `Winter14_V5` jet energy corrections at CMS: the uncertainty in dependence of η for jets with $p_{\mathrm{T}} = 30\,\mathrm{GeV}$ (left), and the p_{T}-dependent uncertainty for central jets (right).

ple by their signed flavor-dependent uncertainties corresponds to changing both the sample of interest and the reference sample for derivation of JEC from Pythia to Herwig++, with regard to the jet response. The large uncertainty for gluon jets of up to 2.5% therefore is a consequence of the sizable response difference found for gluons between Pythia and Herwig++ and of the dominance of light quarks in the data sample, so that the constraints for the gluon jets are weaker. The responses for quark jets are found to be in good agreement between both generators. Together with the good constraint from data, this results in a low uncertainty for light quarks on the order of 0.2%, with an opposite sign compared to gluons. The flavor uncertainty for bottom jets is approximately 0.5% and was recently validated using Z+b data [122]. The flavor uncertainties enter this analysis as a p_{T}-dependent component due to the uncertainty on the flavor composition of the Z+jet sample, and as relative uncertainties on the jet energies of each individual jet flavor in the $t\bar{t}$ sample.

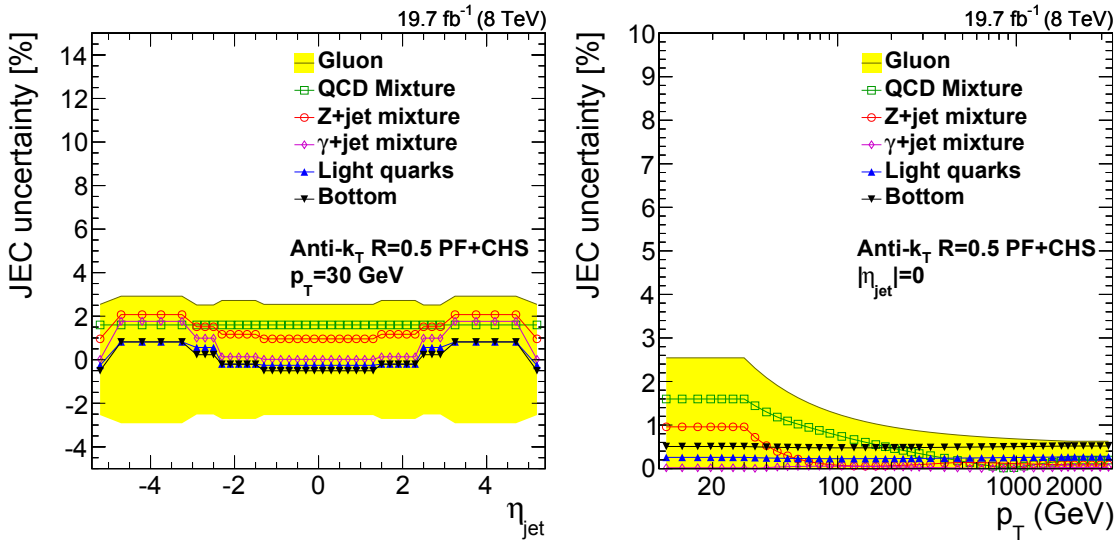

Figure 5.9.: Flavor uncertainties on the `Winter14_V5` jet energy corrections at CMS in dependence of η for jets with $p_{\mathrm{T}} = 30\,\mathrm{GeV}$ (left), and in dependence of p_{T} for central jets (right).

Jet energy resolution

The jet energy resolution (JER) was measured with the *dijet asymmetry* method [118, 123] which uses the p_{T} asymmetry distribution for dijet pairs defined as

$$\mathcal{A} = \frac{|p_{\mathrm{T}}^1 - p_{\mathrm{T}}^2|}{1/2 \cdot (p_{\mathrm{T}}^1 + p_{\mathrm{T}}^2)}.$$

As shown in Figure 5.10 (left), the resolution in data was found to be worse than in simulation by about 10% on average and up to 40%, at most depending on the detector region. The p_{T}-dependence of the resolution in data was found to be well modeled by the simulation. The later is shown in Figure 5.10 (right). Also shown is the impact of pileup events on the jet energy resolution. The JER uncertainty varies from 3% to 5% in the tracker region. A cross-check analysis performed in γ+jet events confirmed the results with slightly larger uncertainties.

B tagging

As $t\bar{t}$ events contain two bottom quarks, the identification of jets containing bottom hadrons is very useful for separating signal and background processes.

Algorithms for b-jet identification exploit the unique properties of the bottom hadrons, i.e., a mass of $m_B \simeq 5\,\mathrm{GeV}$ and a lifetime τ_B in the order of 10^{-12} s. The

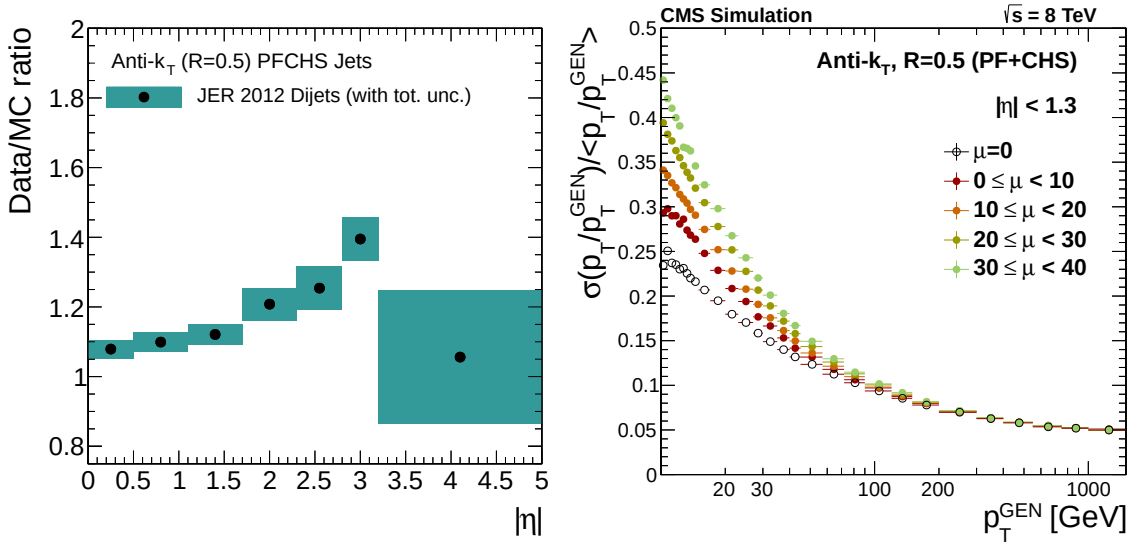

Figure 5.10.: Corrections for the jet energy resolution at CMS in dependency of η (left) [123], and resolution in dependency of p_{T} and the number of pileup events μ (right) [124].

flight distance of a B hadron in the detector is then $d = \gamma_B \beta_B \tau_B \approx 7$ mm with the relativistic Lorentz factor $\beta_B \gamma_B \approx 0.4 \cdot {}^{m_t}/_{m_B}$ as typical for a top-quark decay. The CMS detector is able to resolve the secondary vertices of B hadron decays using the pixel tracker. Typical observables used for b tagging are the flight distance (or its significance) to the secondary vertex, or reconstructed invariant mass of all charged particles assigned to the secondary vertex, which is of order ${}^{m_B}/_2$. Even without a reconstruction of a secondary vertex, observables like the number of tracks with large impact parameters associated to the primary vertex give good discrimination power against other jet flavors.

There are several algorithms implemented in CMS that usually use different sensitive observables combined into a (likelihood) discriminator value. Their performance expected from simulation is shown in Figure 5.11 (left) where the b-tag efficiency is plotted against the probability for falsely tagging a light jet (mistag). Working points are defined as loose/medium/tight, corresponding to 10%/1%/0.1% mistag rate. The *Track Counting High Purity* (TCHP), *Jet Probability* (JP) and *Combined Secondary Vertex* (CSV) algorithms are used in CMS analyses at $\sqrt{s} = 8\,\mathrm{TeV}$. Their performance at the different working points has been evaluated in data and p_{T}-dependent scale factors accounting for the different efficiencies in data and simulation were determined, shown in Figure 5.11 for the CSV algorithm at the medium working point (right) [125].

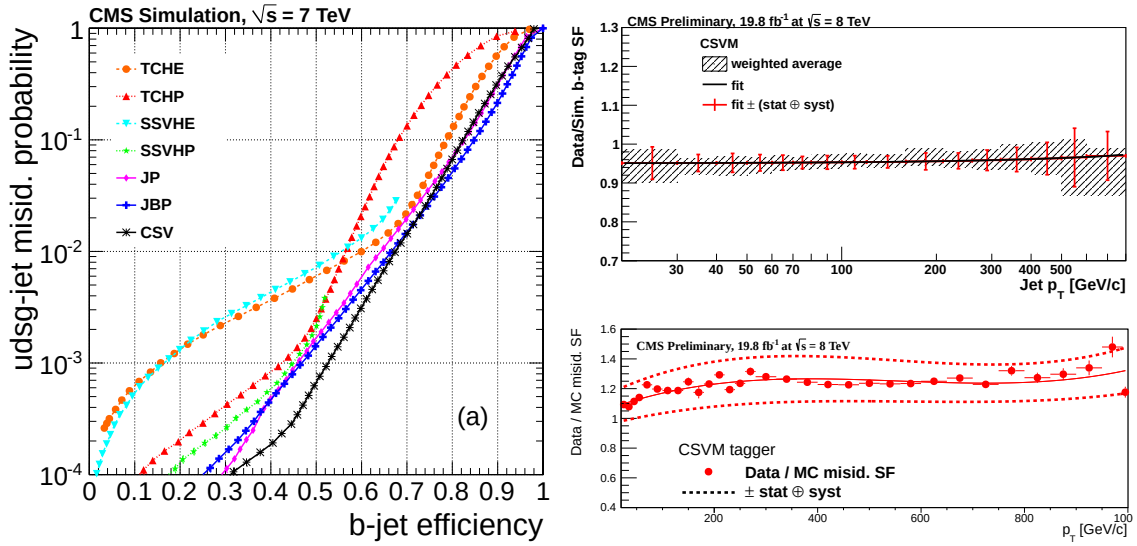

Figure 5.11.: Performance of the b-tag algorithms implemented in CMS [126] (left), data-MC scale factors for the b-tag efficiency (upper right), and mistag rate (lower right) of the CSV tagger at the medium working point in dependence of the jet p_{T} [125].

5.5. Missing transverse momentum

Neutrinos from the W-boson decays interact only weakly and thus leave no direct evidence in the detector. Still their p_{T} can be estimated by summing the vectorial transverse momenta of all other objects in the event. The imbalance

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{i \in \mathrm{particles}} \vec{p}_{T,i}$$

is called the *missing transverse momentum* or *energy* (MET). The MET reconstruction is improved by the so-called *Type-I corrections*, where effectively the momenta of the uncalibrated jets in the expression above are replaced with the momenta of the corresponding calibrated jets if their p_{T} is larger than 10 GeV.

As a significant amount of MET is also a signature in many searches for BSM physics, the MET performance of CMS has been carefully validated [127]. This analysis uses the MET as an estimate for the transverse momentum of the undetected neutrino from the leptonic W decay. No selection requirement on MET is imposed as the other selection steps already yield a very clean sample of $t\bar{t}$ events, as shown in Chapter 6.

6. Lepton+jets selection and top-quark reconstruction

The goal of the event selection is to obtain a pure sample of $t\bar{t}$ events in the muon+jets and electron+jets channels. Such events are expected to contain one high-energetic charged lepton, one undetected neutrino, and at least four quarks that give rise to jets. The charged lepton is required to be isolated, i.e., the amount of event activity in its vicinity is not allowed to exceed some threshold. Due to the large phase space available at the LHC, additional initial-state radiation jets are very likely to occur, potentially adding ambiguities to the event reconstruction. The identification of jets stemming from the bottom quarks in the top-quark decay is an important step to reduce the background from W+jets production. Finally, a kinematic fit is employed that checks the compatibility of a selected event with the $t\bar{t}$ hypothesis, using the precise knowledge of the W-boson mass but no a-priori assumption on the top-quark mass.

6.1. Datasets

The analysis is performed on both real collision data and simulated events, described in the following.

Collision data

For the presented measurement of the top-quark mass, the full 2012 data sample at $\sqrt{s} = 8$ TeV is analyzed. The runs where all detector systems were working as expected* correspond to an integrated luminosity of $19.7\,\text{fb}^{-1}$. Events are required to pass a single-muon trigger or a single-electron trigger, so that the analysis is

*The datasets are filtered to contain only certified runs using the following JSON file: `Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt`

performed on the following SingleMu and SingleElectron datasets that were reconstructed with the latest detector calibration and alignment after the data taking:

- /SingleMu/Run2012A-22Jan2013-v1/AOD
- /SingleMu/Run2012B-22Jan2013-v1/AOD
- /SingleMu/Run2012C-22Jan2013-v1/AOD
- /SingleMu/Run2012D-22Jan2013-v1/AOD
- /SingleElectron/Run2012A-22Jan2013-v1/AOD
- /SingleElectron/Run2012B-22Jan2013-v1/AOD
- /SingleElectron/Run2012C-22Jan2013-v1/AOD
- /SingleElectron/Run2012D-22Jan2013-v1/AOD

Simulation

Simulated events are used to develop and validate the analysis method. The simulation of $t\bar{t}$ signal events is done with the MadGraph 5.1.5.11 matrix element generator [64], MadSpin [128] for the decay of heavy resonances, Pythia 6.426 parton showering [56] using the $Z2^*$ tune [129], and Tauola [130] for decays of τ leptons. The nominal top-quark mass used for MC production is 172.5 GeV. Additional $t\bar{t}$ samples are produced with the nominal m_t and different generator setups in order to explore systematic variations. W/Z+jets background events are generated with MadGraph 5.1.3.30. The single-top-quark background is simulated using Powheg 1.0 r1380 [48,80,131–133], assuming a top-quark mass of 172.5 GeV. All samples are part of the `Summer12` central production campaign at CMS and the detector response is simulated using Geant4 [134]. The $t\bar{t}$, W/Z+jets, and single-top-quark samples are normalized to the theoretical predictions described in Refs. [37, 135–139]. The simulation includes effects of additional overlapping minimum-bias events (pileup) that are weighted to match the distribution expected in data. The complete list of MC samples used for analysis is given in Table 6.1.

Table 6.1.: Simulated samples used for the analysis. The top-quark mass is $m_t = 172.5\,\text{GeV}$ if not stated otherwise.

Dataset name	Number of events [M]	$\mathcal{L}\ [\text{fb}^{-1}]$	σ [pb]	Calculation
$t\bar{t}$ signal (MadGraph + MadSpin + Pythia 6 + Tauola)				
`TTJets_MSDecays_central`	62.1	249.4	249.1	NNLO+NNLL [37]
`TTJets_MSDecays_mass166_5` ($m_t = 166.5\,\text{GeV}$)	27.1	98.1	276.1	"
`TTJets_MSDecays_mass169_5` ($m_t = 169.5\,\text{GeV}$)	39.5	150.8	262.1	"
`TTJets_MSDecays_mass171_5` ($m_t = 171.5\,\text{GeV}$)	24.4	96.5	253.3	"
`TTJets_MSDecays_mass173_5` ($m_t = 173.5\,\text{GeV}$)	26.5	108.1	245.0	"
`TTJets_MSDecays_mass175_5` ($m_t = 175.5\,\text{GeV}$)	40.2	169.8	237.0	"
`TTJets_MSDecays_mass178_5` ($m_t = 178.5\,\text{GeV}$)	24.4	107.9	225.8	"
$t\bar{t}$ systematic variations (MadGraph + MadSpin + Pythia 6 + Tauola)				
`TTJets_MSDecays_scaleup`	41.9	168.2	249.1	"
`TTJets_MSDecays_scaledown`	39.2	157.7	249.1	"
`TTJets_MSDecays_matchingup`	37.1	148.9	249.1	"
`TTJets_MSDecays_matchingdown`	34.1	136.7	249.1	"
Table continues on next page ↓				

Table 6.1 continued from previous page

Dataset name	Number of events [M]	$\mathcal{L}$ $\left[\text{fb}^{-1}\right]$	σ [pb]	Calculation
$t\bar{t}$ systematic variations (MadGraph + Pythia 6 + Tauola)				
TTJets_*MGDecays	25.0 (ℓ+jets)	228.7	249.1	NNLO+NNLL [37]
TTJets_*MGDecays_TuneP11	12.0 (ℓ+jets)	110.0	249.1	"
TTJets_*MGDecays_TuneP11noCR	24.9 (ℓ+jets)	228.6	249.1	"
TTJets_*MGDecays_TuneP11mpiHi	8.0 (ℓ+jets)	73.1	249.1	"
TTJets_*MGDecays_TuneP11TeV	7.9 (ℓ+jets)	72.0	249.1	"
$t\bar{t}$ systematic variations (NLO)				
TT_CT10_TuneZ2star_8TeV-powheg	21.7	87.0	249.1	"
TT_8TeV-mcatnlo	32.9×0.77[†]	101.8	249.1	"
V+jets background (MadGraph + Pythia 6 + Tauola)				
DYJetsToLL_M-50	30.5	8.7	3503.7	NNLO [136]
WJetsToLNu			36257.0	"
– W(0)JetsToLNu	42.9	1.6	74.4%	LO [64]
– W1JetsToLNu	23.1	3.6	17.6%	"
– W2JetsToLNu	34.0	16.7	5.6%	"
– W3JetsToLNu	15.5	25.4	1.7%	"
– W4JetsToLNu	13.4	53.4	0.7%	"

Table continues on next page ↓

[†]In the MC@NLO method, the effective number of events is reduced by negative weights.

Table 6.1 continued from previous page

Dataset name	Number of events [M]	$\mathcal{L}$ $\left[\text{fb}^{-1}\right]$	σ [pb]	Calculation
Single top background (POWHEG (NLO) + PYTHIA 6 + TAUOLA)				
`TBarToDilepton_tW-channel-DR{_scaleup/down}`	3.0	2550.1	1.2	NNLO [137]
`TBarToLeptons_s-channel{_scaleup/down}`	2.0	3506.7	0.6	"
`TBarToLeptons_t-channel{_scaleup/down}`	1.7	172.1	9.9	NLO [138, 139]
`TBarToThadWlep_tW-channel-DR{_scaleup/down}`	1.5	614.1	2.4	NNLO [137]
`TBarToTlepWhad_tW-channel-DR{_scaleup/down}`	1.5	603.2	2.4	"
`TToDilepton_tW-channel-DR{_scaleup/down}`	3.0	2554.4	1.2	"
`TToLeptons_s-channel{_scaleup/down}`	3.9	3202.6	1.2	"
`TToLeptons_t-channel{_scaleup/down}`	3.9	223.8	17.5	NLO [138, 139]
`TToThadWlep_tW-channel-DR{_scaleup/down}`	1.5	599.6	2.4	NNLO [137]
`TToTlepWhad_tW-channel-DR{_scaleup/down}`	1.5	612.9	2.4	"

Table continues on next page ↓

Table 6.1 continued from previous page

Dataset name	Number of events [M]	$\mathcal{L}$ $\left[\text{fb}^{-1}\right]$	σ [pb]	Calculation
QCD multijet background (Pythia 6)				
QCD_Pt_20_MuEnrichedPt_15	21.5	0.2	134680.0	LO [56]
QCD_Pt_20_30_BCtoE	1.7	0.0	167388.0	"
QCD_Pt_30_80_BCtoE	2.0	0.0	167040.0	"
QCD_Pt_80_170_BCtoE	1.9	0.1	12981.9	"
QCD_Pt_170_250_BCtoE	1.9	3.1	632.0	"
QCD_Pt_250_350_BCtoE	2.0	19.7	103.0	"
QCD_Pt_350_BCtoE	1.9	81.5	23.9	"
QCD_Pt_20_30_EMEnriched	35.0	0.0	2914860.0	"
QCD_Pt_30_80_EMEnriched	33.1	0.0	4615893.0	"
QCD_Pt_80_170_EMEnriched	34.5	0.2	183295.0	"
QCD_Pt_170_250_EMEnriched	31.7	6.9	4587.0	"
QCD_Pt_250_350_EMEnriched	34.6	62.1	557.0	"
QCD_Pt_350_EMEnriched	34.1	382.5	89.1	"
Diboson background (Pythia 6 + Tauola)				
WW	10.0	182.5	54.8	NLO [135]
WZ	10.0	301.2	33.2	"
ZZ	9.8	1209.9	8.1	"

6.2. Trigger

The minimum trigger threshold on the transverse momentum (p_{T}) of an isolated muon is 24 GeV.[‡] The electron trigger requires one isolated electron with $p_{\mathrm{T}} > 27$ GeV.[§]

Scale factors for the lepton trigger, reconstruction, and selection efficiencies have been determined via the tag-and-probe method in $Z \to \ell^+\ell^-$ events in [140, 141] in the context of differential $t\bar{t}$ cross section measurements. The *tag* lepton is required to fire the trigger and fulfill tight selection criteria. The *probe* lepton candidate is then reconstructed using the inner tracker only and required to be oppositely-charged, where the invariant mass of the lepton pair has to be within a window of 15 GeV around the Z-boson mass. Efficiencies are determined relative to these candidates by applying identification, selection, and trigger criteria to the *probe* leptons. The data-to-MC scale factors are the ratios of the efficiencies, and depend on the lepton p_{T} and η. The combined average scale factor is determined to be 0.966 ± 0.001 (stat.) for muons and 0.976 ± 0.001 (stat.) for electrons.

6.3. Preselection

Events pass the preselection if they contain one high-energetic lepton associated to the primary vertex. In the muon+jets channel, a single muon trigger is required, and an isolated muon with transverse momentum $p_{\mathrm{T}} > 33$ GeV and pseudorapidity $|\eta| < 2.1$ is selected in the offline reconstruction. The selection in the electron+jets channel requires a single electron trigger and an electron with $E_{\mathrm{T}} > 33$ GeV and $|\eta| < 2.1$ in the offline reconstruction.

In order to suppress the contributions from Z+jets processes and the $t\bar{t}$ dilepton decay channel, events fail selection in both sub-channels if at least one additional isolated lepton is detected. The veto is applied if an additional muon is reconstructed globally or in the tracker with $p_{\mathrm{T}} > 10$ GeV, $|\eta| < 2.5$ and an isolation $I\,(R = 0.4) < 0.2$. The thresholds for veto electrons are $E_{\mathrm{T}} > 20$ GeV, $|\eta| < 2.5$ and an isolation $I\,(R = 0.3) < 0.15$, with relaxed requirements on the electron identification.

An event passes the preselection if at least 4 jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$ are present. No b tag is required at this stage.

[‡]Trigger path `HLT_IsoMu24_eta2p1_v*`
[§]Trigger path `HLT_Ele27_WP80_v*`

Table 6.2 shows the preselection results in the column "preselection". In data, 582 584 events are selected. The fraction of $t\bar{t}$ signal as estimated from simulation is 57%. The largest background source is the W+jets process that has a similar signature and contributes a fraction of 29%.

Lepton observables are well described by the simulation within the systematic uncertainties, as shown in Figure 6.1 for p_{T}, η, ϕ, and the flavor of the lepton. The lepton η distribution is wider for the background processes than for $t\bar{t}$ because the leptonically decaying W bosons can be produced at smaller $\hat{s} = x_1 x_2 s$. Therefore, the ratio x_1/x_2 can take more extreme values and the produced W bosons can get a higher boost in ±z–direction. The pp initial state also results in a larger cross section for W^+ than for W^- production, leading to an enhancement of positively charged leptons. This predicted charge asymmetry is in good agreement with the data. At this stage, more electron than muon events are selected due to non-prompt electrons from QCD multijet background. Individual simulated QCD events with large event weights lead to spikes in the flat lepton ϕ distribution.

Figure 6.2 shows the number of b-tagged jets among the four leading in p_{T} (that will later be used for top quark reconstruction), and the number of jets with $p_{\mathrm{T}} > 20$ GeV. The jets are tagged with the CSV algorithm that combines reconstructed secondary vertices and track-based lifetime information, using the medium working point (cf. Section 5.4). Figure 6.3 shows the distribution of missing transverse momentum, and the distance in η-ϕ space between the selected lepton and the nearest jet.

The preselection plots include uncertainties from the number of pileup events and from the cross section of signal and background processes. The following uncertainties on the cross sections are assumed: 5.3% for $t\bar{t}$, 20% for V+jets (from the heavy flavor component), 10% for single top, 100% for QCD multijet, and 5% for diboson background [142]. The discrepancies at higher jet multiplicities are expected to be covered by shape uncertainties of the $t\bar{t}$ signal, as shown later in Section 6.4.

The data excess at low ΔR between lepton and nearest jet is a feature of the electron+jets channel and can have different sources. One possibility is a contribution from processes with a very high cross-section, e.g. QCD multijet background. Isolated electrons overlapping with jets can be faked by topologies where a high-momentum track is surrounded by additional particle-flow candidates within $0.3 < \Delta R < 0.5$. Another possible source is spread of bremsstrahlung from a real electron due to interactions with the tracker that might be larger than simulated. These topologies are fostered by the relatively small isolation cone used for electrons.

Table 6.2.: Event yields and fractions for signal and background processes. The uncertainties are given by the quadratic sum of statistical uncertainties from finite MC sample sizes, as well as systematic uncertainties from process cross sections and number of pileup events. For the $t\bar{t}$-dominated "baseline" and "final" selections, the uncertainties contain in addition the contributions from jet energy scale and resolution, renormalization and factorization scales, matching threshold, ME generator, and top-p_{T} weighting (cf. Chapter 8).

Sample	preselection (1 ℓ, 4 jets)			baseline selection (+ 2 b tags, $\Delta R(\ell, \mathrm{jet})$)			final selection (+ P_{gof})		
	events	$\pm$	[%]	events	$\pm$	[%]	events	$\pm$	[%]
$t\bar{t}$ +jets	316810.0	16794.2	56.6	94356.2	5003.9	92.6	26966.3	890.6	96.3
Z+jets	24762.8	4977.5	4.4	447.0	110.9	0.4	64.8	27.8	0.2
W+jets	163213.5	32648.0	29.2	2108.5	426.6	2.1	211.7	46.9	0.8
Single top	17749.2	1775.9	3.2	4089.1	409.9	4.0	724.0	73.2	2.6
QCD multijet	33603.0	34174.3	6.0	777.0	948.3	0.8	17.3	24.5	0.1
Diboson	3602.7	184.1	0.6	112.1	8.2	0.1	15.2	2.4	0.1
MC total	559741.3	50435.9	100	101889.8	5128.4	100	27999.5	895.6	100
Data	582584.0	763.0	104	104580.0	323.5	103	28199.0	168	101

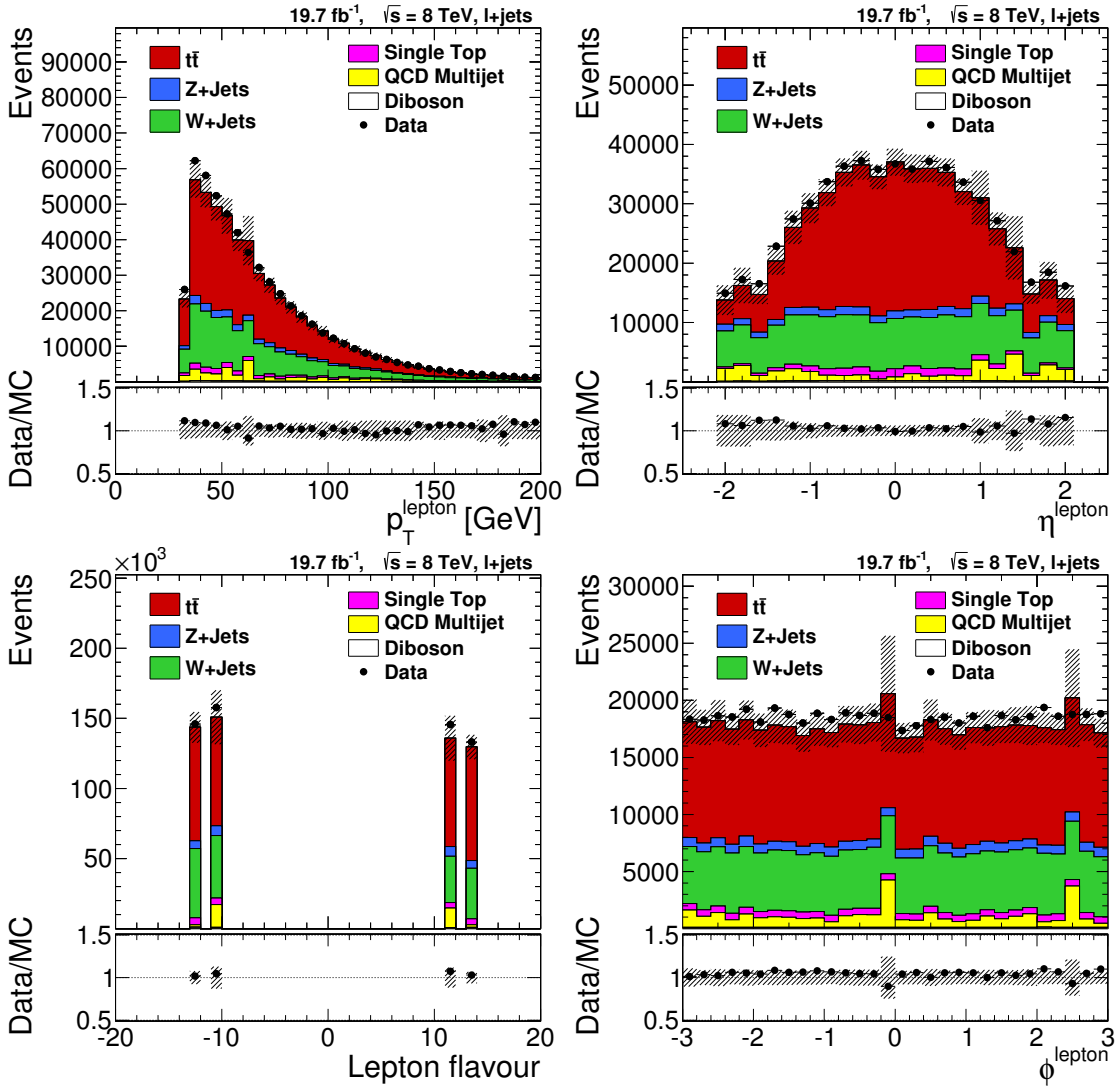

Figure 6.1.: Preselection: Upper row: reconstructed lepton p_{T} (left) and η (right). Lower row: lepton flavor (left; $-13 = \mu^+, -11 = e^+, 11 = e^-, 13 = \mu^-$) and ϕ (right). The plots are normalized to luminosity. The hatched area indicates the quadratic sum of statistical uncertainties from finite MC sample sizes, as well as systematic uncertainties from process cross sections and number of pileup events.

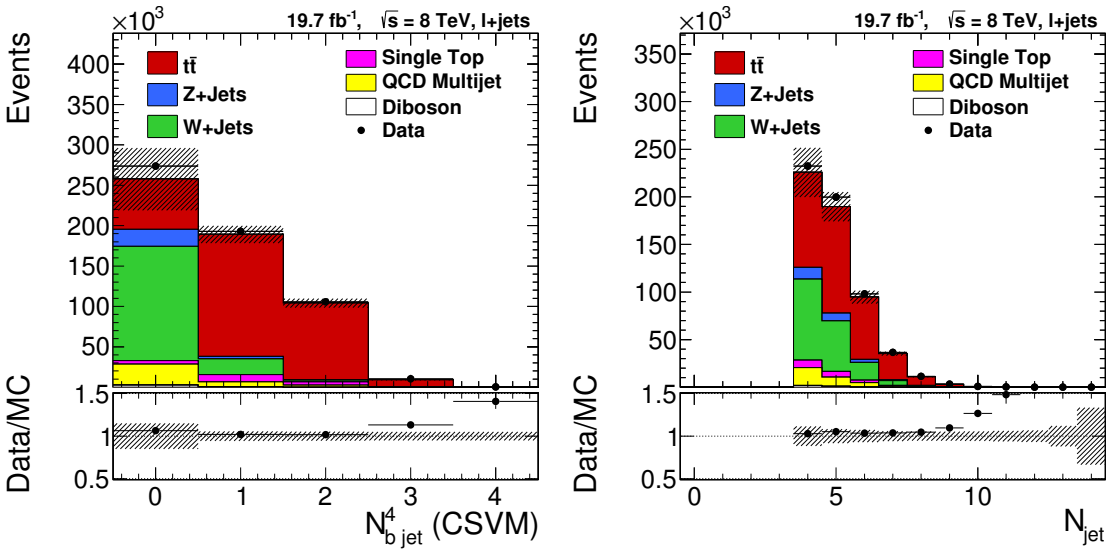

Figure 6.2.: Preselection: Number of b-tagged jets among the four leading jets with $p_{\mathrm{T}} > 30$ GeV, (left), and number of jets with $p_{\mathrm{T}} > 20$ GeV.

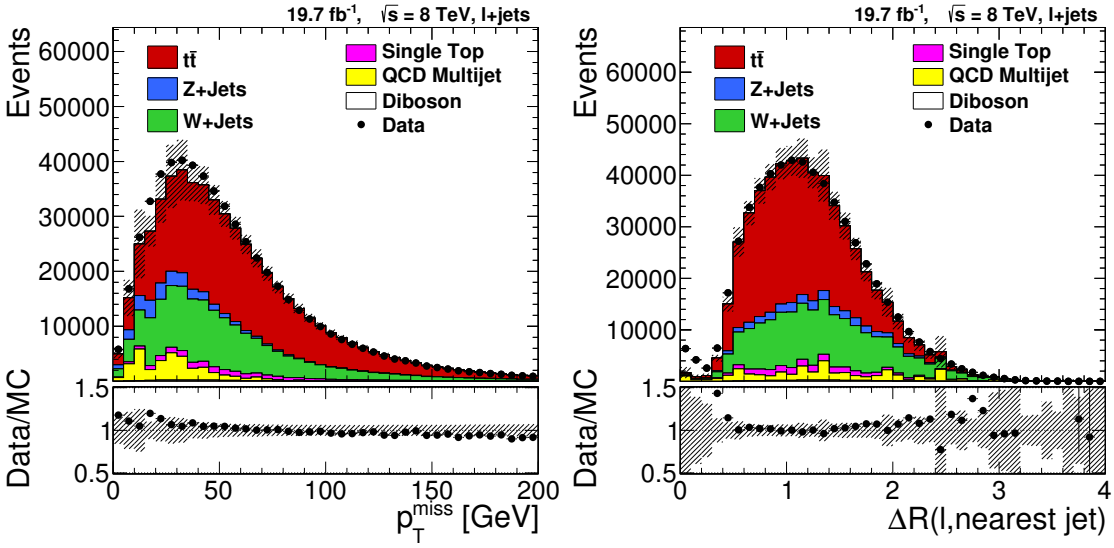

Figure 6.3.: Preselection: Missing transverse momentum (left), and minimum distance between the selected lepton and a jet with $p_{\mathrm{T}} > 20$ GeV (right).

As the jet axis is usually close to the electron direction, an additional selection criterion of $\min(\Delta R(\ell, \text{jet})) > 0.3$ with $p_{\text{T}}^{\text{jet}} > 20\,\text{GeV}$ is applied to remove this class of events.

6.4. Baseline selection

As shown in Figure 6.2, the signal fraction can be enhanced by identifying jets stemming from a bottom quark. Exactly two b-tagged jets are required among the four leading jets. The event weight applied to simulation for the b-tag efficiency correction is derived as $w = P(\text{data})/P(\text{MC})$ with

$$\begin{aligned} P(\text{MC}) &= \prod_{i=\text{tagged}} \varepsilon_i \times \prod_{j=\text{not tagged}} (1-\varepsilon_j), \\ P(\text{data}) &= \prod_{i=\text{tagged}} \text{SF}_i\varepsilon_i \times \prod_{j=\text{not tagged}} (1-\text{SF}_j\varepsilon_j), \end{aligned}$$

where ε_i is the tagging efficiency in simulation, and SF_i the data/MC scale factor shown in Figure 5.11. SF_i and ε_i are functions of the jet flavor, p_{T} and η. The indices i, j run over the four leading jets.

After baseline selection, 104 580 $t\bar{t}$ candidate events are selected in data, while 101 890 events are expected from simulation. The estimated selection efficiency for the $t\bar{t}$ signal is 2.2%. From simulation, the event composition due to $t\bar{t}$ is expected to be $> 90\%$, as shown in Table 6.2, column "baseline selection". Hence, the selection leads to a very clean sample of $t\bar{t}$ events (see also [143, 144]).

The same control distributions as for the preselection are shown in Figs. 6.4 – 6.6, and demonstrate that a very pure sample of top-quark pairs is obtained. The origin of the data excess at large $\Delta R(\ell, \text{jet})$ is presently unknown. The transverse momenta of the four leading jets are shown in Figure 6.7. The uncertainty bands take into account statistical uncertainties from finite MC sample sizes, as well as systematic uncertainties from process cross sections, the number of pileup events, jet energy scale and resolution, renormalization and factorization scales, ME-PS matching threshold, ME generator, and top-p_{T} weighting described in detail in Chapter 8.

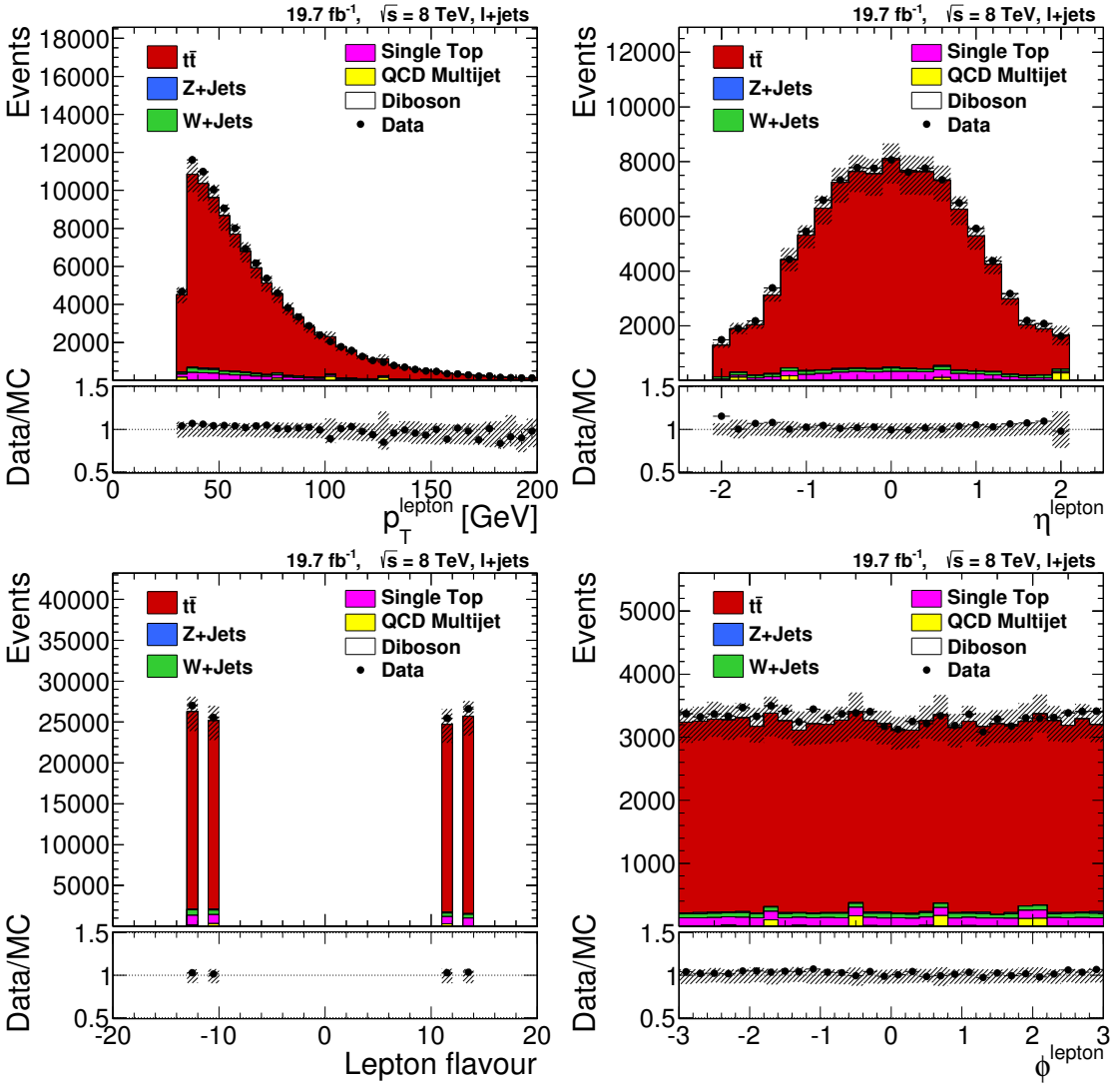

Figure 6.4.: Baseline selection: Upper row: reconstructed lepton p_{T} (left) and η (right). Lower row: lepton flavor (left; $-13 = \mu^+, -11 = e^+, 11 = e^-, 13 = \mu^-$) and ϕ (right). The plots are normalized to luminosity. The hatched area indicates statistical uncertainties from finite MC sample sizes, as well as systematic uncertainties from process cross sections, the number of pileup events, jet energy scale and resolution, renormalization and factorization scales, ME-PS matching threshold, ME generator, and top-p_{T} weighting.

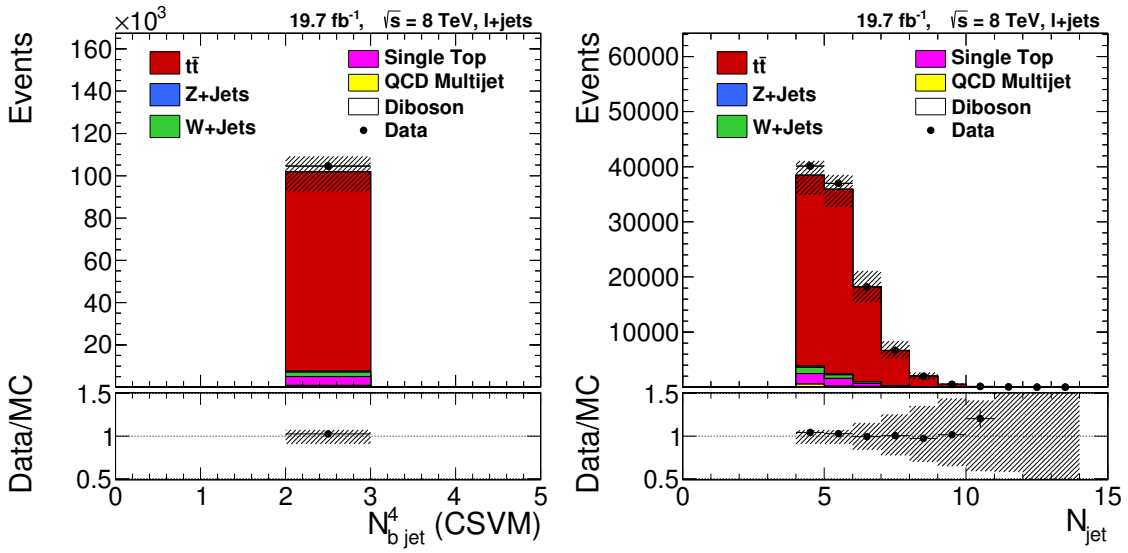

Figure 6.5.: Baseline selection: Number of b-tags among the four leading jets with $p_{\mathrm{T}} > 30$ GeV, (left), and number of jets with $p_{\mathrm{T}} > 20$ GeV.

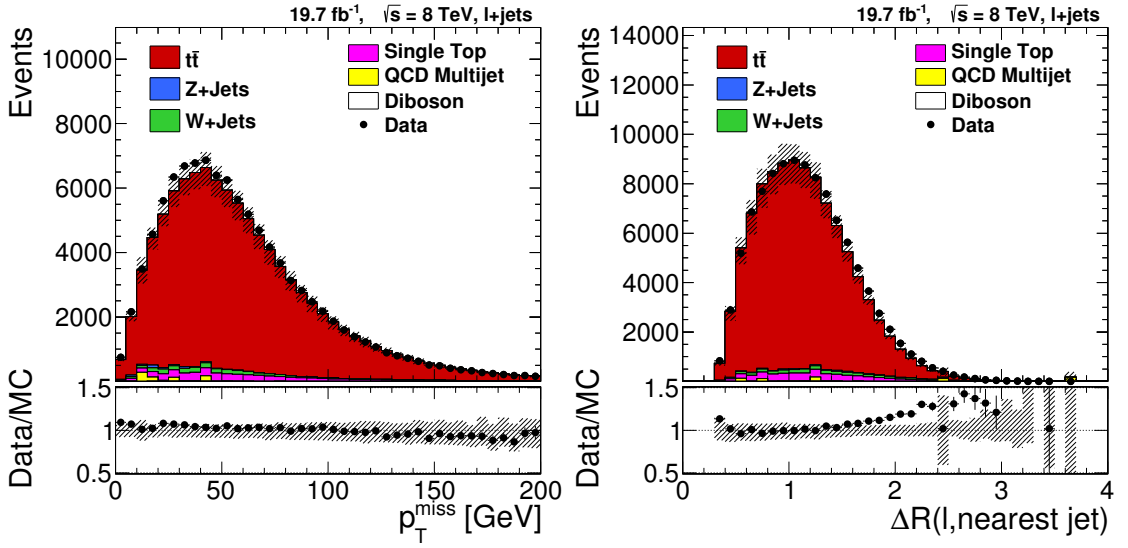

Figure 6.6.: Baseline selection: Missing transverse momentum (left), and minimum distance between the selected lepton and a jet with $p_{\mathrm{T}} > 20$ GeV (right).

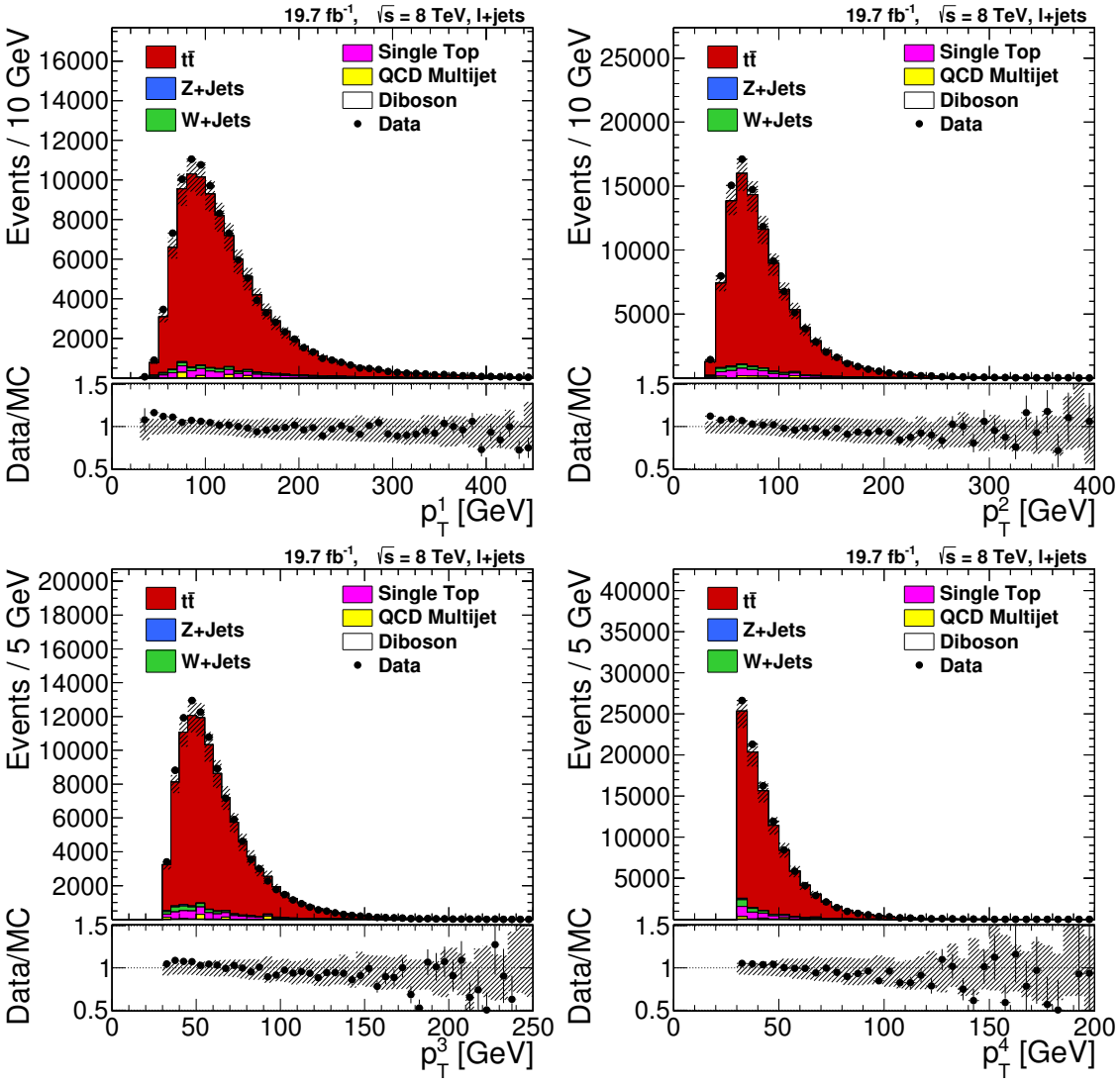

Figure 6.7.: Baseline selection: Reconstructed transverse momenta of the four leading jets.

6.5. Mass reconstruction

The invariant mass of the jets stemming from the hadronically decaying top quark is the most sensitive observable for measuring the top-quark mass. The four leading jets are used for the reconstruction of the top quarks in the $t\bar{t}$ hypothesis. The two b-tagged jets serve as candidates for the b quarks, while the two untagged jets build a W-boson candidate. This leads to two possible jet-parton permutations per event as the W candidate can be associated to either of the bottom jets to form a top-quark candidate. Both permutations are considered from now on. The mass distributions of reconstructed W candidates m_W^{reco} and of reconstructed top-quark candidates m_t^{reco} are shown in Figure 6.8. For simulated $t\bar{t}$ events, the four-momenta of the top-quark decay products are known. The jet-parton assignments then can be classified as

- *correct permutations* (cp): The four quarks i from the $t\bar{t}$ decay can be matched unambiguously to the four leading jets j within a distance $\Delta R_{ij} < 0.3$. The correct W candidate is identified and associated to the correct bottom jet to form the top-quark candidate.

- *wrong permutations* (wp): The correct jets are selected but not correctly associated to the partons, e.g. by confusing the two bottom jets. Events can completely consist of wrong permutations if one or both bottom jets are not tagged and one or two mistags (especially from charm jets) occur instead.

- *unmatched permutations* (un): Events and their permutations are unmatched if one of the four leading jets cannot be matched to a quark from the $t\bar{t}$ decay. This happens when high-energetic initial state radiation leads to one of the four leading jets, two or more partons from the $t\bar{t}$ decay are within the matching radius of a jet, at least one of the correct jets fails the selection criteria, or the initial parton direction is out of the detector acceptance.

It is observed that the mass distributions for the correct permutations have Gaussian shapes around the true masses of the W boson and the top quark. The wrong and unmatched permutations tend to contribute to a large tail at higher masses.

As the classification is based on parton-level information, it is not accessible in data. It is however useful to optimize the selection for correct permutations that are expected to be closest to the LO parton-level configuration for reconstructed masses and other observables.

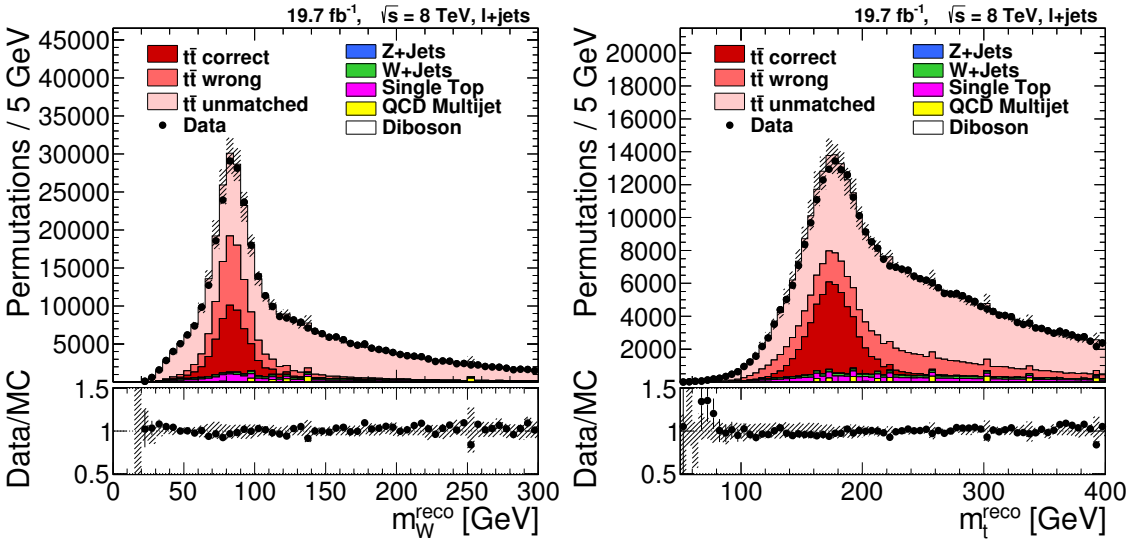

Figure 6.8.: Baseline selection: Reconstructed masses of the W bosons decaying to $q\bar{q}$ pairs (left) and the corresponding top quarks (right). The number of permutations in simulation is normalized to the number observed in data.

6.6. Kinematic fit and final selection

A constrained kinematic fit [145,146] is employed to suppress wrong and unmatched permutations, and to improve the resolution of the reconstructed observables. The inputs to the fit are the four-momenta of the 4 leading jets, the charged lepton ℓ, and

$$\vec{k}_{\mathrm{T}} = \vec{E}_{\mathrm{T}}^{\mathrm{miss}} + \vec{p}_{\mathrm{T}}\,(\ell) + \sum_{\mathrm{jets}} \vec{p}_{\mathrm{T}}\,(\mathrm{jet})\,,$$

where the sum runs over the 4 leading jets. The $\vec{k}_{\mathrm{T}}$ quantity effectively removes the lepton and the leading jets from the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ definition (c.f. Section 5.5), so that the fit input quantities are uncorrelated.

The reconstructed $t\bar{t}$ system is described by the six four-vectors of the top-quark decay products. Six constraints are given by the known particle masses, where charged leptons, neutrinos and light quarks are assumed to be massless, and the mass of the b quark is set to 4.7 GeV. Additional 5×3 constraints are provided by the measured momenta of the charged lepton and the jets, and 2 constraints stem from the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ of the event. The invariant mass of the two untagged jets m_{jj} and that of the lepton and the neutrino $m_{\ell\nu}$ are fixed to the mass of the W boson of $m_W = 80.4\,\mathrm{GeV}$, yielding two additional constraints. The natural width of the W

Table 6.3.: Permutation fractions for different signal classes.

Permutation type	baseline	final $P_{\text{gof}} > 0.2$	effective $\times P_{\text{gof}}$
$t\bar{t}$ correct	12.8 %	41.7 %	43.8 %
$t\bar{t}$ wrong	15.9 %	21.1 %	20.8 %
$t\bar{t}$ unmatched	71.3 %	37.2 %	35.4 %

boson is neglected as it is smaller than the experimental resolution. The uncertainty on m_W of 0.02% is negligible in the context of this analysis. The last constraint is the requirement of equal top-quark masses in the event which also neglects the finite decay width of the top quark but is true when averaging over a larger sample. In total, there are 26 constraints for 24 quantities, yielding 2 degrees of freedom.

For technical reasons, the kinematic fit needs starting values for all quantities, including the unmeasured z–component of the neutrino momentum, p_z^ν. Imposing the equality of the top-quark mass at both decay branches results in a quadratic equation for p_z^ν that has two solutions. The number of permutations in the event is doubled by the possibility of 2 initial p_z^ν values for each b-jet assignment.

The HitFit program [145] then minimizes $\chi^2 = (\mathbf{x} - \mathbf{x}^m)^T G (\mathbf{x} - \mathbf{x}^m)$ where $\mathbf{x}^m$ is the vector of measured observables, $\mathbf{x}$ the vector of fitted observables, and G is the inverse error matrix which is given by the resolutions of the observables. The goodness-of-fit (gof) probability for the kinematic fit with two degrees of freedom is $P_{\text{gof}} = \exp\left(-\frac{1}{2}\chi^2\right)$.

Figure 6.9 shows χ^2 and P_{gof} of the kinematic fit, the fraction of combination types, and signal-to-background performance. The number of permutations in simulation is normalized to the number observed in data, so that the uncertainty band is given solely by shape uncertainties. The *correct* permutations have a flat P_{gof} distribution, while the two other permutation types cluster at lower values of P_{gof}. Imposing a minimum requirement of P_{gof} therefore enhances the fraction of correct permutations. The *signal fraction* $S/(S+B)$ and the *signal significance* $S/\sqrt{B}$ are shown as a function of the selection requirement on P_{gof}, where S denotes the number of correct permutations, while B is the sum of wrong and unmatched permutations. Both measures are shown with and without weighting by P_{gof}. The signal fraction rises constantly while the significance has a maximum and steeply falls as the number of permutations goes to 0 for $P_{\text{gof}} \to 1$.

In order to increase both the signal (correct $t\bar{t}$ permutations) fraction and significance, the goodness-of-fit probability is required to be at least 0.2. This *final*

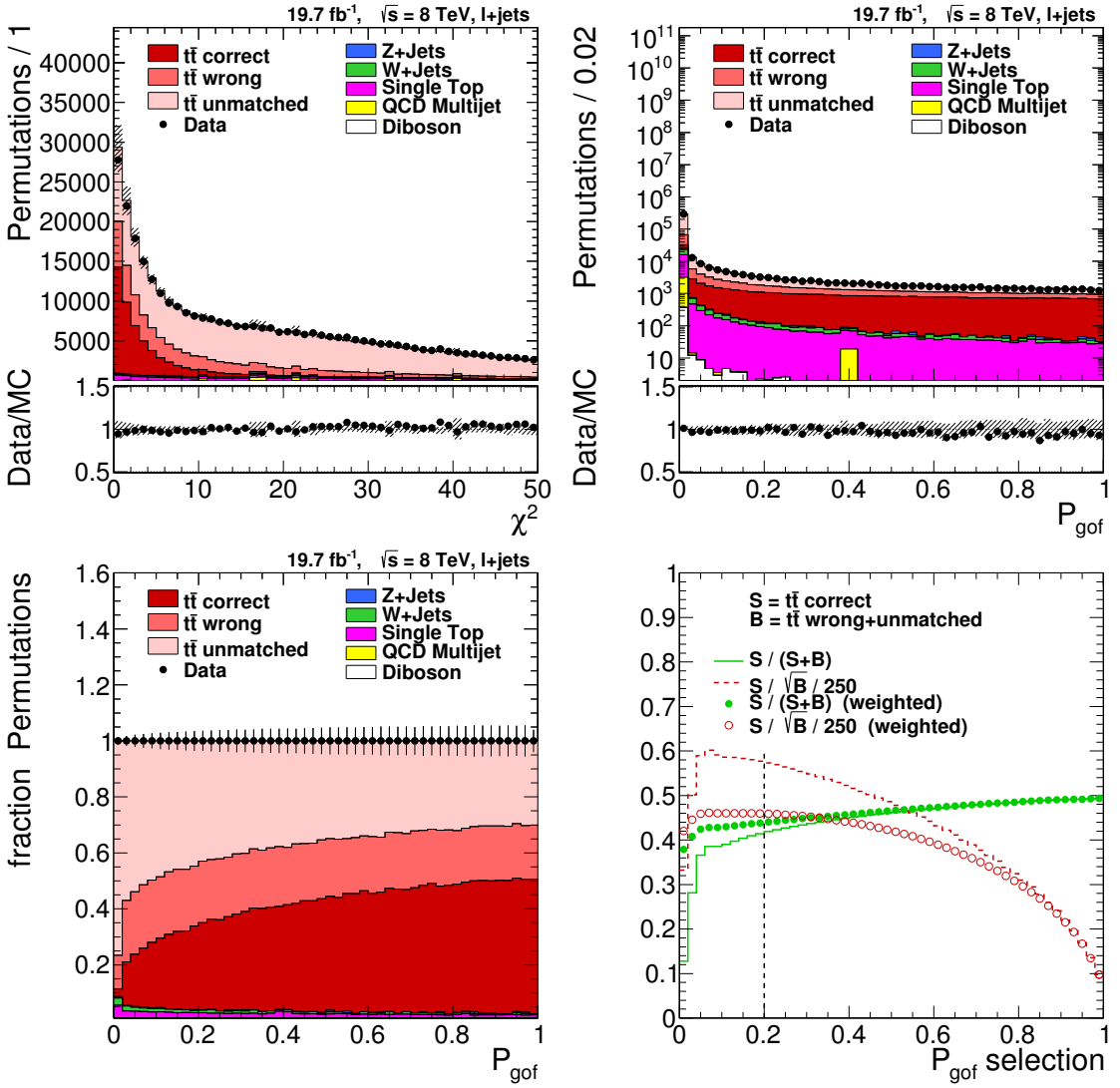

Figure 6.9.: Baseline selection: χ^2 of the kinematic fit (upper left), goodness-of-fit probability P_{gof} (upper right), fraction of permutation types in dependency of P_{gof} (lower left), and signal-to-background performance for a selection requirement on P_{gof} (lower right). The number of permutations in simulation is normalized to the number observed in data. The signal significance $S/\sqrt{B}$ is scaled by a factor $1/250$ to be shown on the same axis as the signal fraction $S/(S+B)$.

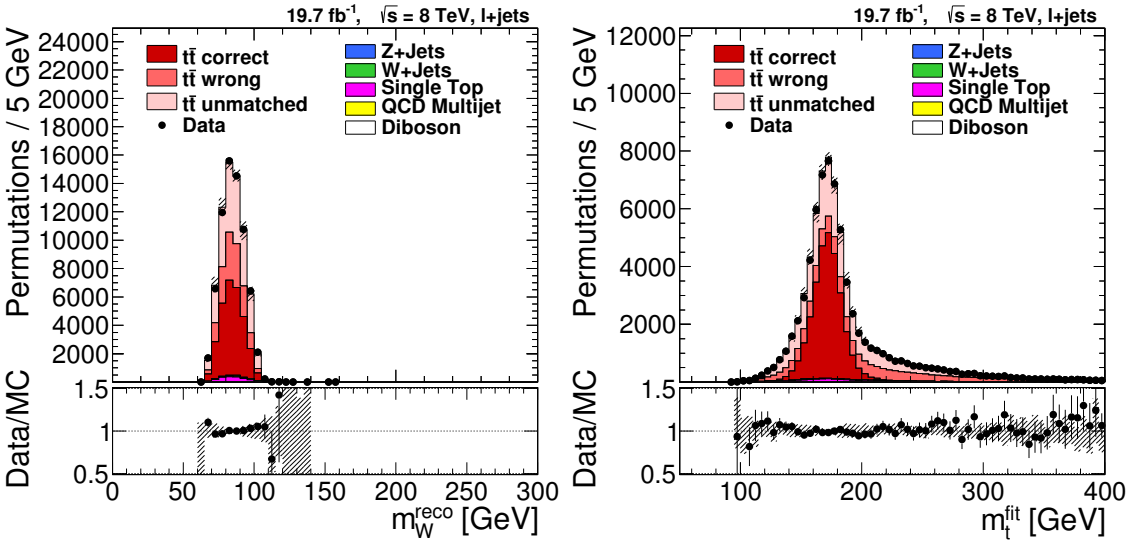

Figure 6.10.: Final selection: The reconstructed W-boson masses (left) and the fitted top-quark masses (right) after the goodness-of-fit selection. The binning is identical to that used in Figure 6.8. The number of permutations in simulation is normalized to the number observed in data.

selection yields 28 199 events for the mass measurement, containing 69 880 permutations passing this requirement. As shown in Table 6.3, the fraction of correct permutations estimated from simulation improves from 13% to 42%. The non-$t\bar{t}$ background is reduced to 3.6%, see Table 6.2. In the mass extraction, the permutations are weighted by their P_{gof} values, yielding an *effective* fraction of correct permutations of 44%.

Figure 6.10 shows the distributions of the W boson mass as reconstructed from the $q\bar{q}$ pair m_W^{reco} and the top-quark mass after the kinematic fit m_t^{fit} and the P_{gof} selection, exhibiting an improved reconstruction compared to the distributions shown in Figure 6.8. The *reconstructed* W-boson mass after the kinematic fit (but using the original four-momenta of the light jets) is chosen as observable here as the *fitted* W-boson mass m_W^{fit} is always equal to 80.4 GeV due to the mass constraint. Figure 6.11 zooms in at the peak regions and demonstrates good agreement between data and simulation. These distributions will be used to extract the top-quark mass as described in the following chapters. No P_{gof} weighting is applied in the figures.

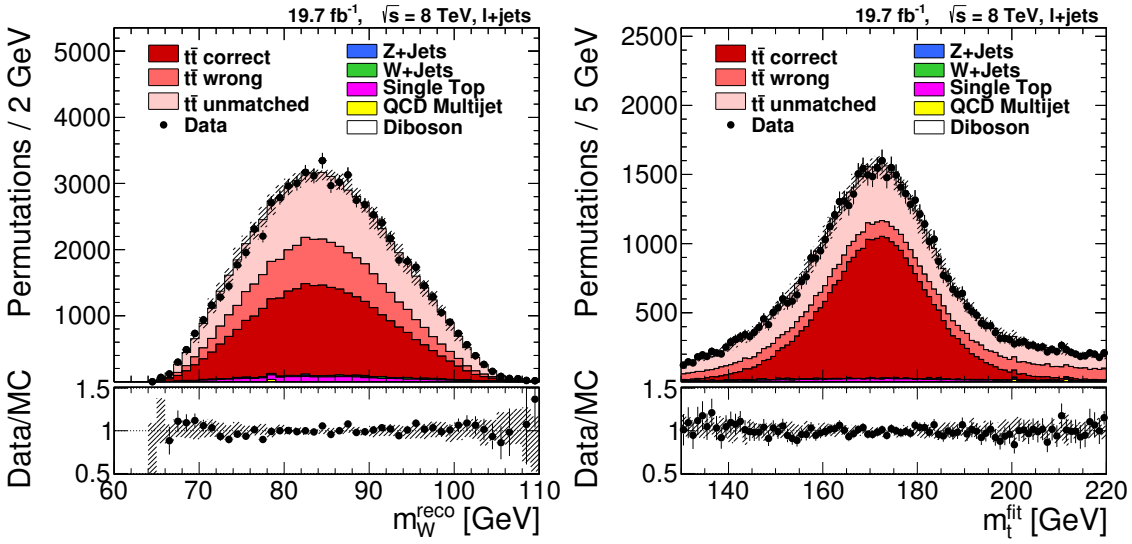

Figure 6.11.: Final selection: Reconstructed W-boson masses (left) and the fitted top-quark masses (right) with finer binning after the goodness-of-fit selection. The number of permutations in simulation is normalized to the number observed in data.

7. Methodology

Measurements of the top-quark mass are typically very sensitive to the uncertainties on the jet energy calibration if the invariant mass of several jets is used as main observable. In previous experiments and in the beginning of the CMS data taking, the jet energy scale was known only with an accuracy of about 2%. Scaling all jet energies by 2% directly yields a 2% uncertainty on the top-quark mass reconstructed from correct permutations (cf. Figure 6.8), which is about 3.3 GeV. Using the knowledge of the W-boson mass in a kinematic fit improves the precision. Still, a 2% shift in jet energy scale moves the mean of the fitted top-quark mass m_t^{fit} by 1.5 GeV. Therefore, the reconstructed mass of the W boson m_W^{reco} shall be used here to extract a *jet scale factor* (JSF). Both the top-quark mass and the JSF are then determined simultaneously with a joint-likelihood fit taking into account multiple permutations per event, called the *ideogram* method. This is called the "2D approach" in the following, in contrast to the "1D approach" that does not use the m_W^{reco} information. A "hybrid approach" is developed for maximum use of the information available.

7.1. Ideogram method

The ideogram method was previously employed for measuring the W-boson mass at the CERN LEP collider [147], and for a measurement of the top-quark mass at the Fermilab Tevatron collider [148, 149]. The method provides a convenient framework for including estimators from different sources, combining them into one likelihood. In this analysis, the likelihood in the ideogram method is built of analytic expressions that are obtained from simulation.

The probability distribution for measuring the top-quark mass m_t and the jet scale factor JSF given a data sample is given by Bayes' theorem:

$$P\left(m_t, \text{JSF}|\text{sample}\right) = P\left(\text{sample}|m_t, \text{JSF}\right) \times \frac{P\left(m_t\right) P\left(\text{JSF}\right)}{P\left(\text{sample}\right)}.$$

In the *2D approach*, no prior knowledge is included, and the second factor becomes a constant.

However, a useful prior knowledge is the jet energy scale as determined from γ/Z+jet events using the methods described in Section 5.4. By setting $P\,(\mathrm{JSF}) = \delta\,(1)$, the JSF is fixed to unity and the *1D approach* is obtained.

In a *hybrid approach*, the prior knowledge about the jet energy scale is incorporated by choosing $P\,(\mathrm{JSF})$ to be a Gaussian constraint with $\mathrm{JSF} = 1 \pm \sigma^{\mathrm{c}}_{\mathrm{JSF}}$. The relative weight of the 1D approach is then given by $w^{\mathrm{1D}}_{\mathrm{hyb}} = \left(\sigma^{\mathrm{2D}}_{\mathrm{JSF}}\right)^2 / \left\{\left(\sigma^{\mathrm{2D}}_{\mathrm{JSF}}\right)^2 + \left(\sigma^{\mathrm{c}}_{\mathrm{JSF}}\right)^2\right\}$, where $\sigma^{\mathrm{2D}}_{\mathrm{JSF}}$ is the statistical uncertainty obtained for the JSF in the unconstrained 2D approach and depends on the integrated luminosity of the analyzed sample. It is evident that the 2D approach is recovered in the limit $\sigma^{\mathrm{c}}_{\mathrm{JSF}} \to \infty$, and the 1D approach in the limit $\sigma^{\mathrm{c}}_{\mathrm{JSF}} \to 0$, so that the choice of $\sigma^{\mathrm{c}}_{\mathrm{JSF}}$ allows for extrapolating between both approaches. A default weight $w^{\mathrm{1D}}_{\mathrm{hyb}} = 0.5$ will be used for the hybrid approach in this thesis, giving equal weight to the JSF determined from W-boson decays and the jet energy scale from γ/Z+jet events.

Denoting an unnormalized probability a "likelihood", the likelihood used to estimate the top-quark mass m_t and the JSF from the data sample is defined as the product over all single-event likelihoods

$$\begin{aligned}
\mathcal{L}\,(m_t, \mathrm{JSF}|\mathrm{sample}) &= \mathcal{L}\,(\mathrm{sample}|m_t, \mathrm{JSF}) = \prod_{\mathrm{events}} \mathcal{L}\,(\mathrm{event}|m_t, \mathrm{JSF})^{C \times w_{\mathrm{event}}}\,, \\
w_{\mathrm{event}} &= \sum_{i=1}^{n} P_{\mathrm{gof}}\,(i)\,, \quad P_{\mathrm{gof}}\,(i) > 0.2, \\
C &= N_{\mathrm{events}} \,/\, \sum_{\mathrm{events}} w_{\mathrm{event}}
\end{aligned}$$

where i runs over the n selected permutations of an event, and C is a normalization constant. The event weight w_{event} is introduced to reduce the impact of events without correct permutations.

The ideogram method can easily be illustrated as shown in Figure 7.1. Single ideograms can include multiple distinct solutions for the event reconstruction, often visible by eye. The multiplication of the ideograms results in a very precise determination of the parameters under study.

The event likelihoods or *ideograms* are given by the sum of two-dimensional *permutation probability densities* P that encapsulate the dependency of the observed quantities m_t^{fit} and m_W^{reco} on the input top-quark mass and JSF obtained from simulated events. Each permutation is weighted by the goodness-of-fit probability P_{gof}

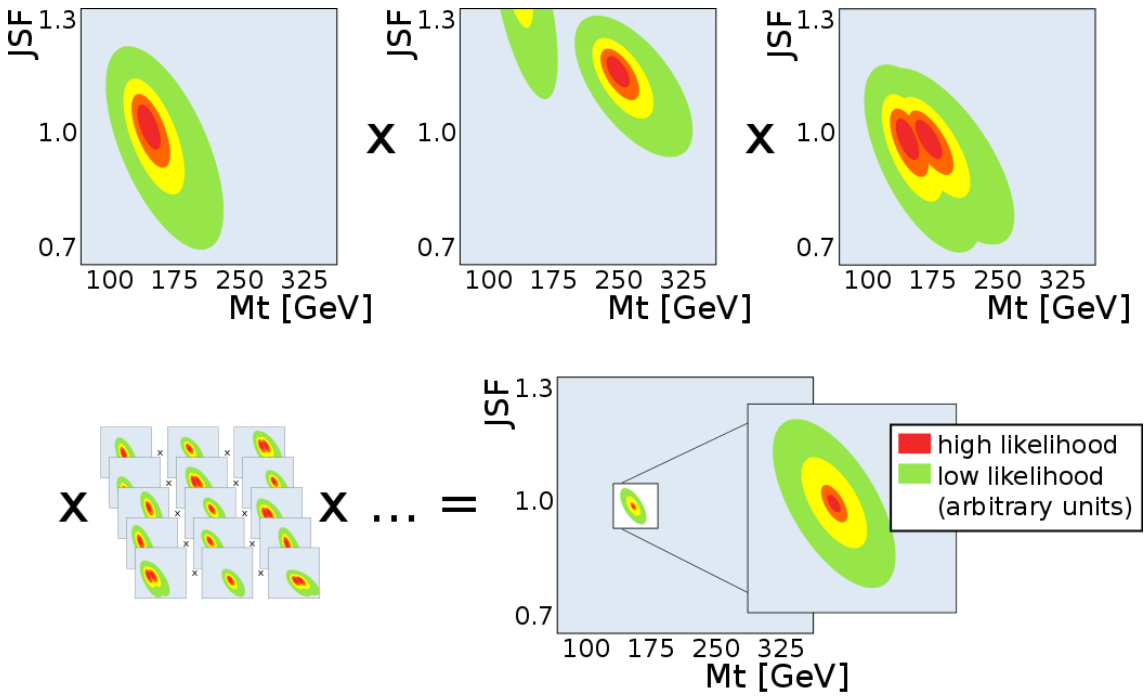

Figure 7.1.: Illustration of the ideogram method showing the multiplication of event likelihoods (*ideograms*) to form the sample likelihood $\mathcal{L}$. The color-coded z–axis has arbitrary units.

so that the expression reads

$$\mathcal{L}\left(\text{event}|m_t, \text{JSF}\right) \;=\; \sum_{i=1}^{n} P_{\text{gof}}\left(i\right) P\left(m_{t,i}^{\text{fit}}, m_{W,i}^{\text{reco}}|m_t, \text{JSF}\right).$$

The observables m_t^{fit} and m_W^{reco} have a low correlation coefficient of 0.02 as the light-quark jets in m_t^{fit} are constrained to an invariant mass of 80.4 GeV by the kinematic fit. Therefore, the permutation probability density P can be factorized into one-dimensional expressions and is given by the sum over the different permutation types:

$$P\Big(m_{t,i}^{\text{fit}}, m_{W,i}^{\text{reco}}|m_t, \text{JSF}\Big) \;=\; \sum_j f_j P_j\Big(m_{t,i}^{\text{fit}}|m_t, \text{JSF}\Big) \times P_j\Big(m_{W,i}^{\text{reco}}|m_t, \text{JSF}\Big),$$

where j represents the correct (cp), wrong (wp) and unmatched (un) permutation types, and f_j their relative fraction taken from the simulated sample with $m_{t,\text{gen}} = 172.5\,\text{GeV}$. Similarly, a background term could be included – without dependency on the true top-quark mass – but the background fraction of 3.7% can

safely be neglected at this stage of the analysis and left to the calibration of the method described in Section 7.2.

The relative fractions f_j and probability density distributions P_j are determined separately for the muon and electron channels from simulated $t\bar{t}$ events. These were generated for nine different top-quark mass values $m_{t,\mathrm{gen}}$ between 166.5 GeV and 178.5 GeV using the default MadGraph + Pythia 6 setup. The jet energies are scaled to obtain for all masses samples with different JSF values (0.96, 0.98, 1.00, 1.02, and 1.04).

The weighted m_t^{fit} distributions of correct permutations in the muon+jets channel are shown in Figure 7.2 for different values of $m_{t,\mathrm{gen}}$ and JSF. They can be fitted with a Voigt distribution (Breit-Wigner distribution convolved with a Gaussian resolution) characterized by the mean μ, the Gaussian resolution σ and the width Γ, where the latter is fixed to a value of 2 GeV, accounting for the natural top-quark width and non-Gaussian resolution effects:

$$\begin{aligned}
\text{Voigt: } V(x;\mu,\sigma,\Gamma) &= \int dx' BW(x';\mu,\Gamma) \times G(x;x',\sigma)\,, \\
\text{Breit-Wigner: } BW(x;\mu,\Gamma) &= \frac{1}{2\pi} \cdot \frac{\Gamma}{\left((x-\mu)^2 + \Gamma^2/4\right)}, \\
\text{Gauss: } G(x;\mu,\sigma) &= \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right).
\end{aligned}$$

The two free fit parameters are determined as a function of both $m_{t,\mathrm{gen}}$ and JSF by first fitting both parameters as a linear function of (JSF − 1) for fixed $m_{t,\mathrm{gen}}$, shown in Figure 7.3, where the dependence of both mean and resolution on the JSF is clearly visible. The constant and slope parameters of the first iteration of linear fits are then again fitted linearly as a function of $(m_{t,\mathrm{gen}} - 172.5\,\mathrm{GeV})$ as shown in Figure 7.4. As expected and needed for the analysis, the mean of the distribution is extremely sensitive to the value of the input top-quark mass. Also the total value of the resolution depends on $m_{t,\mathrm{gen}}$. The parameters extracted from the second set of linear fits determine the expected Voigt distribution of m_t^{fit} for correct permutations as a function of $m_{t,\mathrm{gen}}$ and JSF, and can be identified with the term $P_{cp}\left(m_{t,i}^{\mathrm{fit}} | m_t, \mathrm{JSF}\right)$ in the likelihood expression.

Similarly, the probability densities for the wrong and unmatched permutations (Figures 7.5 and 7.6) are parametrized using the asymmetric Crystal-Ball distribu-

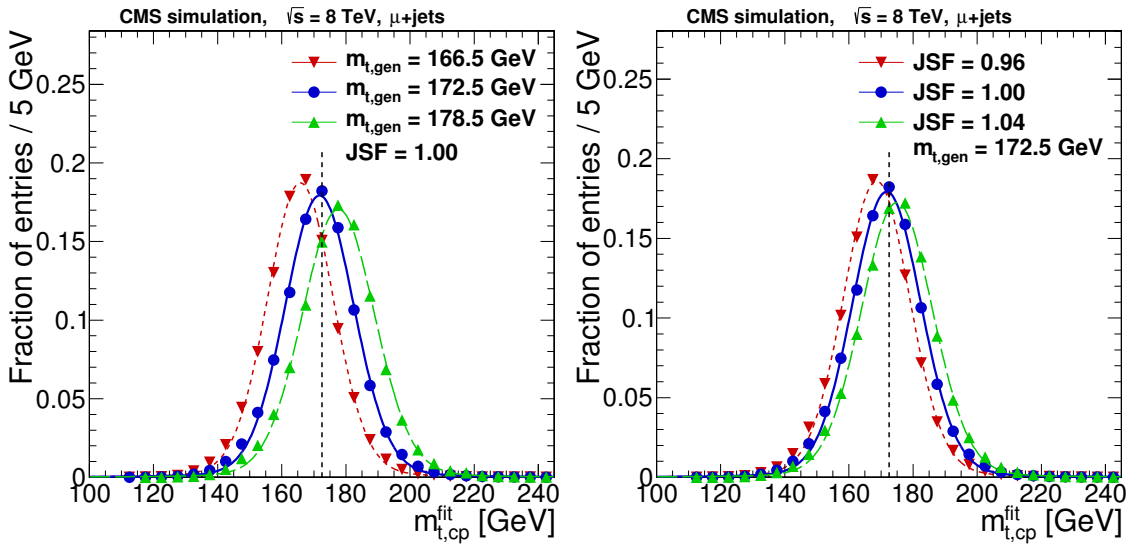

Figure 7.2.: Simulated m_t^{fit} distributions of correct $t\bar{t}$ permutations, for three generated masses $m_{t,\text{gen}}$ with JSF = 1 (left), and for three jet scale factors with $m_{t,\text{gen}} = 172.5\,\text{GeV}$ (right). The vertical dashed line corresponds to $m_t^{\text{fit}} = 172.5\,\text{GeV}$.

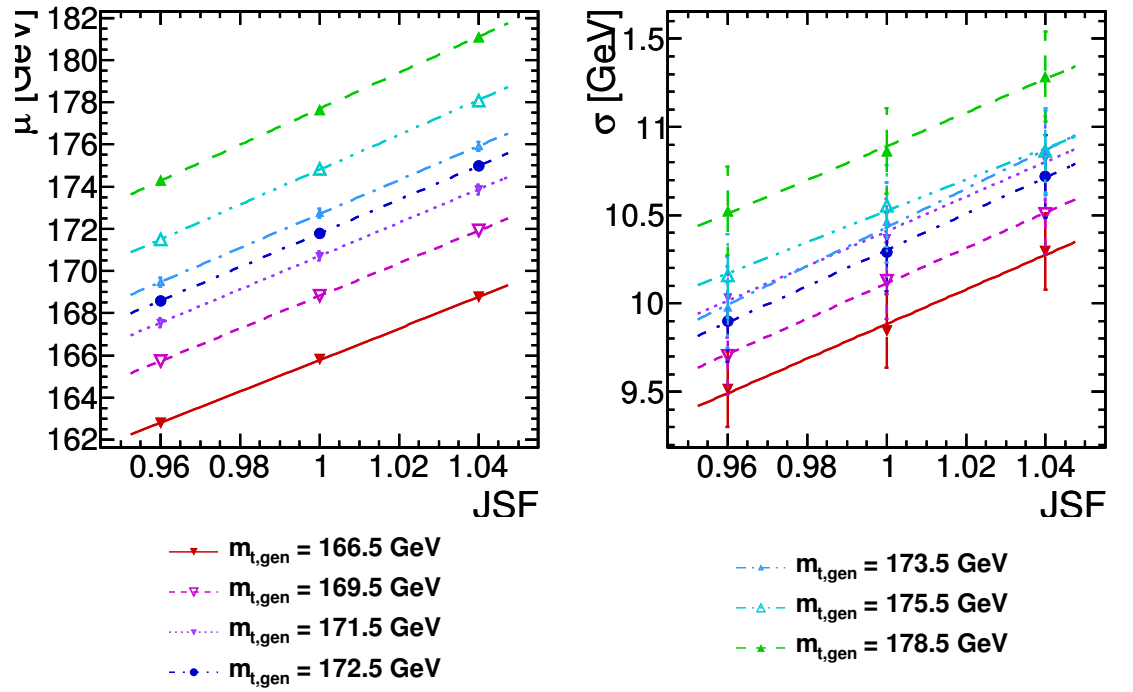

Figure 7.3.: The Gaussian mean μ (left) and resolution σ (right) of the m_t^{fit} distribution of correct permutations in dependency of the input top-quark mass and JSF.

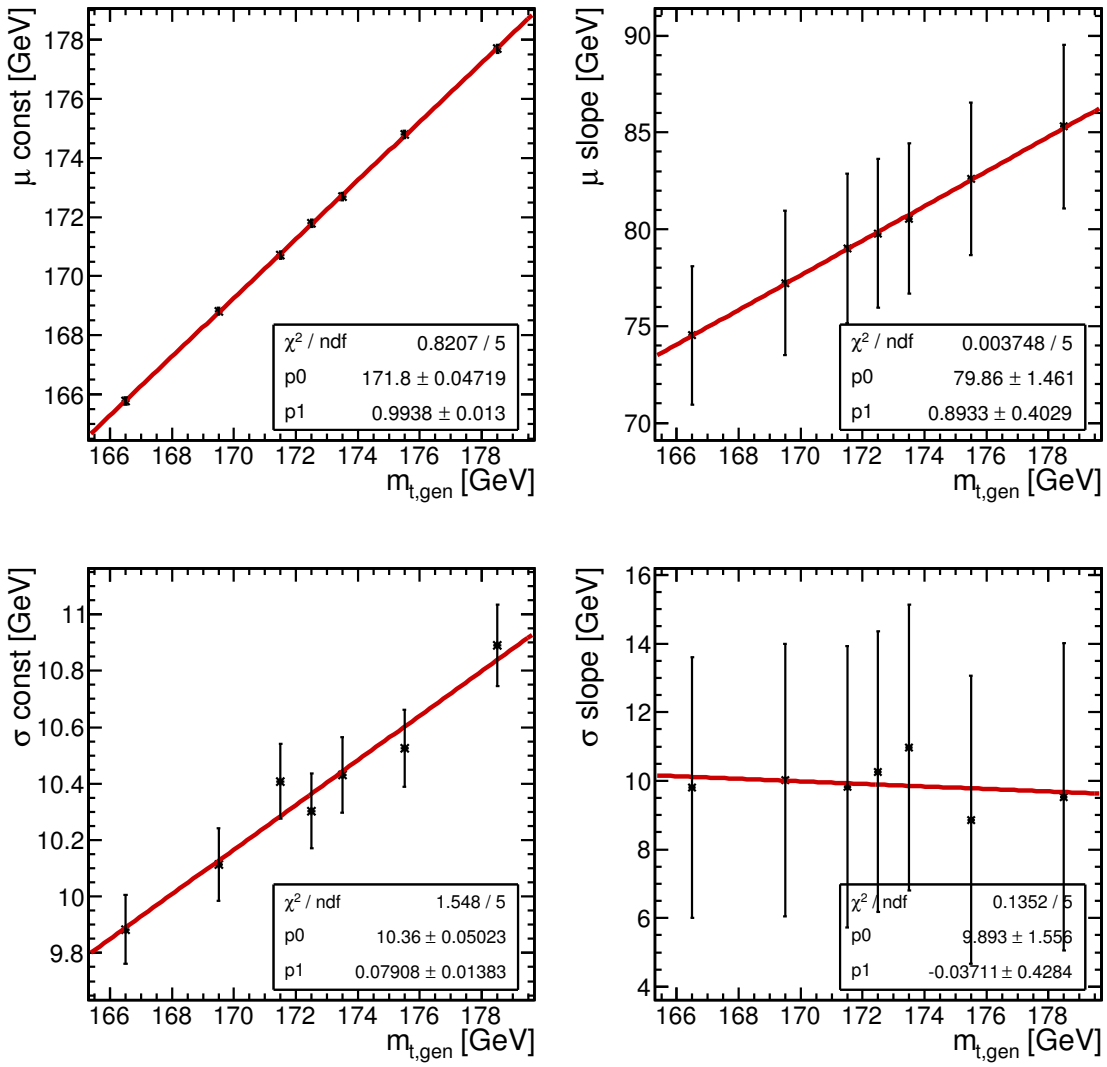

Figure 7.4.: The Gaussian mean μ (upper left) and resolution σ (lower left), and their respective dependency on JSF (upper/lower right), of the m_t^{fit} distribution of correct permutations in dependency of the input top-quark mass.

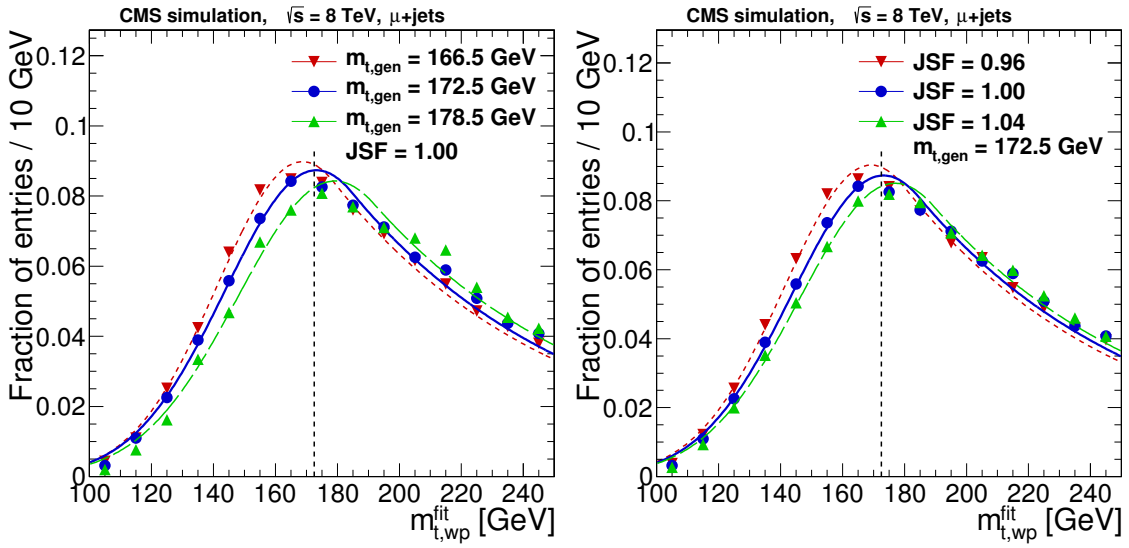

Figure 7.5.: Simulated m_t^{fit} distributions of wrong $t\bar{t}$ permutations, for three generated masses $m_{t,\text{gen}}$ with JSF = 1 (left), and for three jet scale factors with $m_{t,\text{gen}} = 172.5\,\text{GeV}$ (right). The vertical dashed line corresponds to $m_t^{\text{fit}} = 172.5\,\text{GeV}$.

tion defined as

$$C\left(x;\mu,\sigma,\alpha,n\right) = N\cdot\begin{cases}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & \text{for } \frac{x-\mu}{\sigma} > -\alpha\\ A\cdot\left(B-\frac{x-\mu}{\sigma}\right)^{-n}, & \text{for } \frac{x-\mu}{\sigma} \leqslant -\alpha\end{cases},$$
$$A = \left(\frac{n}{|\alpha|}\right)^n\cdot\exp\left(-\frac{|\alpha|^2}{2}\right),\ B = \frac{n}{|\alpha|}-|\alpha|\,,$$

where the Gaussian mean μ and resolution σ as well as the exponential threshold α are left free floating with the fixed exponents $n_{wp} = 15$ and $n_{un} = 3$.

The distribution of the reconstructed W-boson mass, shown in Figures 7.7 and 7.8 for different JSF, is distorted by the P_{gof} selection. It is found that it can be described well by an asymmetric Gauss distribution

$$\text{AG}\left(x;\mu,\sigma_1,\sigma_2\right) = \begin{cases}\frac{1}{\sigma_1\sqrt{2\pi}}\cdot\exp\left(\frac{(x-\mu)^2}{2\sigma_1^2}\right), & \text{for } \frac{x-\mu}{\sigma_1} < 0\\ \frac{1}{\sigma_2\sqrt{2\pi}}\cdot\exp\left(\frac{(x-\mu)^2}{2\sigma_2^2}\right), & \text{for } \frac{x-\mu}{\sigma_1} \geqslant 0\end{cases},$$

where the mean μ and the resolutions σ_1 and σ_2 are again parametrized in terms of $m_{t,\text{gen}}$ and JSF, with a negligible dependence on the top-quark mass.

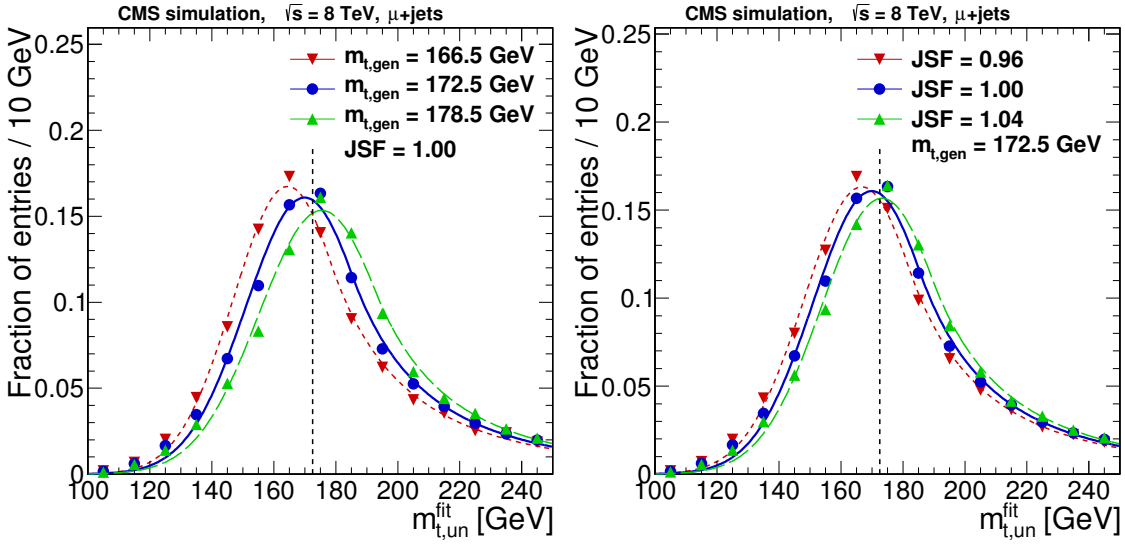

Figure 7.6.: Simulated m_t^{fit} distributions of unmatched $t\bar{t}$ permutations, for three generated masses $m_{t,\text{gen}}$ with JSF $= 1$ (left), and for three jet scale factors with $m_{t,\text{gen}} = 172.5\,\text{GeV}$ (right). The vertical dashed line corresponds to $m_t^{\text{fit}} = 172.5\,\text{GeV}$.

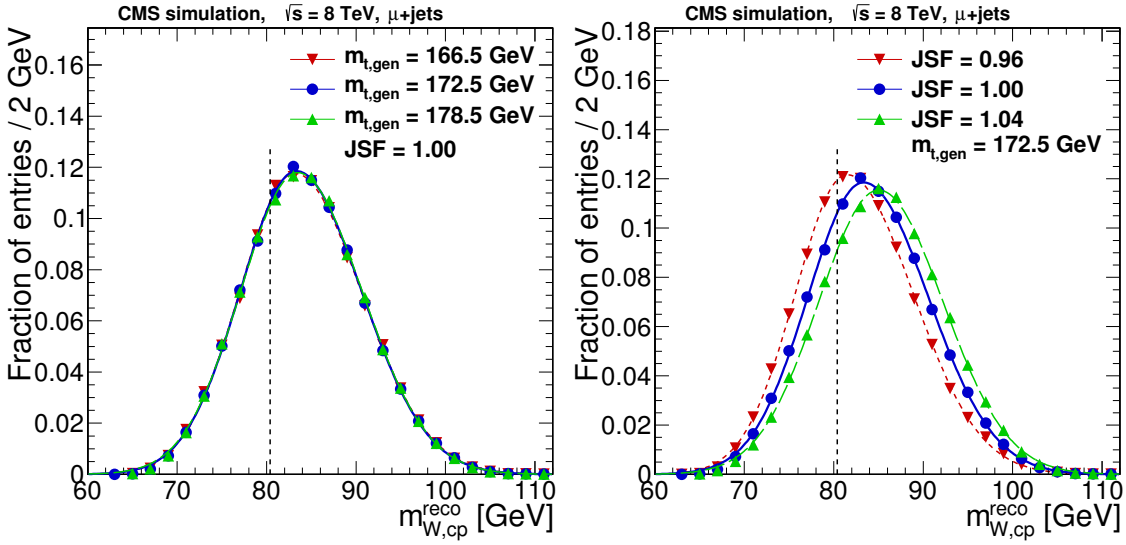

Figure 7.7.: Simulated m_W^{reco} distributions of correct $t\bar{t}$ permutations, for three generated masses $m_{t,\text{gen}}$ with JSF $= 1$ (left), and for three jet scale factors with $m_{t,\text{gen}} = 172.5\,\text{GeV}$ (right). The vertical dashed line indicates the known W-boson mass of $m_W = 80.4\,\text{GeV}$.

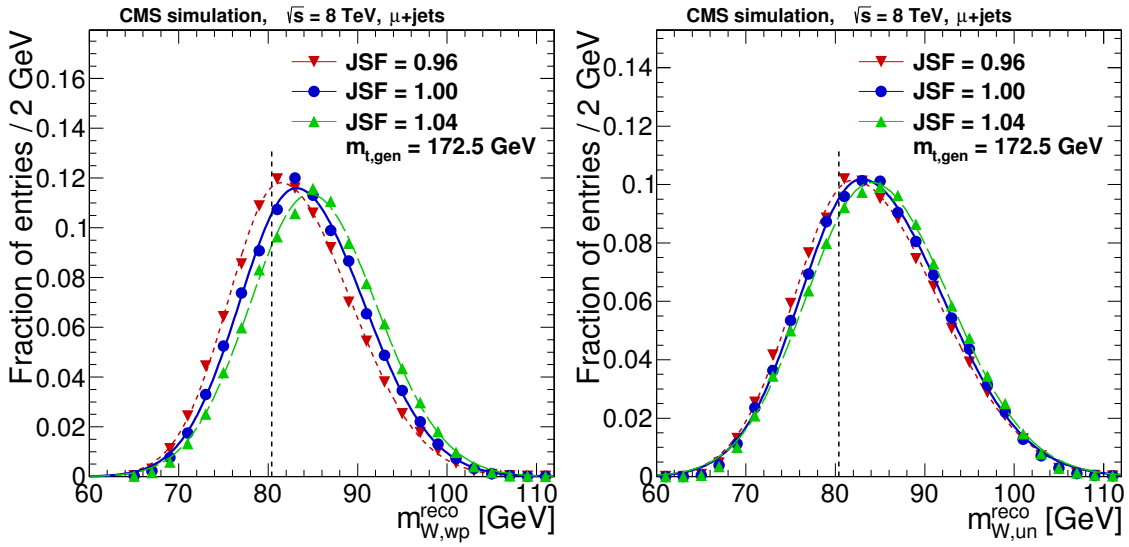

Figure 7.8.: Simulated m_W^{reco} distributions of wrong (left) and unmatched (right) $t\bar{t}$ permutations, for three jet scale factors with $m_{t,\text{gen}} = 172.5\,\text{GeV}$. The vertical dashed line indicates the known W-boson mass of $m_W = 80.4\,\text{GeV}$.

Separate probability densities are obtained for the muon+jets and electron+jets channels with very similar parameters. In order to combine the two channels, the likelihoods are multiplied: $\mathcal{L}_\ell = \mathcal{L}_\mu \times \mathcal{L}_e$. In the 2D approach, the most likely top-quark mass m_t^{2D} and JSF are obtained by maximizing the two-dimensional likelihood. For the 1D and hybrid approaches the corresponding JSF constraints are taken into account. Technically, a minimization of the negative logarithm of the likelihood is performed.* The statistical uncertainty is then given by the condition $-2\Delta \ln \mathcal{L}\,(m_t, \text{JSF}|\text{sample}) = 1$.

7.2. Calibration of the ideogram method

The presented ideogram method contains some simplifications, like the fixed values of f_j and the neglection of the background contribution. Therefore, the method has to be checked for possible biases and for the correct estimation of the statistical uncertainty. This calibration is performed for the 2D approach. The calibration is also valid for the 1D and hybrid approaches as these are special cases of the 2D approach.

*Using the `ROOT::Math::Minimizer` class.

For each combination of the nine $m_{t,\mathrm{gen}}$ values and three JSF scales (0.98, 1.00, 1.02), 10 000 pseudo-experiments are conducted, resulting in 27 calibration points in the (m_t, JSF) plane. The number of simulated events contained in each pseudo-experiment is given by a Poisson distribution around the number of events observed in data for the individual or the combined channel, respectively. All background processes are included in the pseudo-experiments. The following steps are done to draw events from a combined sample of N signal and background events that were normalized to luminosity and have some maximum MC event weight $\max(w_{\mathrm{MC}})$[†]:

1. Draw a random integer number in the interval $[0, N-1]$ to choose an event.
2. Draw a random number in the interval $[0, \max(w_{\mathrm{MC}})]$. Pick the event for the pseudo-experiment if its event weight is larger than the random number.
3. Increase the counter of drawn events by 1, or decrease if the event has a negative weight given by the MC@NLO method.
4. Repeat until the number of drawn events is equal to the desired number of events.

This prescription ensures that every pseudo-experiment is based on a correctly unweighted MC sub-sample that can be treated exactly like data by the ideogram method.

The distribution of the extracted values $m_{t,\mathrm{extr}}$ and $\mathrm{JSF}_{\mathrm{extr}}$ follows a Gaussian around the generated values, smeared by the statistical uncertainty of the method, as shown in Figure 7.9. The biases are defined as

$$\begin{aligned} \text{mass bias} &= \Big\langle m_{t,\mathrm{extr}} - m_{t,\mathrm{gen}} \Big\rangle, \\ \text{JSF bias} &= \Big\langle \mathrm{JSF}_{\mathrm{extr}} - \mathrm{JSF} \Big\rangle. \end{aligned}$$

The biases in mass and in JSF are shown in Figure 7.10 as a function of $m_{t,\mathrm{gen}}$ for the three values of JSF, together with linear fits for each value of JSF. A constant fit of the calibration points serves as a quality estimator of the overall calibration. The extracted values are found to be within 0.5 GeV and 0.5% of the input top-quark mass and JSF, respectively, and the biases are nearly independent of the input values.

[†]Weight for simulated events containing luminosity and cross section information, pileup, trigger and b-tag scale factors. Not to be confused with the analysis weight based on P_{gof}.

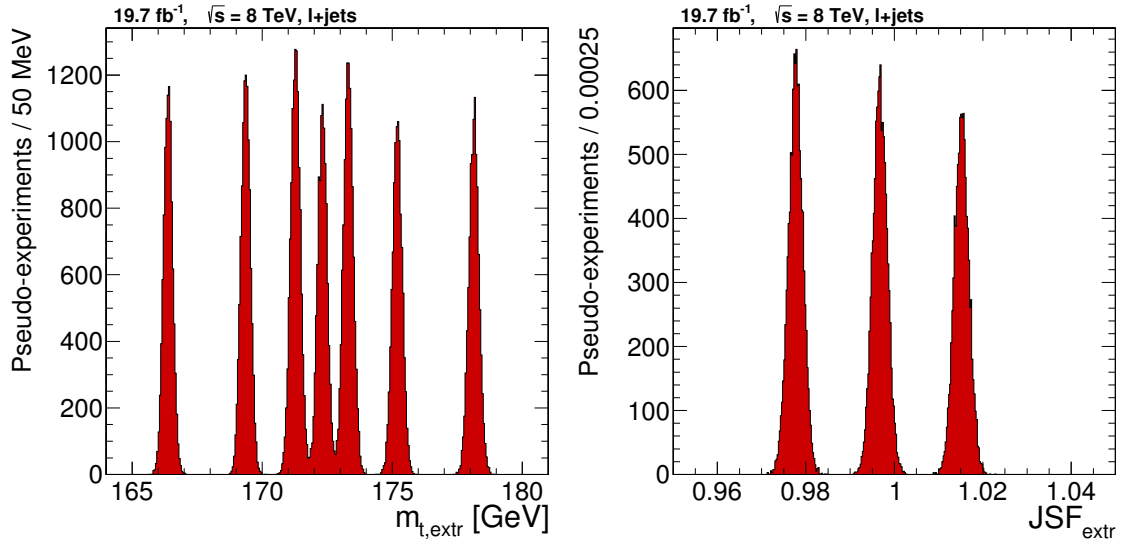

Figure 7.9.: Extracted top-quark masses $m_{t,\mathrm{extr}}$ for different $m_{t,\mathrm{gen}}$ and JSF = 1 (left), and extracted jet scale factors $\mathrm{JSF}_{\mathrm{extr}}$ for different JSF and $m_{t,\mathrm{gen}} = 172.5\,\mathrm{GeV}$ (right), from pseudo-experiments in the combined lepton+jets channel *before* the calibration.

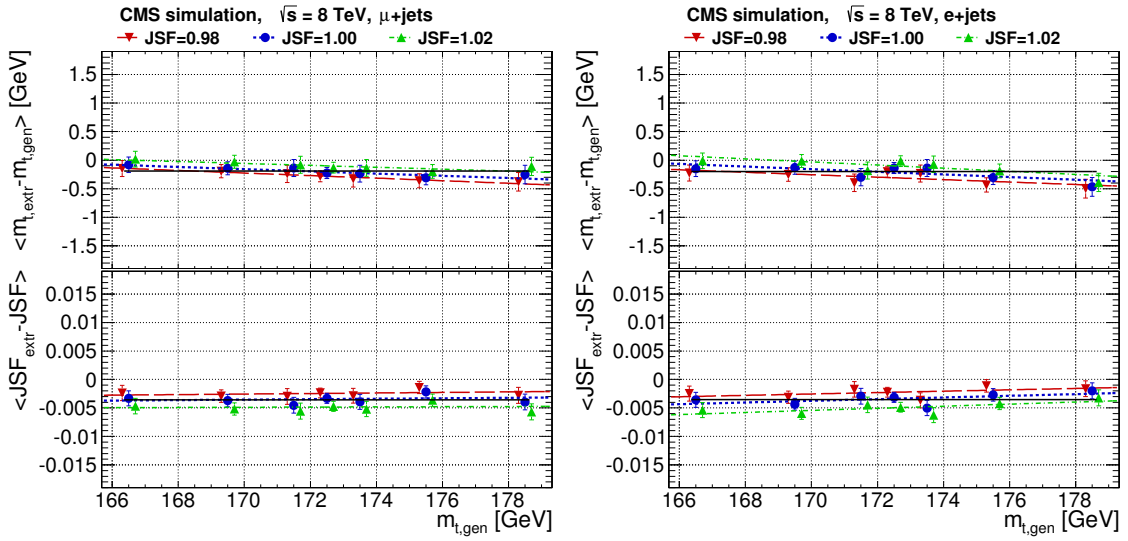

Figure 7.10.: Mean difference between the extracted mass $m_{t,\mathrm{extr}}$ and each generated mass $m_{t,\mathrm{gen}}$ and between $\mathrm{JSF}_{\mathrm{extr}}$ and $\mathrm{JSF}_{\mathrm{gen}}$ for the muon+jets (left) and the electron+jets channel (right), *before* the calibration, as a function of different $m_{t,\mathrm{gen}}$ and three values of JSF. The colored dashed lines correspond to straight line fits. The black solid line corresponds to an assumption of a constant calibration for all mass and JSF points in each channel.

Corrections for the extracted values are obtained by fitting the biases with a 2-dimensional function

$$\begin{aligned} f\left(m_{t,\mathrm{extr}}, \mathrm{JSF}_{\mathrm{extr}}\right) &= p_0 + p_1 \times \left(m_{t,\mathrm{extr}} - 172.5\right) + p_2 \times \left(\mathrm{JSF}_{\mathrm{extr}} - 1\right) \\ &\quad + p_3 \times \left(m_{t,\mathrm{extr}} - 172.5\right) \times \left(\mathrm{JSF}_{\mathrm{extr}} - 1\right), \end{aligned}$$

where the p_i are the calibration parameters. Applying this calibration to the extracted values yields the final calibrated top-quark mass $m_{t,\mathrm{cal}}$ and the jet scale factor $\mathrm{JSF}_{\mathrm{cal}}$ that agree well with the input values leading to biases close to zero, as shown in Figure 7.11.

For validating the statistical uncertainty returned by the method, the *pull* distribution for the top-quark mass is defined as:

$$\mathrm{pull} = \frac{m_{t,\mathrm{cal}} - m_{t,\mathrm{gen}}}{\sigma\left(m_{t,\mathrm{cal}}\right)},$$

where $\sigma\left(m_{t,\mathrm{cal}}\right)$ is the statistical uncertainty of an individual $m_{t,\mathrm{cal}}$ for a pseudo-experiment generated at $m_{t,\mathrm{gen}}$, and accordingly for the JSF. Using the calibrated pseudo-experiments, a Gaussian function is fitted to the pull distributions, as for instance shown in Figure 7.12. A pull width smaller/larger than unity means that the mean statistical uncertainty returned by the extraction method is larger/smaller than the variance of the bias which is the true statistical uncertainty. As depicted in Fig. 7.13, mass pull widths of 1.042 for the muon+jets channel and 1.030 for the electron+jets channel are found. The method therefore slightly underestimates the statistical uncertainty of the measurement, and the required corrections are incorporated into the evaluation of the final likelihood by scaling $\ln \mathcal{L}$ with a factor $(\text{pull width})^{-2}$.

After applying the full single-channel calibration for biases and the pull width, additional pseudo-experiments are generated for the combination of the two channels. As shown in Fig. 7.14, no additional corrections are needed.

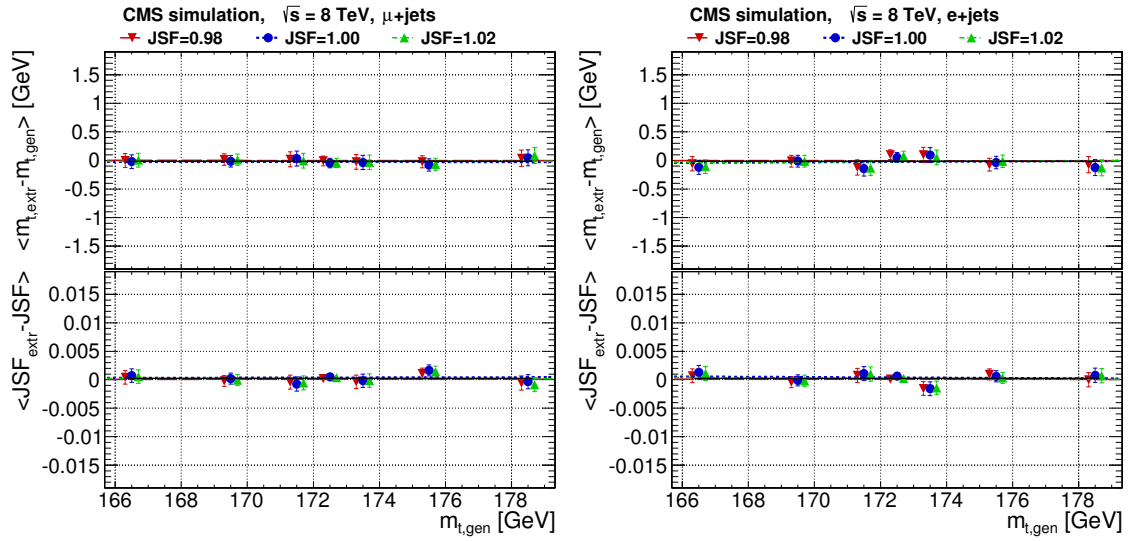

Figure 7.11.: Mean difference between the extracted mass $m_{t,\mathrm{extr}}$ and each generated mass $m_{t,\mathrm{gen}}$ and between $\mathrm{JSF_{extr}}$ and $\mathrm{JSF_{gen}}$ for the muon+jets (left) and the electron+jets channel (right), *after* the calibration, as a function of different $m_{t,\mathrm{gen}}$ and three values of JSF.

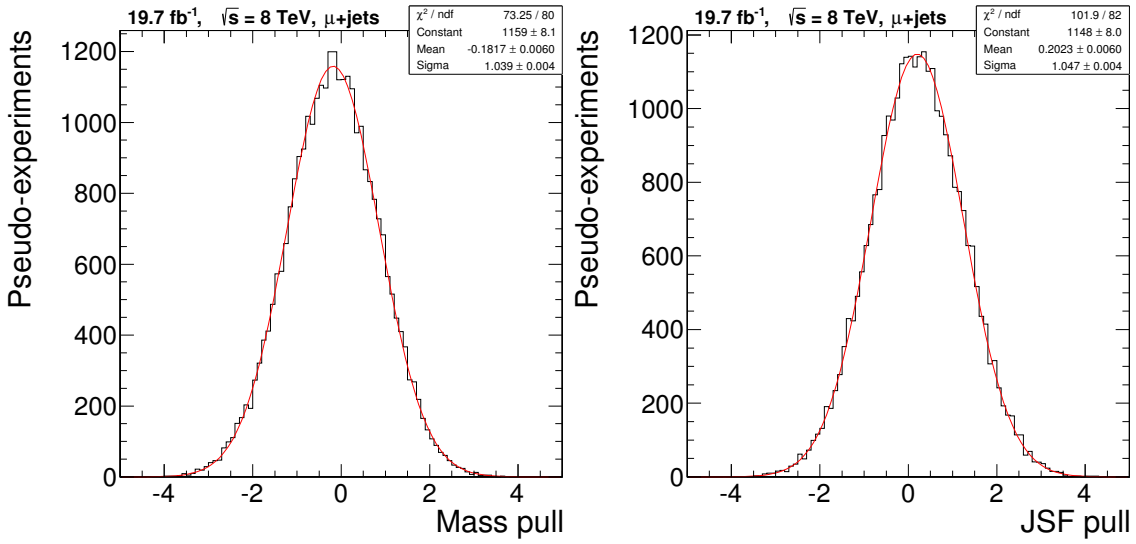

Figure 7.12.: Mass (left) and JSF (right) pull distributions in the muon+jets channel for $m_{t,\mathrm{gen}} = 172.5\,\mathrm{GeV}$ and $\mathrm{JSF} = 1$.

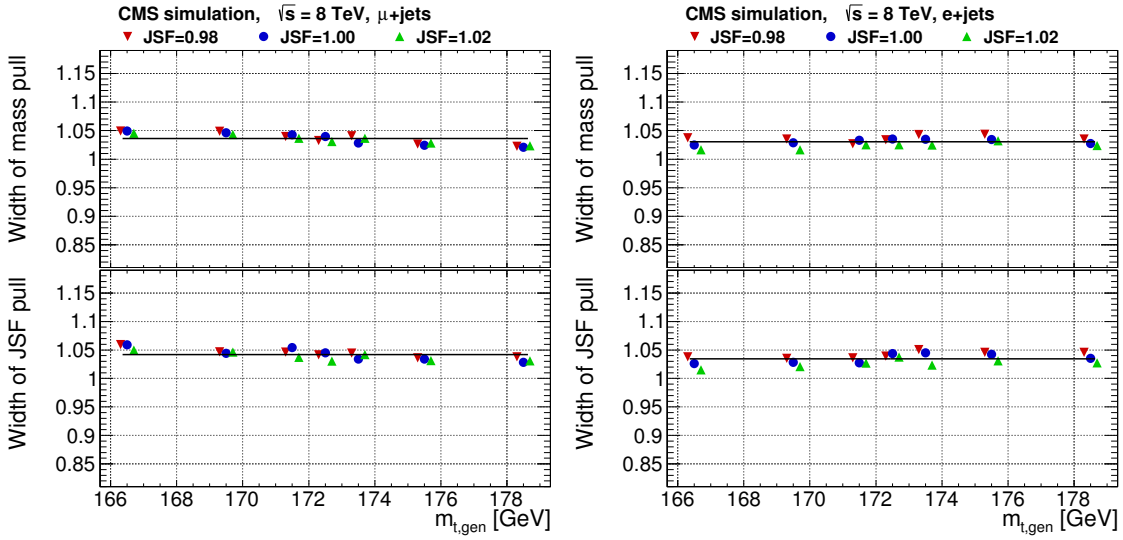

Figure 7.13.: Width of the pull distribution for the calibrated measurement of m_t and JSF as a function of different $m_{\mathrm{t,gen}}$ and three values of JSF in the muon+jets (left) and the electron+jets channel (right). The black solid lines correspond to fits of constants to all calibration points, assuming no dependence on m_t or JSF.

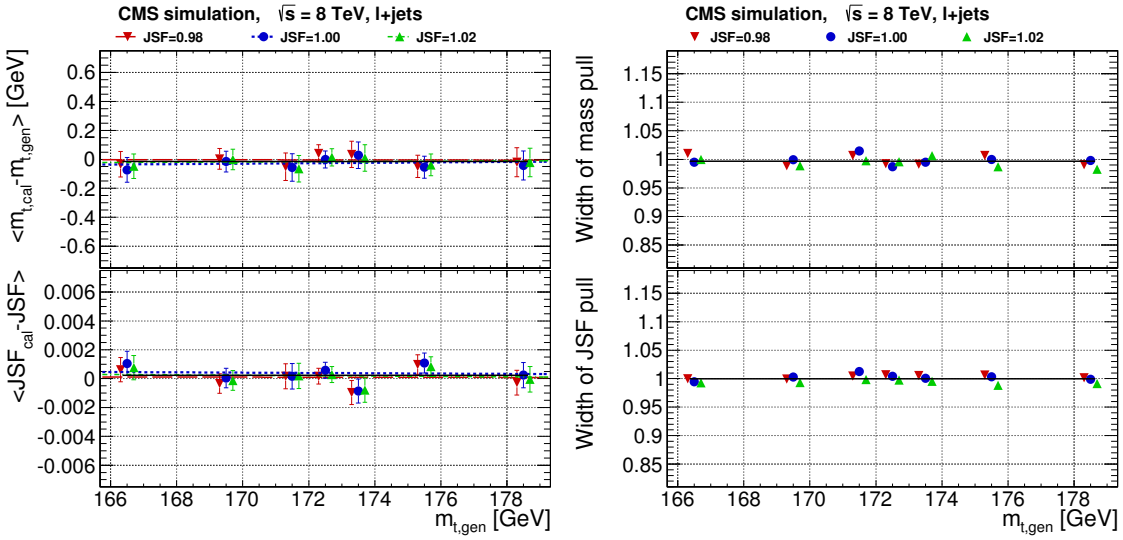

Figure 7.14.: Mean difference between the calibrated and generated values of m_t and JSF as a function of different $m_{t,\mathrm{gen}}$ and three values of JSF for combined lepton+jets events (left). Note the different y–scales compared to the previous figures; width of the pull distributions for the combined channel after the single-channel calibration (right).

8. Systematic uncertainties

The systematic uncertainties considered as relevant for this measurement, and the methods used to evaluate them are described in this chapter. The contributions from the different sources of systematic uncertainties are shown in Table 8.1, for the combined fit to the entire lepton+jets data set using the 2D, 1D, and hybrid (using $w_{\text{hyb}}^{\text{1D}} = 0.5$) approaches.

In general, the absolute values of the largest observed shifts in m_t and JSF, determined from changing the parameters of interest by ± 1 standard deviations in simulated pseudo-experiments, are assigned as systematic uncertainties on the final measurement.

In order to avoid accidentally small uncertainties from samples with limited size, the statistical significance is calculated for each comparison between uncorrelated samples. The statistical uncertainties on the mass and JSF extracted from pseudo-experiments based on a simulated sample containing N_{sample} weighted events is given by

$$\sigma_{\text{sample}} = \overline{\sigma}_{\text{PE}} \times \sqrt{\frac{\overline{N}_{\text{PE}}}{N_{\text{sample}}} \times \text{mean}\left(w_{\text{MC,norm}}^2\right)},$$

where $\overline{\sigma}_{\text{PE}}$ and $\overline{N}_{\text{PE}}$ are the mean uncertainty and number of events used for conducting the pseudo-experiments corresponding to $19.7\,\text{fb}^{-1}$. The mean of the squared normalized MC event weight $w_{\text{MC,norm}}$ accounts for the degradation of statistical precision stemming from the usage of weights for b-tagging, trigger and pileup. The statistical uncertainty on the observed difference in mass or JSF is used as systematic uncertainty if it is larger than the shift itself.

8.1. Experimental uncertainties

Method calibration: The method is developed and calibrated on simulated samples. This uncertainty accounts for the limited size of the calibration samples and the residual biases after calibration.

Table 8.1.: List of systematic uncertainties for the combined fit to the entire lepton+jets data set. Uncertainties evaluated from uncorrelated samples are determined with limited statistical precision. The numbers in brackets are not added to the total uncertainty. A minus-sign (–) in the $\delta {m_t}^{\mathrm{1D}}$ column indicates that the uncertainty is anti-correlated to the corresponding one on ${m_t}^{\mathrm{2D}}$. The "JEC: Flavor" subcomponents are added linearly, therefore the sign is given for the up variation of each source. The hybrid approach is performed with $w_{\mathrm{hyb}}^{\mathrm{1D}} = 0.5$ and effectively takes the average of the 1D and 2D approaches.

	2D approach		1D approach	Hybrid
	$\delta {m_t}^{\mathrm{2D}}$ (GeV)	δJSF	$\delta {m_t}^{\mathrm{1D}}$ (GeV)	$\delta {m_t}^{\mathrm{hyb}}$ (GeV)
Experimental uncertainties				
Method calibration	0.04	0.001	0.04	0.04
Jet energy corrections (quad. sum)	(0.16)	(0.006)	(–0.42)	(0.19)
– JEC: InterCalibration	<0.01	(<0.001)	–0.02	0.01
– JEC: MPFInSitu	0.01	(0.003)	–0.24	0.12
– JEC: Uncorrelated non-pileup	0.11	(0.005)	–0.32	0.13
– JEC: Uncorrelated pileup	0.12	(0.002)	–0.12	0.06
Lepton energy scale	0.01	<0.001	0.01	0.01
MET	0.04	<0.001	0.03	0.04
Jet energy resolution	0.11	0.002	–0.05	0.03
b tagging	0.06	<0.001	0.04	0.06
Pileup	0.12	0.002	–0.05	0.04
Non-$t\bar{t}$ background	0.05	<0.001	0.01	0.03
Modeling of hadronization				
JEC: Flavor (linear sum)	(0.40)	(0.001)	(0.30)	(0.35)
– light quarks (uds)	+0.11	–0.002	–0.02	+0.05
– charm	+0.03	<0.001	–0.01	+0.01
– bottom	–0.32	<0.001	–0.31	–0.32
– gluon	–0.22	+0.003	+0.05	–0.08
b-jet modeling (quad. sum)	(0.17)	(0.001)	(0.17)	(0.16)
– b fragmentation	0.06	0.001	–0.06	<0.01
– Semi-leptonic B hadron decays	0.16	<0.001	0.15	0.16
Modeling of perturbative QCD				
PDF	0.09	0.001	0.06	0.04
Ren. and fact. scale	0.17±0.08	0.004±0.001	–0.24±0.06	0.09±0.07
ME-PS matching threshold	0.11±0.09	0.002±0.001	–0.07±0.06	0.03±0.07
ME generator	0.07±0.11	0.001±0.001	0.16±0.07	0.12±0.08
Top-quark transverse momentum	0.16	0.003	–0.11	0.02
Modeling of soft QCD				
Underlying event	0.15±0.15	0.002±0.001	0.07±0.09	0.08±0.11
Color reconnection modeling	0.11±0.13	0.002±0.001	–0.09±0.08	0.01±0.09
Total systematic	0.61	0.007	0.65	0.49
Statistical (expected)	0.20	0.002	0.12	0.16
Total (expected)	0.64	0.007	0.66	0.52

Jet energy corrections: The uncertainty on the jet energy corrections is evaluated by scaling the energies of all jets up and down according to their individual data/MC uncertainties [118] (cf. Subsection 5.4), that are given by 23 individual sources and depend on both p_{T} and η of the jet. The uncertainty sources are mapped to the correlation groups defined by the top LHC working group [150]. The "JEC: Intercalibration" and "JEC: MPFInSitu" groups contain the modeling uncertainties of the L2 and L3 residual corrections, respectively, that are partially correlated between different experiments. The remaining flavor-independent JEC uncertainties, e.g., the statistical uncertainties of the residual corrections and detector-specific effects, are uncorrelated between experiments. This "JEC: Uncorrelated" group also contains uncertainties due to the effect of additional pileup events. As the pileup component has a specifically large p_{T}-dependence, it is listed separately here. The largest difference in the measured top-quark mass is taken as a systematic uncertainty. The corresponding shifts on the JSF in the 2D approach are given for information only, as the JSF is measured for the specific flavor composition and p_{T} spectrum of the selected $t\bar{t}$ sample.

Lepton energy scale: The muon [103] energies in simulation are scaled up and down by their uncertainty of 0.2%. The electron [102] energies are known with uncertainties of 0.6% in the barrel and 1.5% in the endcap region.

Missing transverse momentum: All changes in jet and lepton energies are propagated to the MET. For the remaining MET uncertainty evaluated here, the energy of particle-flow candidates not clustered into a jet with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ is scaled up and down by 10% [151].

Jet energy resolution: The jet energy resolution in simulation is degraded by 8 to 25% depending on η to match the resolutions measured in data shown in Subsection 5.4. In order to account for the resolution uncertainty, the jet energy resolution in the simulation is modified by ± 1 standard deviations with respect to the degraded resolution [123].

b tagging: The events are weighted to account for the p_{T}-dependent uncertainty of the b-tag efficiencies and mis-identification rates of the CSVM tagger [125] shown in Subsection 5.4. As an alternative approach, the discriminator value is changed from the nominal CSVM working point of 0.679 to 0.655 or 0.71, respectively. The b-tag efficiency in the nominal $t\bar{t}$ sample is altered by 1.2%

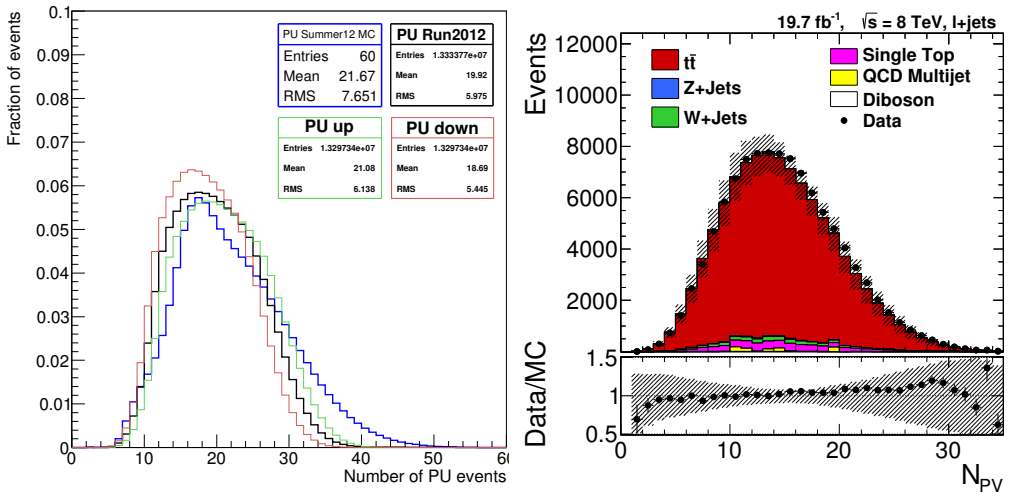

Figure 8.1.: Number of pileup events in Summer12 simulation and data with up/down uncertainties (left), and number of reconstructed primary vertices after baseline selection (right). The hatched area indicates the uncertainties from signal/background normalization and pileup.

and the mis-tag efficiency by 15%, both corresponding to a 1σ variation of the scale factor. The resulting variation of the c-tag efficiency is 5%, corresponding to two standard deviations. As the second method results in a larger shift on the result, this is quoted as a systematic uncertainty.

Pileup: The number of pp interactions per bunch crossing is given by the instantaneous luminosity multiplied with the total inelastic pp cross section. The inelastic pp cross section was determined from the number of reconstructed vertices in $Z \to \mu^+\mu^-$ events as 69.4 ± 4.2 mb. The uncertainty of that measurement is propagated to the expected number of pp interactions in simulation. Figure 8.1 shows the number of pileup events in simulation and the number of reconstructed vertices N_{PV} in both data and simulation. The influence of pileup on the energy measurement of low-p_{T} jets is accounted for by JEC uncertainties (see above).

Non-$t\bar{t}$ background: The normalizations of the background samples are varied by their inclusive cross-section uncertainties obtained from scale variations in the theory calculation, i.e., 20% for W/Z+jets (uncertainty for $W/Z + b\bar{b}$ production), 10% for single top, and 5% for diboson background. The amount of QCD multijet background is varied by 100%. For the dominating background,

single-top production, in addition the renormalization and factorization scales in the simulation are varied by factors of 1/2 and 2 to take into account the uncertainties on the extrapolation from the inclusive to the selected phase space. For the second-largest background, W+jets production, no fully simulated systematics samples were available. It is found that more than 90% of the selected W+jets events stem from the W+4 partons subprocess at matrix-element level, that gives a contribution of 0.7% to the total cross section. Therefore, the contribution of the $W + 4$ partons subprocess was determined in particle-level samples with variations of renormalization/factorization scales and matching threshold. Scale factors of 0.56 (scale up), 2.24 (scale down), 0.61 (matching up), and 2.11 (matching down) were determined. Applying these scale factors as additional uncertainties to the W+jets contribution gives a more realistic uncertainty estimate in the selected phase space. Shape uncertainties are still neglected but the impact of the matching threshold variations is likely to be overestimated as reducing/enlarging the phase space for matrix-element emissions is enlarging/reducing the phase space for corresponding parton shower emissions.

8.2. Modeling of hadronization

Similar tunings of string and cluster fragmentation models to e^+e^- data may still lead to different predictions in top-quark events [88]. The comparison of the reference implementations in Pythia and Herwig is non-trivial due to further differences in the parton shower, ME-PS matching and the underlying event, that may add up or compensate each other, as discussed in Section 10.1.

In this analysis, the impact of switching between Pythia and Herwig is important for the jet energy corrections, because the detector response depends on the momenta and types of stable particles. In addition, the impact of b-fragmentation hardness and of the neutrino fraction in b jets is evaluated using weights calculated on particle level.

Flavor-dependent JES: The difference in jet energy response between Pythia 6 and Herwig++ is determined for each jet flavor with respect to the nominal calibration point, as described in Subsection 5.4. In order to evaluate possible differences between the measured JSF (from light quarks with gluon contamination) and the b-JES, the flavor uncertainties for jets stemming from light

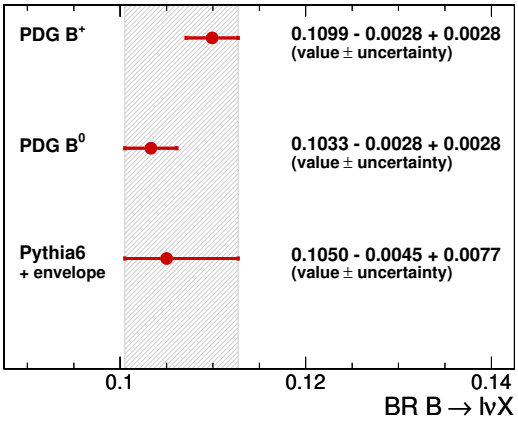

Figure 8.2.: Semi-leptonic branching ratio of B hadrons and the corresponding uncertainty.

quarks (uds), charm quarks, bottom quarks, and gluons are evaluated separately and added linearly. In the framework defined by the top LHC working group, this is called the "JEC: Flavor" correlation group and is assumed to be largely correlated between the experiments. The jet energy response in $Z + b$ events was found to agree with the expectation from MadGraph + Pythia 6 with similar uncertainties [122].

b fragmentation: The Bowler-Lund fragmentation function for B hadrons is retuned to agree with the x_B data measured by the ALEPH [152] and DELPHI [153] collaborations, see Sec. 10.3. The difference between Pythia Tune Z2* and the retuning is taken as the systematic uncertainty, as it is larger than the uncertainties on the retuning.

Semi-leptonic B hadron decays: The semi-leptonic branching ratio of B hadrons corresponds directly to the abundance of undetected neutrinos inside b jets that lower the response with respect to the original bottom quark. The branching ratio $B \to \ell\nu X$ (where ℓ is defined in the literature as either muon or electron) is 0.105 for B^0/B^+ decays in Pythia 6. The branching ratio is varied by -0.45% and $+0.77\%$ to give an envelope of the existing measurements and their uncertainties [87], shown in Figure 8.2.

8.3. Modeling of the hard scattering process

Parton distribution functions: The simulated signal events are generated using the CTEQ6L1 parton distribution functions (PDFs) [154]. The uncertainty due to the choice of PDF is calculated using the PDF4LHC prescription [155], creating an envelope of the CT10 [156], MSTW2008 [63], and NNPDF 2.3 [157] PDFs, including their individual PDF and α_s uncertainties. In total, the mass and JSF shifts for 147 PDF variations are evaluated so that the speed of the ideogram method is crucial. The maximum difference between CTEQ6L1 and the lower or upper boundary of the envelope, respectively, is taken as a systematic uncertainty. The α_s uncertainty is not propagated to the matrix element generator, as the Q^2 scale is varied separately.

Renormalization and factorization scales: The amount of QCD radiation in $t\bar{t}$ events affects top-quark reconstruction and selection efficiencies. Theory-inspired scale variations by factors of $1/2$ and 2 are found to be generous envelopes of jet multiplicity and gap fraction measurements in $t\bar{t}$ events performed by ATLAS [158,159] and CMS [160,161], shown in Figures 8.3 and 8.4. Final-state radiation inside resonance decays is described at NLO accuracy and is tightly constrained by measurements of event shapes at LEP [162–165]. A measurement of jet shapes in $t\bar{t}$ events has been performed by the ATLAS collaboration [166], where the results can be compared to simulation in the Rivet framework, as done for Figure 8.5.

ME-PS matching threshold: In the $t\bar{t}$ simulation, the QCUT matching thresholds used for interfacing the matrix-elements prediction generated with MadGraph and the Pythia parton showering are changed from the default value of 40 GeV down to 30 GeV and up to 60 GeV. These variations cover the parameter range in which the p_{T} distributions of the first three additional jets match smoothly on both sides of the threshold. Figures 8.3 and 8.4 show only small impact on jet observables.

ME generator: The guidelines of the Top LHC working group [167] recommend the comparison of central predictions from different MC generators, using at least one multi-leg and one NLO generator setup. Therefore, the difference between simulating the hard scattering process with MadGraph (LO multi-leg, $t\bar{t}$ with up to 3 jets) and Powheg (NLO $t\bar{t}$) is evaluated. Both matrix-element generators are interfaced to Pythia 6.

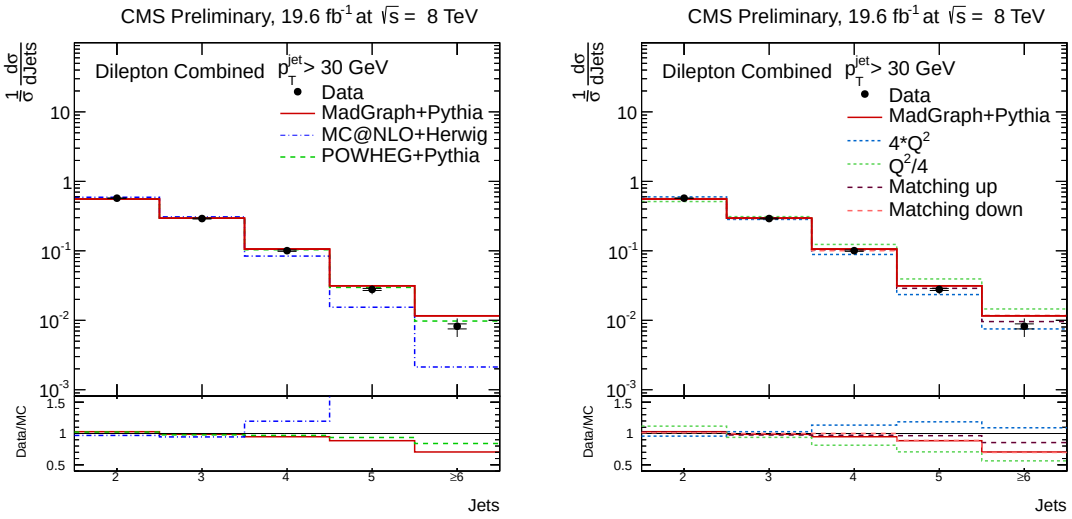

Figure 8.3.: Multiplicity of jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ in $t\bar{t}$ dilepton events comparing CMS data to different generator setups (left) and variations of the MadGraph simulation (right). Plots from [161].

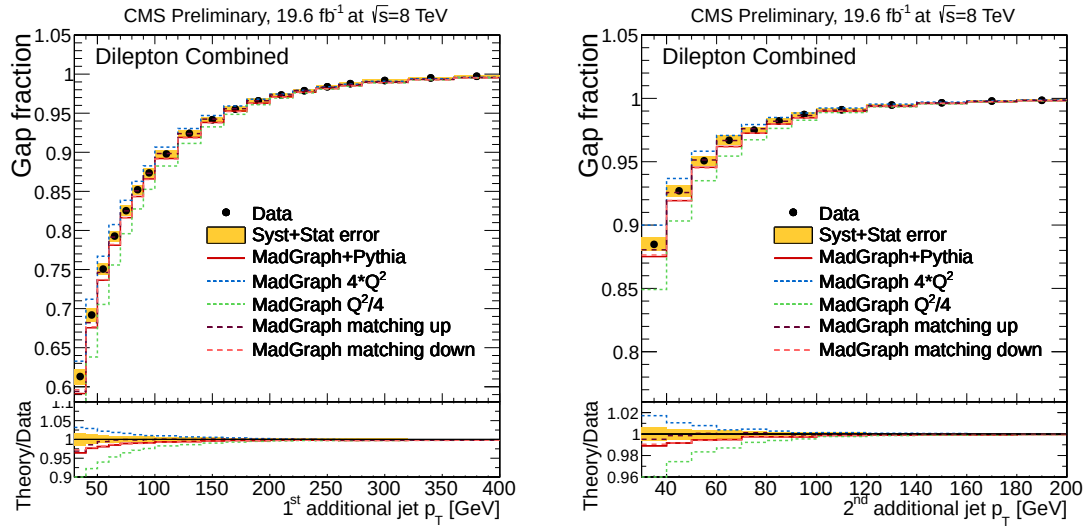

Figure 8.4.: The gap fraction defined as the fraction of events with no additional jet above a given p_{T} threshold, shown for the first (left) and the second (right) additional jet in $t\bar{t}$ dilepton events, comparing CMS data and MadGraph. Plots from [161].

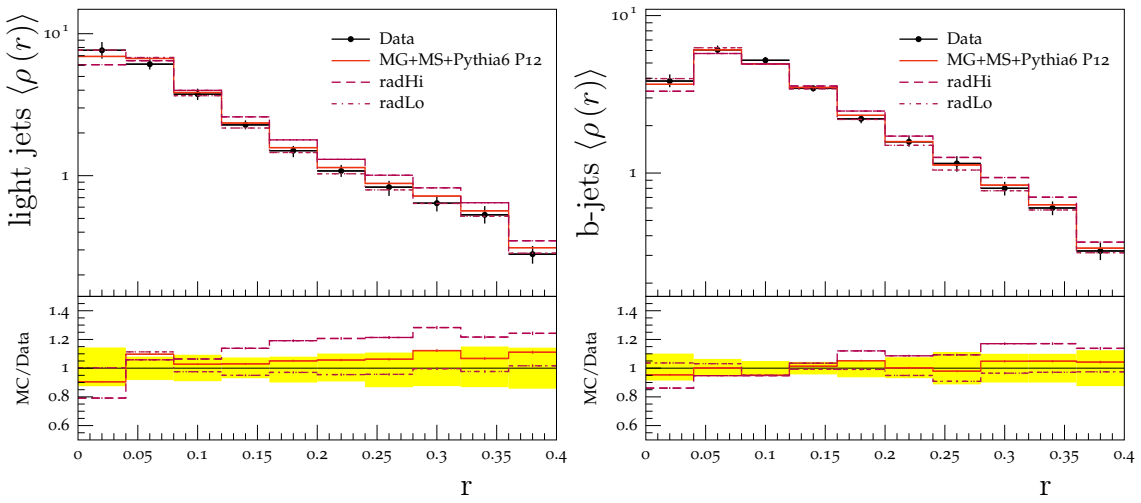

Figure 8.5.: Momentum distribution of particles inside jets ("jet shapes") measured by the ATLAS collaboration in $t\bar{t}$ events, given by the mean p_{T} fraction $\langle \rho(r) \rangle$ at the jet radius $r \pm \Delta r$ for light (left) and bottom (right) jets, compared to MadGraph + Pythia 6 simulation. Data from [166].

Top-quark transverse momentum: Recent calculations suggest that the top-quark p_{T} spectrum is affected by NNLO effects [168]. Also, differential $t\bar{t}$ cross-section measurements indicate a softer transverse momentum (p_{T}) distribution than the MadGraph predictions [140, 141, 169–172]. It was suggested that the agreement of data and simulation can be improved by choosing a different reshuffling scheme in the parton shower for assigning virtuality to the massless parton from the matrix element [173]. Different schemes are implemented in Herwig++ and Pythia 8. For evaluating the uncertainty, the top-quark p_{T} in simulation is varied to match the distribution measured by CMS [170–172].

8.4. Modeling of soft QCD

Underlying event: Soft QCD effects are taken into account by tuning Pythia to measurements of the underlying event [129]. The uncertainties are estimated by comparing the Perugia 2011 tune with the Perugia 2011 mpiHi and Perugia 2011 Tevatron (TeV) tunes [174]. As shown in comparison to CMS data in Figure 8.6 (left), the underlying event model was tuned to a larger summed transverse momentum in the mpiHi variation, while the TeV tune has a decreased underlying event activity.

Color reconnection: The uncertainties that arise from ambiguities in the modeling of color reconnection effects [175] are estimated by comparing in simulation PYTHIA 6 underlying event tunes with and without color reconnection: the Perugia 2011 and Perugia 2011 noCR tunes [174]. The right plot of Figure 8.6 shows variations of the Perugia 2012 tune for the description of the mean transverse momentum of charged particles in dependence of their number in minimum-bias events. It is equivalent to the Perugia 2011 family for this observable but features an additional "loCR" tune with medium color reconnection that agrees with the presented CMS data. The loCR variation is not used in this analysis as it fails to match the ATLAS data measured at a higher minimum particle p_{T} [176], suggesting that the model fails to described higher moments of the distribution. The new models implemented in PYTHIA 8 [92] will be studied in the future as the experimental collaborations are moving to PYTHIA 8 for the LHC run 2.

This uncertainty is expected to be closely related to the inherent ambiguity of $\mathcal{O}(\Lambda_{\mathrm{QCD}})$ in determination of m_t from kinematic reconstruction, that is caused by the ambiguous assignment of radiation due to the color connection between the bottom jet and the beam remnant. This suggests a dependence of the measurement of m_t on the event kinematics that is further studied in Sec. 11.

8.5. Discussion

Improvement with regard to previous measurements

The total uncertainties expected for the 2D and 1D approaches are of similar size, 0.64 GeV and 0.66 GeV, respectively. The published 2011 analysis at $\sqrt{s} = 7$ TeV [13] is based on the work in this thesis and has a total uncertainty of 1.07 GeV in the 2D approach. Apart from the reduction in statistical uncertainty due to the larger data set, this is achieved by the following improvements: The jet energy corrections and the jet energy resolution are determined more precisely, and larger simulated samples improve the statistical significance of systematic variations and reduce statistical fluctuations. Improving the understanding and the factorization of the systematic uncertainties is highly beneficial. So the hadronization uncertainty is reduced by deriving the individual uncertainty associated to each jet flavor from the comparison of PYTHIA 6 and HERWIG++, while the 2011 analyis used the full difference between

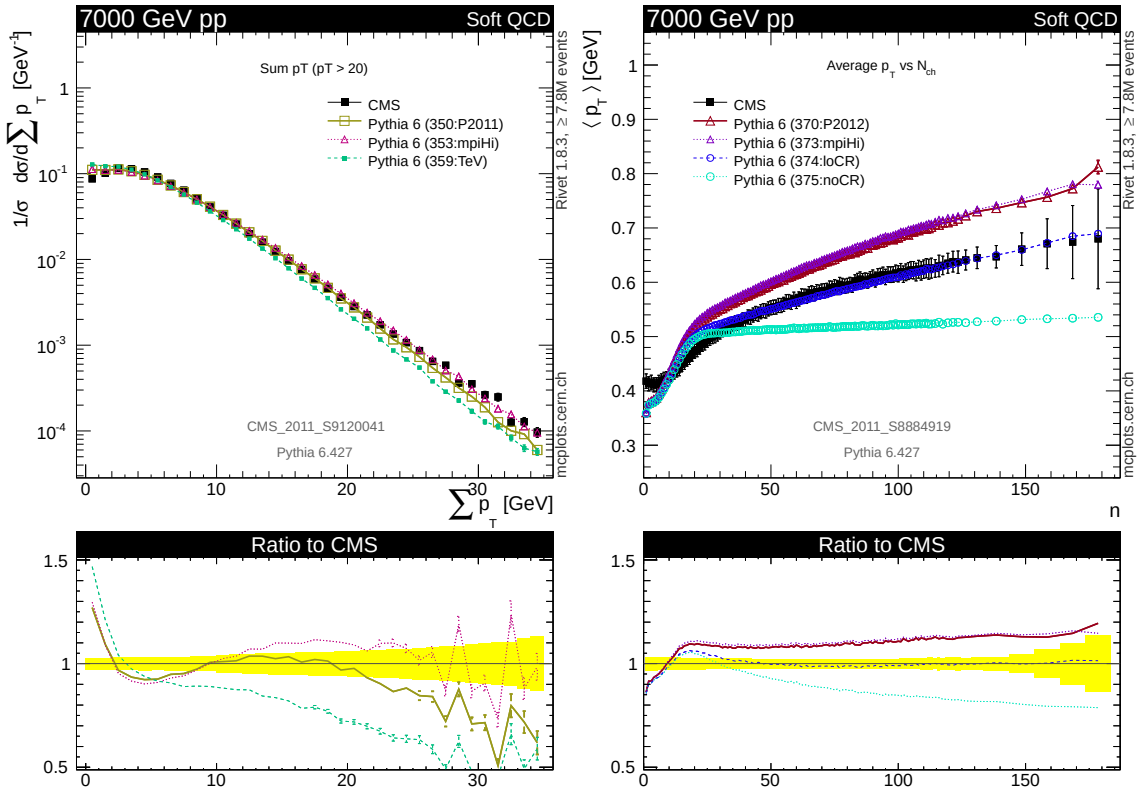

Figure 8.6.: The summed transverse momentum (left) and the average p_{T} vs. the charged particle multiplicity measured in minimum-bias events at CMS [129, 177]. Plots created using mcplots.cern.ch [178].

light-quark and gluon jets as an upper estimate of the uncertainty for bottom jets. In turn, b-jet specific uncertainties in the Lund string model are added to this new result.

The preliminary CMS result at $\sqrt{s} = 8\,\text{TeV}$ [11], also based on this thesis, has a total uncertainty on the top-quark mass of 0.77 GeV in the 2D approach. The improvements in the analysis presented here are due to the usage of the (close-to) final jet energy corrections and resolutions, as well as due to a correction for the estimation of the statistical significance of the systematic samples.

Comparison of 2D, 1D and hybrid approaches

Comparing the 2D and 1D approaches, the uncertainties from jet energy corrections are reduced for the 2D approach as expected. As the JEC uncertainties are larger for low-p_T jets (cf. Section 5.4) and the p_T spectrum of the light jets from the W boson is softer than the spectrum of the bottom jets from the top-quark decays, the top-quark mass in the 2D fit tends to be slightly overcorrected by the flat JSF. This yields an opposite sign for the JEC uncertainties of the 1D and 2D approaches.

A similar effect is observed for the radiation uncertainties, i.e., renormalization/factorization scales and matching threshold. Additional radiation enhances the fraction of unmatched permutations that have on average larger reconstructed invariant masses by picking wrong jets for the $t\bar{t}$ reconstruction. As the shift in m_W^reco is larger than needed to compensate the effect on m_t^fit, the top mass result in the 2D approach gets overcorrected, so that the radiation uncertainties acquire opposite signs with regard to the 1D approach.

In the 1D approach, the flavor-dependent JEC uncertainty is mainly driven by the bottom uncertainty because the invariant mass of the two untagged jets is fixed to $m_W^\text{fit} = 80.4\,\text{GeV}$. In the 2D approach, a significant contribution from the flavor uncertainties of the light quarks and gluons enters the analysis through their contribution to the reconstructed W-boson mass.

The hybrid approach has an expected overall uncertainty of 0.52 GeV as it makes use of all information available on the jet energy scale and thus reduces the overcompensation effects in the 2D approach. The choice of setting $w_\text{hyb}^\text{1D} = 0.5$ is a posteriori justified by

- the similar precision of the 2D and 1D approaches

- the similar precision of the JSF extracted in this analysis (± 0.007) compared to the jet energy scale from γ/Z+jet events, as seen by evaluating the JEC uncertainty on the JSF (± 0.006)

Therefore, assigning an equal weight to the 2D and 1D approaches is a natural choice. While the JEC uncertainty becomes slightly larger compared to the 2D approach, there is a net gain for the jet energy resolution, the flavor-dependent JEC, the radiation uncertainties, and the uncertainty on the top-quark transverse momentum. The hybrid approach also yields an improved statistical uncertainty that, while beneficial on its own, reduces the statistical uncertainty on the observed systematic shifts.

Hadronization uncertainties

In all approaches, the sum of the modeling uncertainties is larger than the experimental ones. The hadronization uncertainties are most frequently debated, c.f. [3, 179]. In principle, they can be considered to be semi-experimental as they are related to the jet composition in terms of stable particle types and momenta. Different predictions for the jet composition result in different extrapolation from the well-measured Z+jet reference flavor composition to pure flavor samples. Experimentally, this issue can be tackled either by an improved determination of the jet compositions or by constraining the jet response for samples enriched with certain jet flavors. The first approach has extremely high demands on hadron discrimination, while the latter requires efficient flavor tagging and a large integrated luminosity. As demonstrated by CMS, a good precision can be achieved from $Z + b$ events with the existing b-tag algorithms and the data sample collected at $\sqrt{s} = 8\,\text{TeV}$ [122].

At the generated particle level, a comparison of string and cluster fragmentation in Sherpa is performed as shown in Section 10.2. The jet-parton response as well as the reconstructed top-quark and W-boson masses are in good agreement, suggesting no significant genuine hadronization effects if all particles could be measured perfectly.

Differences between Pythia and Herwig generators

The comparison of different Pythia and Herwig setups using the full analysis at reconstruction level yields the shifts documented in Table 8.2. It is evident that the shift, e.g., between Powheg + Pythia 6 Z2* and Powheg + Herwig 6 AUET2

is of the same order of magnitude as the total uncertainty of the presented measurement for all three approaches. Furthermore, the selection efficiency in POWHEG + HERWIG 6 AUET2 is found to be larger by about 15%, spoiling the good agreement of the event yield in data and simulation shown in Table 6.2. A priori, such differences are no surprise, as the jet energy calibration and scale factors for b-tag/mistag efficiencies are determined in CMS with respect to PYTHIA 6 Z2*, so these constraints are not valid for HERWIG 6 AUET2 without additional corrections or scale factors.

In order to exclude the detector effects, different generator setups are studied using a simplified top-quark mass analysis at particle level, as discussed in detail in Section 10.1. One of the findings is that the reconstructed top-quark and W-boson masses differ significantly at particle level. E.g., the difference between POWHEG + PYTHIA 6 Z2* and POWHEG + HERWIG 6 AUET2 for a 2D approach is found to be $\delta m_{t,\text{particle}}^{\text{2D}} = 1.43\,\text{GeV}$, which is even larger than at reconstruction level. The difference is likely to stem from the parton showers that might give different predictions for the radiation from the top and bottom quarks. To study the impact of possible issues in the radiation modeling, the matrix-element corrections to the first emissions in the decays of top quarks and W bosons are switched off in PYTHIA 8. This results in a difference of up to 0.9 GeV, supporting the hypothesis of different $t \to bWg$ radiation patterns in PYTHIA and HERWIG as possible explanation. Comparisons of samples generated without underlying event and hadronization may shed additional light on the issue.

The difference between HERWIG 6 and HERWIG++ interfaced to POWHEG is found to be in the order of 1 GeV, while PYTHIA 6 and PYTHIA 8 agree within 0.3 GeV for different setups. The stability of the PYTHIA-based setups is motivation for defining the top-quark mass implemented in PYTHIA as the mass to be measured in this analysis. The issue of different radiation treatment may be related to the issue of the top-quark mass definition itself (c.f. 2.2) so that advancements in the translation between the MC and field-theory mass definitions could finally compensate the differences between the generator setups used for the measurement.

Table 8.2.: List of additional comparisons for the combined fit to the entire lepton+jets data set. A minus-sign (–) in the $\delta {m_t}^{\mathrm{1D}}$ column indicates that the shift is anti-correlated to the corresponding one on ${m_t}^{\mathrm{2D}}$.

	2D approach		1D approach	Hybrid
	$\delta {m_t}^{\mathrm{2D}}$ (GeV)	δJSF	$\delta {m_t}^{\mathrm{1D}}$ (GeV)	$\delta {m_t}^{\mathrm{hyb}}$ (GeV)
Cross-checks				
MadGraph + Pythia 6 Z2* vs. MadGraph + Pythia 6 P11	0.21±0.10	0.003±0.001	–0.04±0.07	0.08±0.08
Powheg + Pythia 6 Z2* vs. Powheg + Herwig 6 AUET2	0.74±0.14	0.003±0.001	0.50±0.08	0.62±0.10
Powheg + Pythia 6 Z2* vs. MC@NLO + Herwig 6	0.12±0.13	0.010±0.001	0.97±0.08	0.55±0.10
MC@NLO + Herwig 6 vs. Powheg + Herwig 6 AUET2	0.62±0.13	0.013±0.001	–0.48±0.08	0.07±0.10

9. Results

9.1. Top-quark mass result

After full calibration of the measurement method using simulated events and the estimation of all systematic uncertainties, the measurement is performed on the 28 199 lepton+jets events selected in the collision data at $\sqrt{s} = 8\,\text{TeV}$, corresponding to an integrated luminosity of $19.7\,\text{fb}^{-1}$. All three approaches (2D, 1D, and hybrid) are evaluated. The resulting values of the top-quark mass are understood to be the optimal parameters for MC simulations using the PYTHIA parton shower and may be numerically different for other parton showers.

Using the 2D approach, the following result is obtained:

$$\begin{aligned} m_t^{\text{2D}} &= 172.15 \pm 0.19\,(\text{stat.+JSF}) \pm 0.61\,(\text{syst.})\,\text{GeV}, \\ \text{JSF} &= 1.006 \pm 0.002\,(\text{stat.}) \pm 0.007\,(\text{syst.}). \end{aligned}$$

Figure 9.1 shows the 2D likelihood obtained from data and compares the uncertainty of the measurement with the expected precision from the pseudo-experiments. As the top-quark mass and the JSF are measured simultaneously, the statistical uncertainty on m_t combines the statistical uncertainty arising from both components of the measurement. The overall uncertainty is 0.64 GeV on the top-quark mass from adding the components in quadrature. Within systematic uncertainties, the measured JSF is compatible with a value of JSF = 1, so it provides an independent cross-check of the jet energy scale obtained from γ/Z+jet events (c.f. Section 5.4). The measurements using only muon+jets or electron+jets events are found to agree within their statistical uncertainties, as expected from the small size of the lepton uncertainties:

$$\begin{aligned} \mu\text{+jets:} \quad & m_t^{\text{2D}} = 172.06 \pm 0.27\,(\text{stat.+JSF})\,\text{GeV}, \ \text{JSF} = 1.007 \pm 0.003\,(\text{stat.}), \\ \text{e+jets:} \quad & m_t^{\text{2D}} = 172.25 \pm 0.28\,(\text{stat.+JSF})\,\text{GeV}, \ \text{JSF} = 1.004 \pm 0.003\,(\text{stat.}). \end{aligned}$$

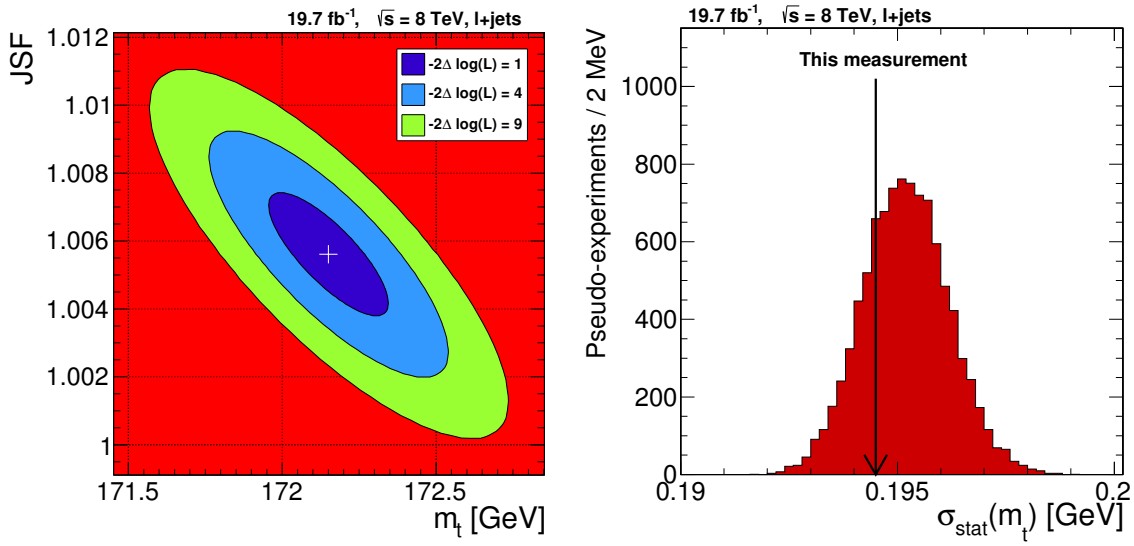

Figure 9.1.: The 2D likelihood ($-2\Delta \log(\mathcal{L})$) measured for the lepton+jets final state (left). The ellipses correspond to statistical uncertainties on m_t and JSF of one, two, and three standard deviations. The statistical uncertainty distribution obtained from 10 000 pseudo-experiments is compared to the uncertainty of the measurement in data (right).

The 1D approach sets JSF = 1, and therefore relies solely on the external jet energy calibration. It yields a result of

$$m_t^{\mathrm{1D}} = 172.63 \pm 0.12\,(\mathrm{stat.}) \pm 0.65\,(\mathrm{syst.})\,\mathrm{GeV}.$$

The total precision of 0.66 GeV is close to that of the 2D approach as the jet energy corrections for the CMS data at $\sqrt{s} = 8\,\mathrm{TeV}$ were determined very precisely by analyzing the complete data set (c.f. Section 5.4).

The hybrid approach combines the information from the simultaneous JSF fit with the external jet energy calibration, yielding a result of

$$\begin{aligned} m_t^{\mathrm{hyb}} &= 172.39 \pm 0.16\,(\mathrm{stat.}+\mathrm{JSF}) \pm 0.49\,(\mathrm{syst.})\,\mathrm{GeV}, \\ \mathrm{JSF}^{\mathrm{hyb}} &= 1.003 \pm 0.001\,(\mathrm{stat.}) \pm 0.005\,(\mathrm{syst.}). \end{aligned}$$

With a total uncertainty of 0.52 GeV, the hybrid approach presented here is the currently most precise determination of the top-quark mass, albeit under the premise that the differences between Pythia and Herwig masses are either understood or obsolete in the future. The two-dimensional hybrid likelihood is shown in Figure 9.2 (left). Figure 9.2 (right) shows the uncertainty on m_t^{hyb} as a function of the weight

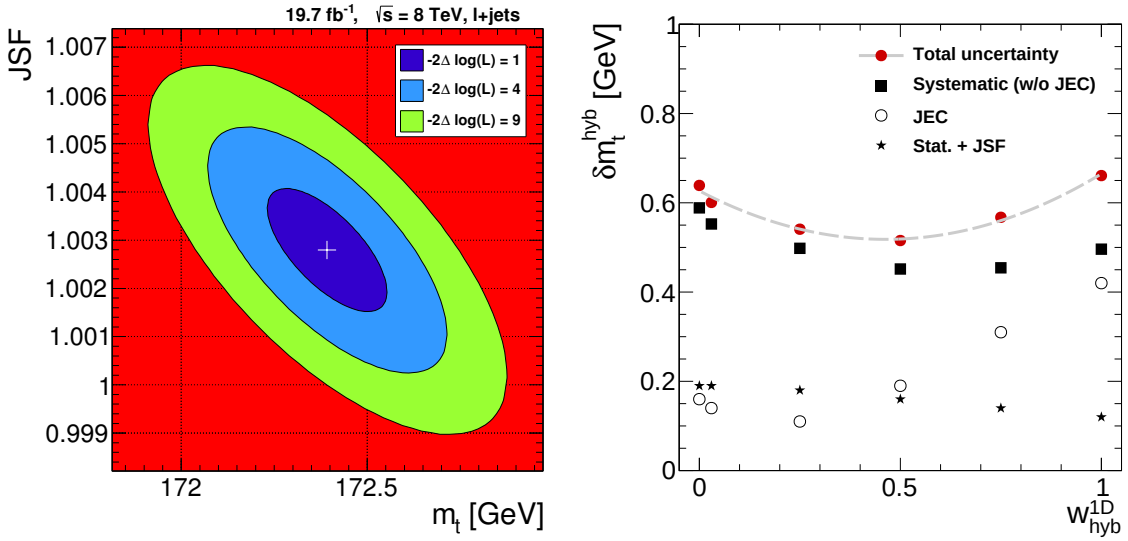

Figure 9.2.: The hybrid likelihood $(-2\Delta \log(\mathcal{L}))$ measured for the lepton+jets final state with a JSF constraint corresponding to a 1D weight in the hybrid approach $w_{\mathrm{hyb}}^{\mathrm{1D}} = 0.5$ (left). The total uncertainty together with three components is displayed as a function of $w_{\mathrm{hyb}}^{\mathrm{1D}}$ (right).

$w_{\mathrm{hyb}}^{\mathrm{1D}}$ associated to the 1D approach. As the statistical uncertainty contains a JSF component, it is largest for $w_{\mathrm{hyb}}^{\mathrm{1D}} = 0$ and constantly falling for higher $w_{\mathrm{hyb}}^{\mathrm{1D}}$. The minimum JEC uncertainty is obtained at low $w_{\mathrm{hyb}}^{\mathrm{1D}}$, where the anti-correlated effects approximately cancel. The sum of the remaining systematic uncertainties is minimized at larger $w_{\mathrm{hyb}}^{\mathrm{1D}}$. It is shown that the initial choice of $w_{\mathrm{hyb}}^{\mathrm{1D}} = 0.5$ is close to the value of minimal total uncertainty.

An overview of the measurements of the top-quark mass in the lepton+jets channel stemming from the work in this thesis is given in Table 9.1. The uncertainties in the 2D approach were mainly reduced by the improvement of details in the estimation of the systematic uncertainties and increased sizes of simulated systematics samples. In contrast, the 1D approach became competitive due to advancements in the determination of the jet energy corrections at CMS. Once the uncertainty in both approaches is of similar size, their combination in the hybrid approach yields a further improvement.

Measurement	Method	m_t [GeV]	stat. [GeV]	syst. [GeV]
CMS [13]	2D	173.49	0.43	0.98
($\sqrt{s} = 7\,\text{TeV}$ published)	1D	172.97	0.27	1.44
CMS [11]	2D	172.04	0.19	0.75
($\sqrt{s} = 8\,\text{TeV}$ preliminary)	1D	172.66	0.11	1.29
This thesis	2D	172.15	0.19	0.61
($\sqrt{s} = 8\,\text{TeV}$)	1D	172.63	0.12	0.65
	hybrid	172.39	0.16	0.49

Table 9.1.: Measurements of the top-quark mass in the lepton+jets channel performed during the course of this thesis.

9.2. Comparison with other measurements

The result using the hybrid approach is compared to other precise determinations of the top-quark mass in Figure 9.3. Following the same numbering scheme as in Figure 9.3, a short overview of each measurement is given.

1. The most precise measurement from the ATLAS collaboration was performed in the lepton+jets channel with a 3D template method using 7 TeV data [180]. In this method, a bottom JSF is fitted in addition to m_t and JSF using the ratio of bottom and light jet transverse momenta. Compared to its 2D version, a reduction of the uncertainties stemming from the relative b-jet energy scale (corresponds to flavor-dependent JEC at CMS) and hadronization (POWHEG + PYTHIA vs. POWHEG + HERWIG in $t\bar{t}$) is achieved, at the cost of a larger b-tag uncertainty. The measured bottom JSF is in excellent agreement with the expected value of 1.
   Result: $m_t = 172.31 \pm 0.75\,(\text{stat.+JSF+bJSF}) \pm 1.35\,(\text{syst.})\ \text{GeV}$
2. Based on the work in this thesis, the CMS measurement in the lepton+jets channel at $\sqrt{s} = 7\,\text{TeV}$ [13] used the ideogram 2D approach. Compared to the new measurement presented here, the jet energy corrections, their flavor-dependency, and the jet energy resolution were less precisely known. Smaller simulated samples lead to larger statistical fluctuations on the systematic uncertainties.
   Result: $m_t = 173.49 \pm 0.43\,(\text{stat.+JSF}) \pm 0.98\,(\text{syst.})\ \text{GeV}$
3. The CMS all-jets analysis at $\sqrt{s} = 7\,\text{TeV}$ [15, 17] uses the same framework of the ideogram method presented in this thesis. The QCD multijet background was suppressed by a tight event selection and modeled using an event mixing technique. Due to large statistical uncertainties on several simulated samples

used for the evaluation of systematic uncertainties, the 1D approach is quoted as main result.
Result: $m_t = 173.49 \pm 0.69$ (stat.) ± 1.21 (syst.) GeV

4. The final CDF measurement in the lepton+jets channel [181] relies on a template fit to the reconstructed mass of the W boson and two top-quark candidates per event. The two dominant systematic uncertainties stem from the jet energy corrections and the generator uncertainty, evaluated by a comparison of Pythia and Herwig in $t\bar{t}$ events. The uncertainty on the detector response for bottom jets is taken to be $\pm 0.2\%$ [182] but no public documentation on the derivation is available.
Result: $m_t = 172.85 \pm 0.71$ (stat.+JSF) ± 0.85 (syst.) GeV
5. This measurement from the D0 collaboration [183] is based on a partial dataset. It uses a matrix-element method (MEM) that calculates the likelihood for m_t and a JSF for a given event based on the LO matrix element for $t\bar{t}$ production and decay.* In a hybrid approach, an external JSF constraint from γ+jet and dijet events is added. A flavor-dependent jet energy correction is obtained from the light-quark- and gluon-enriched γ+jet and dijet samples by a fit of the single-particle responses, and is then propagated to bottom jets with a very small uncertainty. The reponse for central b jets with $p_{\mathrm{T}} = 30\,\mathrm{GeV}$ in data is determined to be $2.3 \pm 0.3\%$ lower than in simulation. The corresponding correction thus shifts the measured top-quark mass to a higher value. The dominant uncertainty, "Hadronization and UE", stems from an Alpgen + Pythia/Herwig comparison in $t\bar{t}$ events.
Result: $m_t = 174.94 \pm 1.14$ (stat.+JSF) ± 0.96 (syst.) GeV
6. The world combination from March 2014 [3] was performed using the BLUE method [184]. The measurements 1–5, including the two CMS results based on the ideogram method presented in this thesis, and six measurements of lower precision served as input. The measurements were found to be in general agreement, indicated by a low χ^2 of 4.3 for 10 degrees of freedom, corresponding to a probability of 93%.
Result: $m_t = 173.34 \pm 0.27$ (stat.) ± 0.71 (syst.) GeV

After the release of the world combination, the following high-precision results appeared:

*The MEM can be understood as a simultaneous fit to all observables that can be built with the four-vectors of the selected objects, yielding a superior statistical precision, but requires long computation times.

7. The preliminary result from the CMS collaboration using data at $\sqrt{s} = 8\,\text{TeV}$ in the lepton+jets channel [11]. The jet energy corrections and resolution were less precisely known compared to the result presented in this thesis. The result uses the 2D approach and the largest uncertainty stems from the flavor-dependence of the jet energy corrections.
   Result: $m_t = 172.04 \pm 0.19\ (\text{stat.+JSF}) \pm 0.75\ (\text{syst.})\ \text{GeV}$
8. Similarly, the CMS measurement in the all-jets channels was updated to 8 TeV [16, 17]. Using the 2D approach, the flavor-dependent JEC were found to be the largest source of uncertainty.
   Result: $m_t = 172.08 \pm 0.36\ (\text{stat.+JSF}) \pm 0.83\ (\text{syst.})\ \text{GeV}$
9. The D0 analysis using the matrix-element method was updated to the full Tevatron dataset [185]. A huge additional correction for the energies of central bottom jets, based on the single-particle responses extracted from light-quark and gluon jets, of $-0.064 \pm 0.003\%$ at $p_\text{T} = 30\,\text{GeV}$ [186] is applied. The "Hadronization and UE" uncertainty was reduced by using the particle-level jet energies after the events were selected using the jet energies at reconstruction-level. In contrast to the previous D0 measurement (5 in this list), no external JSF constraint was used. The measured JSF of 1.025 ± 0.005 (stat.) does not seem to agree well with the jet energy calibration from γ+jet and dijet events, although no systematic uncertainty on the JSF is given. Extrapolating the published 2D likelihood to JSF = 1 would yield a 1D top-quark mass of approximately 179 GeV, with increased uncertainties from jet energy corrections.
   Result: $m_t = 174.98 \pm 0.58\ (\text{stat.+JSF}) \pm 0.49\ (\text{syst.})\ \text{GeV}$
10. The result presented in this thesis, using CMS data in the lepton+jets channel at $\sqrt{s} = 8\,\text{TeV}$ and a hybrid approach that combines the simultaneous JSF measurement with an external constraint from γ/Z+jet events.
    Result: $m_t = 172.39 \pm 0.16\ (\text{stat.+JSF}) \pm 0.49\ (\text{syst.})\ \text{GeV}$

All measurements listed here use a Pythia-based setup as default simulation and thus measure the top-quark mass as defined in Pythia. It is noticeable that the measurements by the ATLAS, CDF, and D0 collaborations quote an uncertainty based on the comparison of $t\bar{t}$ samples generated with Pythia and Herwig. As argued in Section 8.5, the jet energy corrections and other data/MC scale factors are usually determined with respect to Pythia so that their application to Herwig samples discards the constraints from data and introduces potential double-counting of uncertainties.

Excluding detector effects, even larger differences between PYTHIA and HERWIG are expected from the study of top-quark reconstruction at particle level (c.f. Section 10.1) that potentially affect all measurements based on kinematic reconstruction in a similar way. As the "true" differences between both generators are therefore likely to be underestimated by the quoted values, it is more useful to identify the measured top-quark mass with m_t^{PYTHIA} and rely on the on-going theory efforts for connecting m_t^{PYTHIA} (and m_t^{HERWIG}) with the pole and short-distance masses [30].

9.3. Compatibility testing using the BLUE method

This section is devoted to testing the compatibility between selected measurements using the BLUE method [184] that was also used for combining the top-quark mass results from the Tevatron [187], the LHC [188], all of them together [3], or CMS alone [179]. The BLUE method tries to find a linear combination of the input measurements that minimizes the total uncertainty, using the individual uncertainties and their estimated correlations.

BLUE combination of 2D and 1D approaches

As a cross-check of the hybrid approach, a combination of the 2D and 1D results presented in Section 9.1 is performed using the BLUE method. The effects of the systematic uncertainty sources were found to be either fully correlated (correlation coefficient $\rho = 1$) or anti-correlated ($\rho = -1$), as indicated in Table 8.1. The correlation coefficient $\rho_{\text{1D}}^{\text{2D}}$ of the statistical uncertainty between both methods is determined to be $\rho_{\text{1D}}^{\text{2D}} = 0.589$ from pseudo-experiments.

The combined result using BLUE is then

$$m_t = 172.38 \pm 0.14\,(\text{stat.+JSF}) \pm 0.47\,(\text{syst.})\,\text{GeV},$$

having a total uncertainty of 0.49 GeV. The combination has a χ^2 of 0.31 for one degree of freedom, corresponding to a probability of 58% and demonstrating good internal consistency. The 1D measurement has a combination weight of 48.6%, the 2D measurement a weight of 51.4%, so that their combination approximately corresponds to a simple average.

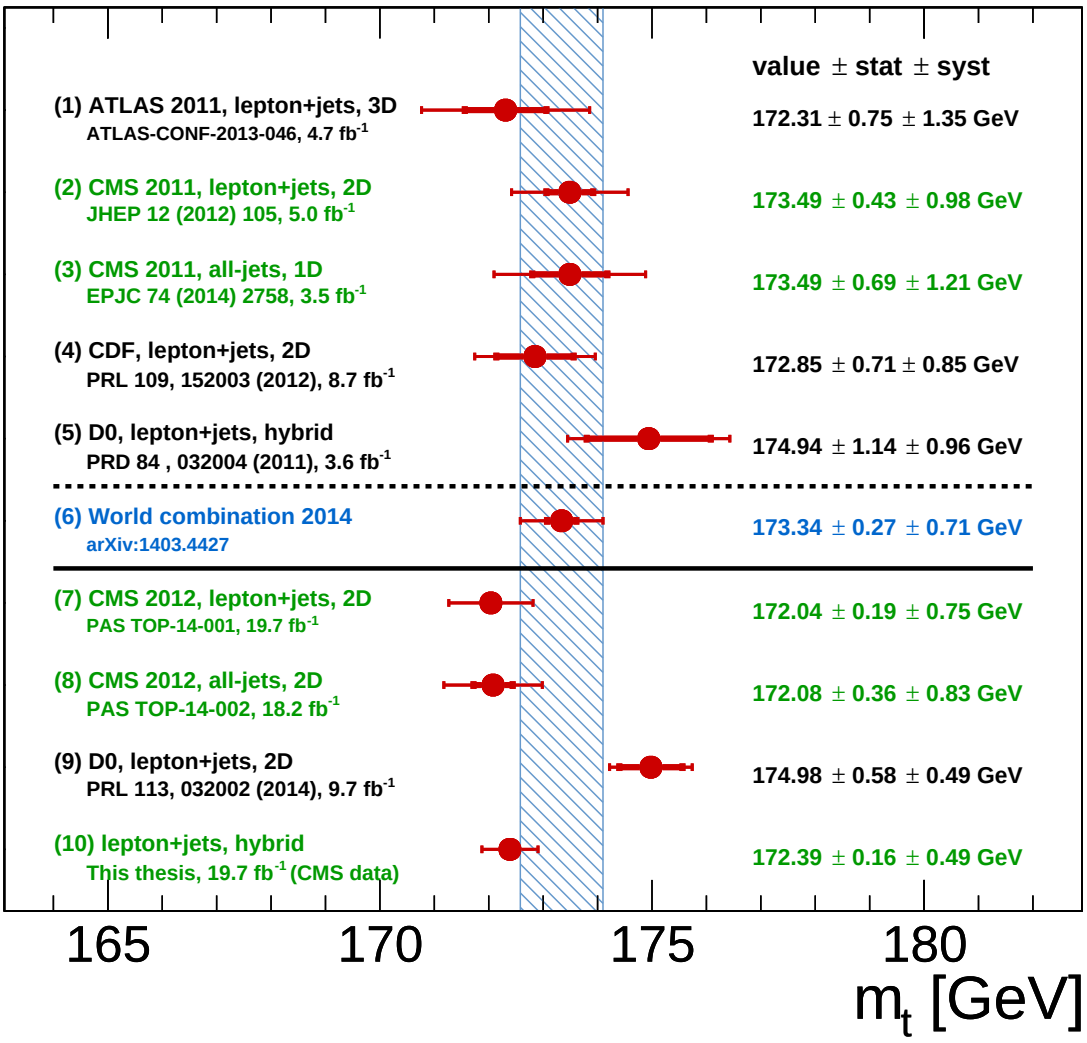

Figure 9.3.: Overview of precise measurements of the top-quark mass. The measurements highlighted in green are performed using the method described in this thesis. The world combination (5) is highlighted in blue, containing (amongst others) the measurements above the dashed line (1–5). The measurements below the solid line (7–10) appeared after the world combination.

If all systematic uncertainty sources are treated as correlated (with no negative signs), a combined result of

$$m_t = 172.36 \pm 0.15\,(\text{stat.+JSF}) \pm 0.60\,(\text{syst.})\,\text{GeV}$$

is obtained. While the central value is stable, this combination has a larger uncertainty and the χ^2 value increases to 1.48, corresponding to a probability of 22%. Therefore, both the total uncertainty and the quality of combining both approaches is improved by a correct treatment of the correlations, as done by construction in the hybrid approach.

CMS 7 TeV lepton+jets and thesis result

The high combination probabilities found in combinations of a large number of measurements are naturally explained by the inclusion of measurements with larger uncertainties that effectively only add degrees of freedom. The compatibility of the two very precise measurements in the lepton+jets channel at different center-of-mass energies ([13] and this thesis) is thus investigated without taking into account other measurements. For the comparison on equal ground, the result of the 2D approach is used for both measurements, shown in Table 9.2.

Apart from the statistical uncertainty components, the "JEC: Uncorrelated", jet energy resolution and pileup uncertainties are considered as uncorrelated between the measurements. This is due to harsher pileup conditions in the 8 TeV run period, as well as due to changes in the JEC/JER determinations using different methods and higher average transverse momenta of the jets in the γ/Z+jet events selected for jet energy calibration.

As shown in Table 9.2, the overall correlation coefficient between both measurements is 0.617 using the standard BLUE method. The 7 TeV measurement contributes only around 1% to the combined value. Different magnitudes of the individual uncertainties at 7 and 8 TeV could stem from statistical fluctuations or may indicate an overestimate of the correlation between the two measurements. Restricting the correlations for each uncertainty source X to

$$\rho_{7\,\text{TeV}}^{8\,\text{TeV},X} = \min\left(\frac{\sigma_{7\,\text{TeV}}^{X}}{\sigma_{8\,\text{TeV}}^{X}}, \frac{\sigma_{8\,\text{TeV}}^{X}}{\sigma_{7\,\text{TeV}}^{X}}\right)$$

Measurement	m_t [GeV]	stat. [GeV]	syst. [GeV]
CMS lepton+jets 7 TeV [13]	173.49	0.43	0.98
This thesis 8 TeV (2D)	172.15	0.19	0.66

Combination	Correlation 7 TeV vs. 8 TeV	Relative weight	
		7 TeV	8 TeV
BLUE default	0.617	-1.35%	101.35%
Limited correlations	0.371	15.32%	84.68%

Combination	m_t [GeV]	unc. [GeV]	χ^2/ndf	Prob	Standard deviations
BLUE default	172.132	0.642	2.56/1	11.0%	1.60
Limited correlations	172.355	0.623	1.73/1	18.9%	1.31

Table 9.2.: Combination of CMS measurements in the lepton+jets channel.

leads to a lower overall correlation of 0.371 and the 7 TeV measurement gets a weight of 15%.

In both cases, the combination results are close to the more precise 8 TeV measurement. The χ^2 probabilities are 11% for the BLUE default and 19% for BLUE with limited correlations. These correspond to an agreement between the two measurements within 1.6 and 1.3 standard deviations, respectively.

CMS 7 TeV lepton+jets, thesis result, and D0 lepton+jets

Adding the D0 measurement [185] to the combination requires additional assumptions on the correlation between the two experiments. Two scenarios will be investigated. In the first scenario, all systematic uncertainties between the two experiments are assumed to be uncorrelated ($\rho^{\mathrm{D0}}_{\mathrm{CMS}} = 0$). In the second scenario, it is assumed that experimental uncertainties are uncorrelated (including the D0 determination of the bottom jet energy scale), while modeling uncertainties are correlated. Again, BLUE is evaluated with both default and restricted correlations.

As expected and shown in Table 9.3, the CMS 7 TeV measurement has a relatively low impact on the combination, while the weight of the 8 TeV result from this thesis is slightly larger than the weight of the D0 measurement. The combinations yield top-quark masses between 173.26 and 173.43 GeV with reduced total uncertainties in the order of 0.5 GeV but the χ^2 probabilities are at the level of 1% or lower. This means that the measurements agree with each other only within 2.5–3.1 standard deviations. It is currently under investigation if the deviation is caused by experi-

Measurement	m_t [GeV]	stat. [GeV]	syst. [GeV]
CMS lepton+jets 7 TeV [13]	173.49	0.43	0.98
This thesis 8 TeV (2D)	172.15	0.19	0.66
D0 lepton+jets [185]	174.98	0.58	0.49

Combination	Correlation		
	CMS 7 TeV vs. this thesis 8 TeV	CMS 7 TeV vs. D0	This thesis 8 TeV vs. D0
BLUE default ($\rho^{\mathrm{D0}}_{\mathrm{CMS}} = 0$)	0.617	0	0
Limited correlations ($\rho^{\mathrm{D0}}_{\mathrm{CMS}} = 0$)	0.371	0	0
BLUE default	0.617	0.146	0.204
Limited correlations	0.371	0.051	0.125

Combination	Relative weight		
	CMS 7 TeV	This thesis 8 TeV	D0
BLUE default ($\rho^{\mathrm{D0}}_{\mathrm{CMS}} = 0$)	-0.79%	58.80%	41.99%
Limited correlations ($\rho^{\mathrm{D0}}_{\mathrm{CMS}} = 0$)	9.12%	50.37%	40.52%
BLUE default	-1.87%	61.87%	39.99%
Limited correlations	10.06%	50.70%	39.24%

Combination	m_t [GeV]	unc. [GeV]	χ^2/ndf	Prob	Standard deviations
BLUE default ($\rho^{\mathrm{D0}}_{\mathrm{CMS}} = 0$)	173.328	0.489	10.81/2	0.45%	2.84
Limited correlations ($\rho^{\mathrm{D0}}_{\mathrm{CMS}} = 0$)	173.419	0.481	8.92/2	1.16%	2.52
BLUE default	173.257	0.536	12.64/2	0.18%	3.12
Limited correlations	173.395	0.509	10.06/2	0.65%	2.72

Table 9.3.: Combination of CMS and D0 measurements in the lepton+jets channel.

mental issues or by different simulated reference distributions at particle level (cf. Section 10.1).

10. Generator studies

The measurement of the top-quark mass presented in this thesis relies on comparison to event generators which implement the hard scattering process, a parton shower, and models for hadronization and the underlying event, as described in Chapter 3. Therefore, different generator setups are compared for top-quark pair production using the Rivet toolkit [189]. Rivet is designed to provide a convenient way for calculating and plotting hadron level observables from HepMC input [190]. The simulated distributions can be compared to reference data, that can either be a well understood reference simulation or unfolded experimental data. In Rivet, unfolded data distributions are available from numerous experiments of 30 years of high energy physics, including electron-positron, electron-proton and proton-(anti)proton collisions. The tuning of the event generators to data from Z boson decays and minimum-bias events is an integral part of the simulation chain. Unfortunately, only very few particle level distributions from top-quark data are available up to now [166, 191], so that no systematic comparison to data is possible. Therefore, this section will compare different event generator setups among each other, and always in comparison to the CMS default setup.

10.1. Top-quark reconstruction at particle level

A study of the properties of jets and top-quark candidates using a partial reconstruction of simulated $t\bar{t}$ lepton+jets events at $\sqrt{s} = 8\,\text{TeV}$ is performed. The particle-level analysis* implements a lepton+jets selection that is close to the one used by CMS. Within a range of $|\eta| < 2.5$, exactly one charged electron or muon with $p_\text{T} > 30$ GeV is required. Jets are clustered from all particles with $|\eta| < 2.5$ using the anti-k_T algorithm with $R = 0.5$. The charged lepton with $p_\text{T} > 30$ GeV and undetectable neutrinos are excluded from the jet clustering. At least four jets are required with $p_\text{T} > 30$. Jets stemming from bottom quarks are identified by

*`MC_TTBAR_HADRON`, based on the `MC_TTBAR` analysis bundled with Rivet.

requiring $\Delta R\,(\text{jet}, \text{B}) < 0.4$, where B denotes a bottom hadron with $p_\text{T} > 5$ GeV in the event history of the generator. The quark content of the B hadron is accessible at generator level[†] and can be exploited to tag jets as stemming from a b or $\bar{b}$ quark. The events are required to contain exactly one bottom and one antibottom jet, as well as at least two untagged jets, in order to enhance the probability for a correct reconstruction of the hadronically decaying top quark.

In order to find a W boson candidate at particle level, the invariant masses of all pairs of untagged jets are calculated. The pair with the invariant mass closest to 80.4 GeV serves as the W boson candidate. To enhance the number of correct candidates, only events with 75 GeV $< m_{jj} <$ 85 GeV are selected. For the reconstruction of a top quark candidate, the W candidate needs to be combined with one of the two bottom jets. If the hard lepton in the event is positively charged, it stems from a $t \to bW^+ \to b\ell^+\nu_\ell$ decay chain. Therefore, the W boson reconstructed from jets must be negatively charged, requiring a pairing with the antibottom jet to reconstruct the decay $\bar{t} \to \bar{b}W^- \to \bar{b}\bar{q}q'$, and vice versa for opposite charges.

Using Gaussian distributions, the reconstructed W candidate mass m_{jj} is fitted within a range of 75–85 GeV, and the top-quark candidate mass m_{jjb} within a range of 160–185 GeV. In addition, the 2D approach presented in Section 7.1 is mimicked, where a correction factor for the top-quark mass is derived using the fixed W mass $m_W = 80.4\,\text{GeV}$, yielding a top-quark mass

$$\left\langle m_t^{2D} \right\rangle = \langle m_{jjb} \rangle \times \frac{\langle m_W \rangle}{\langle m_{jj} \rangle}.$$

The 2D mass is therefore corrected for effects that affect both m_{jj} and m_{jjb} similarly. The results obtained for the different generator setups described in the following are summarized in Table 10.1.

Default simulation

The default $t\bar{t}$ simulation produced by the CMS collaboration is based on the MadGraph 5 matrix-element generator, interfaced with Pythia 6 for parton shower and hadronization, using the Z2* tune [129]. Top-quark pair events are generated with $m_t^{MC} = 172.5$ GeV and up to three additional partons from initial-state radiation (ISR). MLM matching is used to remove double-counting of emissions by the

[†]In a detector-level analysis, the charge of the secondary vertex could be used for distinguishing between hadrons with b and $\bar{b}$ quarks.

Table 10.1.: List of generator setups used in the particle-level analysis. The shifts observed in the reconstructed masses are with regard to the reference setup MadGraph + Pythia 6 Z2*. The different Pythia setups are marked with a closed circle (●), where single/few-parameter variations are excluded.

	Generator setup	δm_{jj}	δm_{jjb}	δm_t^{2D}
1	MadGraph + Pythia 6 Z2*, $m_t = 172.5\,\text{GeV}$ ●			
	MadGraph + Pythia 6 Z2*, $m_t = 169.5\,\text{GeV}$	−0.02	−2.84	−2.81
	MadGraph + Pythia 6 Z2*, $m_t = 175.5\,\text{GeV}$	+0.01	+2.50	+2.47
	MadGraph + Pythia 6 Z2*, scale down	+0.18	+0.21	−0.17
	MadGraph + Pythia 6 Z2*, scale up	−0.15	−0.23	+0.08
2	MadGraph + Pythia 6 P11 ●	−0.12	−0.28	−0.02
	MadGraph + Pythia 6 P11TeV	−0.30	−0.51	+0.13
	MadGraph + Pythia 6 P11mpiHi	−0.10	−0.20	+0.01
	MadGraph + Pythia 6 P11noCR	+0.03	+0.02	−0.04
3	MadGraph + Pythia 8 CUEP8M1 ●	−0.15	−0.23	+0.08
	MadGraph + Pythia 8 CUEP8M1, global recoil	−0.14	−0.20	+0.10
	MadGraph + Pythia 8 CUEP8M1, no ME corr.	−0.12	−0.87	−0.62
4	aMC@NLO FxFx + Pythia 8 CUEP8M1, ME corr. ●	−0.09	−0.27	−0.08
	aMC@NLO FxFx + Pythia 8 CUEP8M1, no ME corr.	−0.07	−0.80	−0.64
	aMC@NLO + Pythia 8 CUEP8M1, ME corr. ●	−0.28	−0.31	+0.28
	aMC@NLO + Pythia 8 CUEP8M1, no ME corr.	−0.22	−1.07	−0.60
	aMC@NLO + Herwig++ EE-3C	−0.27	−2.80	−2.25
5	Pythia 8 CUEP8M1, ME corr.	−0.13	−0.33	−0.07
	Pythia 8 CUEP8M1, no ME corr	−0.07	−0.98	−0.83
	Herwig++ EE-3C	−0.29	−1.63	−1.03
6	Powheg + Pythia 6 Z2* ●	−0.05	−0.23	−0.13
	Powheg + Pythia 8 CUEP8M1 ●	−0.24	−0.42	+0.08
	Powheg + Herwig++ EE-3C	−0.10	−2.60	−2.39
	Powheg + Herwig++ EE-5C	+0.43	−1.79	−2.68
	Powheg + Herwig 6 AUET2	+0.81	+0.14	−1.56
7	Sherpa 2, Lund fragmentation	+0.61	−0.66	−1.94
	Sherpa 2, Cluster fragmentation	+0.65	−1.12	−2.49
	All setups (without mass variations)			
	- Mean difference to default	−0.01	−0.82	−0.67
	- RMS	0.30	0.77	0.94
	Pythia setups (●)			
	- Mean difference to default	−0.13	−0.25	+0.03
	- RMS	0.09	0.12	0.12

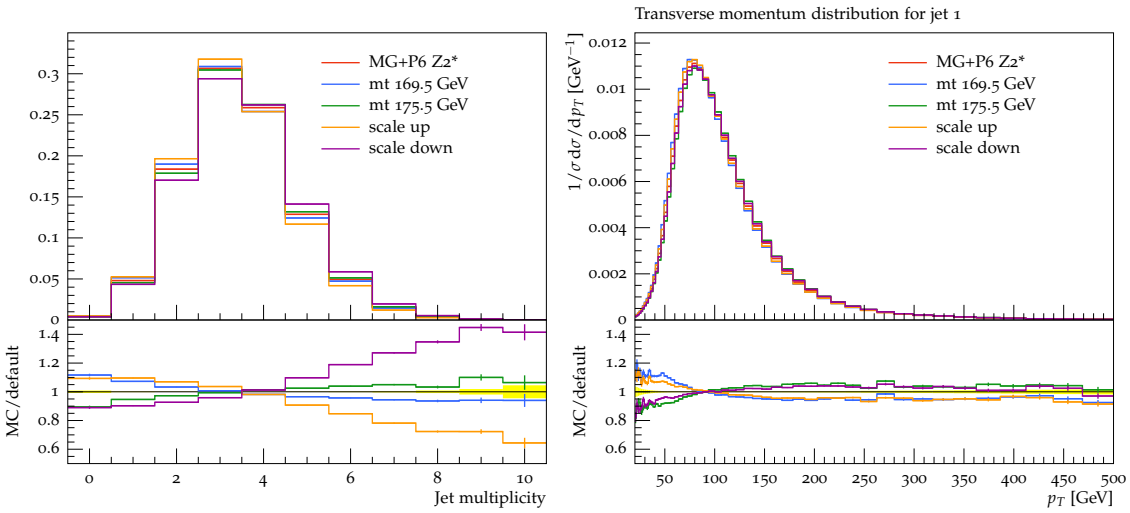

Figure 10.1.: Multiplicity of jets with $p_{\mathrm{T}} > 30$ GeV (left) and the transverse momentum of the hardest jet (right) in simulated $t\bar{t}$ lepton+jets events, showing the default setup MadGraph + Pythia 6 Z2* and variations. Each distribution is normalized to unity.

matrix element and the parton shower, and ensures a smooth transition between the phase-space region where each dominates. The top-quark decays are simulated with MadSpin, taking into account spin correlations between the decay products. The first emissions in the top and W decays are corrected with the NLO matrix element.

The impact of different systematic variations is evaluated, with special attention to the invariant mass of the reconstructed top-quark candidate. The effect of a variation in the input top-quark mass is illustrated by variations of ± 3 GeV. The renormalization and factorization scales are varied simultaneously by factors of $1/2$ and 2, resulting in increased and reduced jet activity. The shifts in reconstructed masses are documented in block 1 of Table 10.1.

Figure 10.1 shows the jet multiplicity and the transverse momentum of the jet leading in p_{T}. The up variation of the renormalization scale μ_R implies a smaller value of $\alpha_s(\mu_R)$. Thus, the jet multiplicity is smaller, and the leading jet p_{T} spectrum is softer due to smaller boost of the $t\bar{t}$ system. Opposite effects are found for the scale down variation. Changing the input mass has only mild effects on the jet multiplicity, while the jet p_{T} is altered: a larger (smaller) top-quark mass leads to harder (softer) decay products.

Figure 10.2 shows the transverse momentum and the invariant mass of the leading bottom jet after the full selection. There is in general a good agreement among the

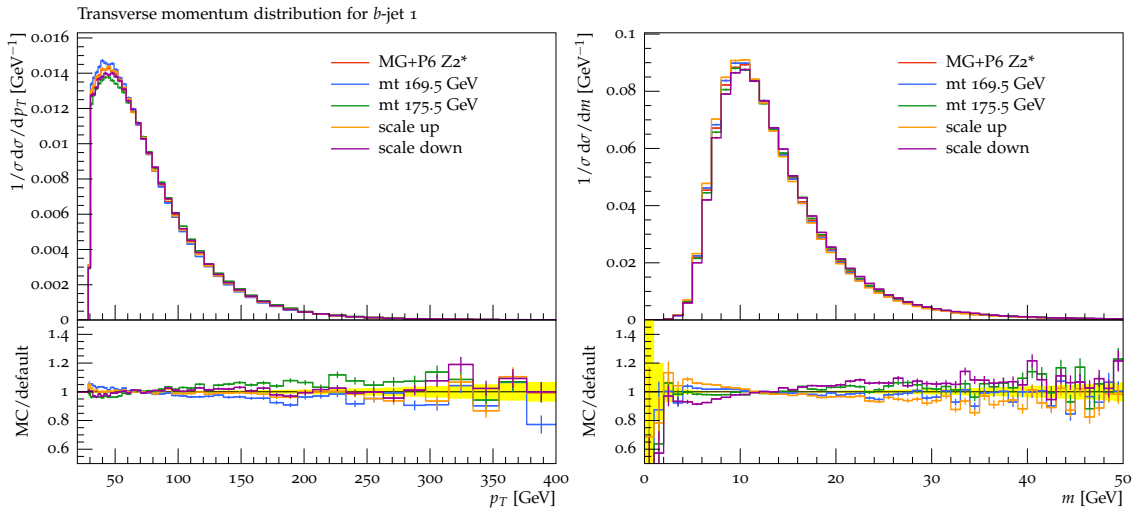

Figure 10.2.: Transverse momentum (left) and mass (right) of the hardest bottom jet in simulated $t\bar{t}$ lepton+jets events, showing the default setup and variations. Each distribution is normalized to unity.

predictions. The jet mass shows some sensitivity on the renormalization scale used in the simulated sample, where a larger α_s leads to slightly more massive jets.

Figure 10.3 indicates that the masses of W boson and top quark reconstructed from jets are affected by scale variations, such that a higher scale (smaller α_s) leads to a lower mass. This can be explained as a sum of two effects: the jets masses are smaller on average and the probability to choose hard initial-state radiation for the reconstruction of the W boson is lower. Therefore, a shift of roughly 200 MeV can be seen for both distribution, leading to a slight overcompensation in the 2D approach. The variation of the input top-quark mass shifts the invariant mass peak of the top-quark candidate by several GeV, as expected.

A different tuning of hadronization parameters is evaluated using the Perugia 2011 tune [174], along with its variations of underlying event activity ("TeV" and "mpiHi") and color reconnection ("noCR"), shown in block 2 of Table 10.1. It is found that the jet multiplicity and p_{T} distributions are not affected by the different tunes. As shown in Figure 10.4, the systematic underlying event variations of the P11 tune cover a similar range in jet mass as the scale variations.

Figure 10.5 shows the invariant mass distributions. Shifts of a few hundred MeV are observed but successfully compensated for m_t^{2D}, showing that a retuning of hadronization and the underlying event in PYTHIA is mostly agnostic of the jet flavor.

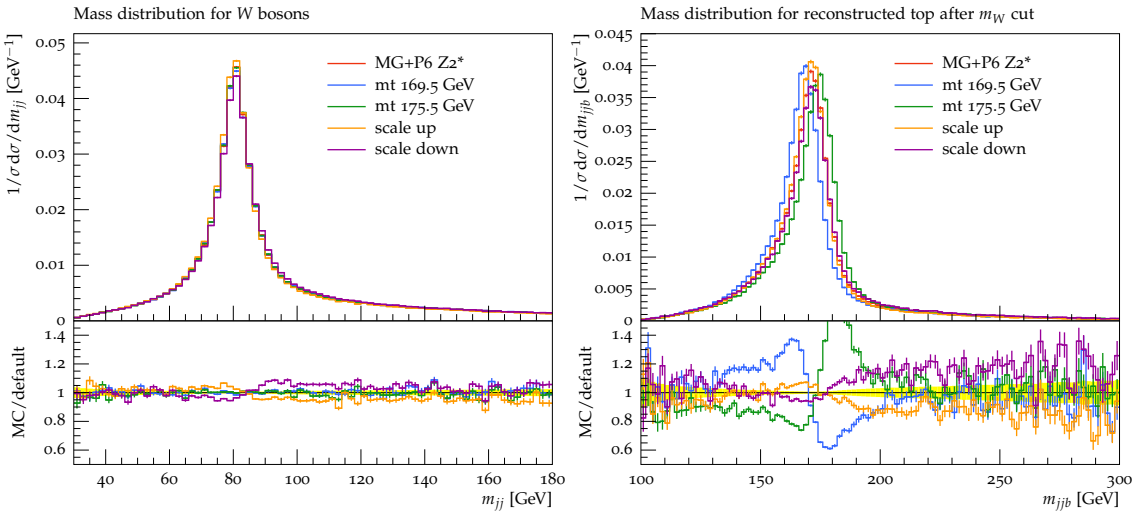

Figure 10.3.: Invariant mass of W boson (left) and top-quark candidates (right) in simulated $t\bar{t}$ lepton+jets events, showing the default setup and variations. Each distribution is normalized to unity.

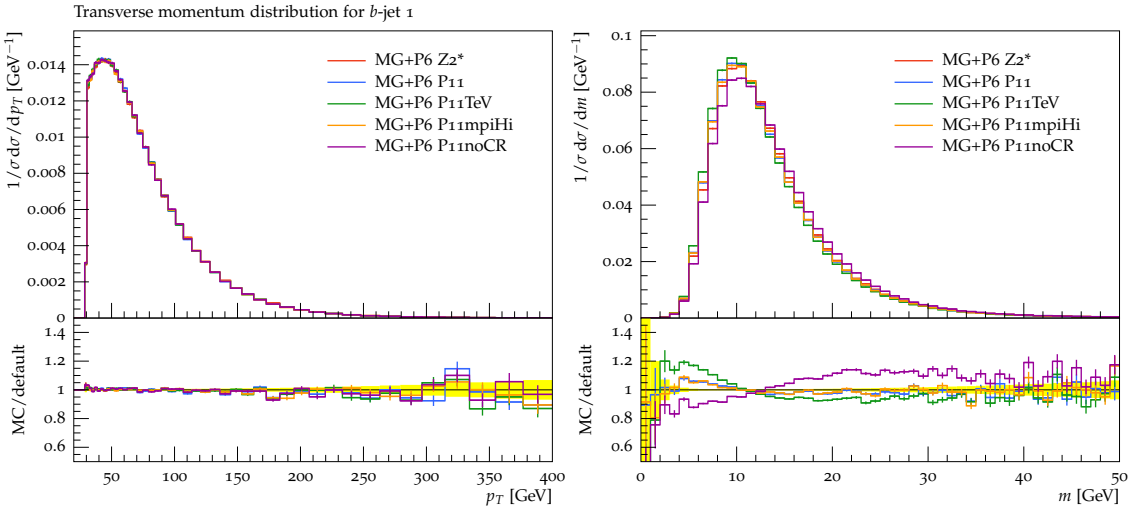

Figure 10.4.: Transverse momentum (left) and mass (right) of the hardest bottom jet in simulated $t\bar{t}$ lepton+jets events, comparing the default setup and the Perugia 2011 tunes. Each distribution is normalized to unity.

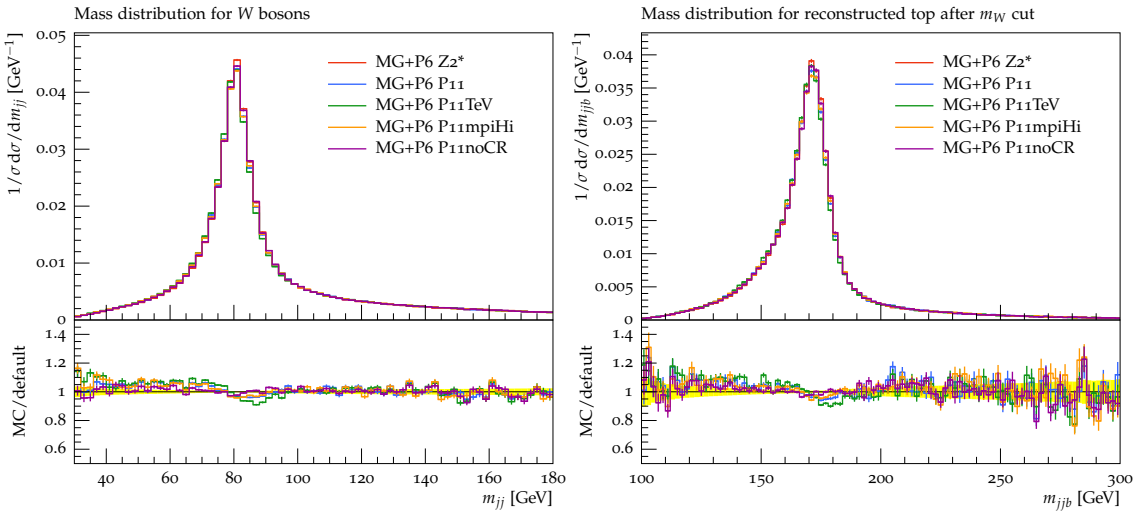

Figure 10.5.: Invariant mass of W boson (left) and top-quark candidates (right) in simulated $t\bar{t}$ lepton+jets events, comparing the default setup and the Perugia 2011 tunes. Each distribution is normalized to unity.

Alternative generators

The generator setups of block 3 in Table 10.1 are created using the same MadGraph matrix element as the previous blocks but interfaced to Pythia 8, using the CMS CUEP8M1 tune, which is based on the Monash tune [192]. As shown in Figure 10.6, a large shift in reconstructed masses for the W boson, the top quark as well as the jet mass is observed when matrix element corrections to the first emission within the decays are turned off in order to study their effect. Activating the "global recoil" instead of the standard "dipole recoil" scheme for momentum reshuffling does not lead to significant differences for the observables under study.

For the LHC Run 2, the CMS collaboration intends to use the NLO successor of MadGraph, aMC@NLO, as default matrix-element generator. It can be interfaced to both Pythia 8 and Herwig++. Block 4 in Table 10.1 lists the studied setups, where "FxFx" indicates a merged prediction for $t\bar{t}$ with up to 2 additional partons at NLO precision. Otherwise, only 1 real emission is contained in the NLO $t\bar{t}$ matrix element. The top-quark mass of the available FxFx sample is 173 GeV instead of 172.5 GeV, the shift in the table is corrected for that. The MC@NLO method requires the matrix-element corrections for the initial state parton shower to be turned off. Here, a preliminary version of Pythia 8 was used that allows for using ME corrections for the radiation inside the W-boson and top-quark decays. It can be noted that the effect of ME corrections in aMC@NLO + Pythia 8 is similar

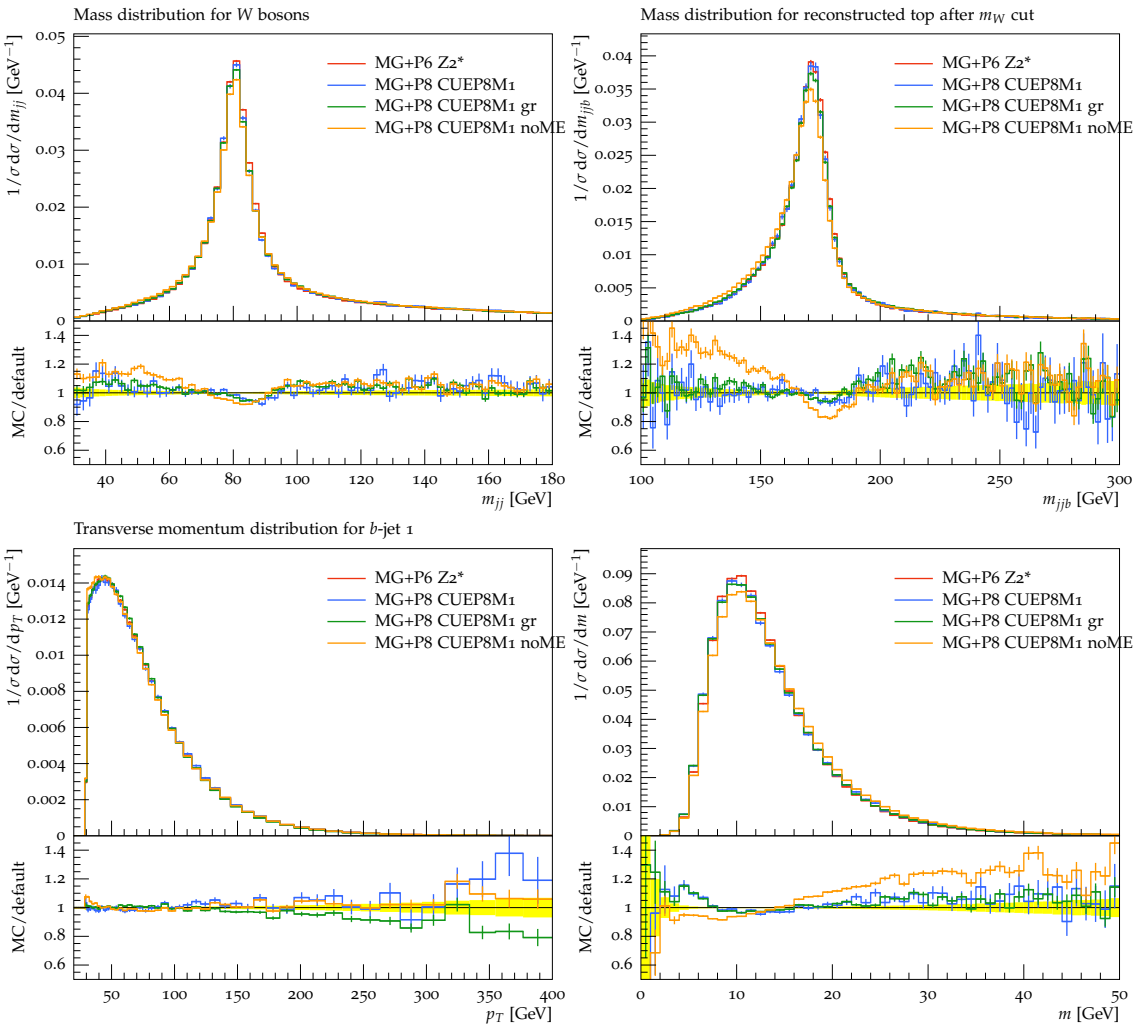

Figure 10.6.: Invariant mass of W-boson (upper left) and top-quark candidates (upper right), and transverse momentum (lower left) and invariant mass (lower right) of the leading bottom jet in simulated $t\bar{t}$ lepton+jets events, comparing the default setup and PYTHIA 8. Each distribution is normalized to unity.

to that in MadGraph + Pythia 8. As visible in Figure 10.7, the samples generated with ME corrections show a good agreement with the default setup. Interfacing the NLO matrix element to Herwig++ leads to a large shift of several GeV in the reconstructed top-quark mass, resembling the default sample with $m_t = 169.5$ GeV although with a further enhancement of masses below 150 GeV. Also, the jet mass deviates significantly from the default setup (that is known to describe CMS data).

In order to exclude effects from the ME–PS interface, Pythia 8 and Herwig++ have been run in standalone mode using their inbuilt LO matrix elements (block 5 in Table 10.1, Figure 10.8). Although the p_{T} spectrum in Pythia 8 is too hard, the invariant mass peaks agree well with the default simulation, as long as ME corrections are activated. The jet mass in Herwig++ is again too low, while the reconstructed masses move closer to the default by about 1 GeV.

As shown in block 6 of Table 10.1 and Figure 10.9, the Powheg NLO generator can be interfaced with Pythia 6/8 and Herwig 6/++. While the Pythia samples show good agreement among each other, different versions and tunes of Herwig induce large shifts in masses of jets, the W boson, and the top-quark candidate. There is unexpectedly large sensitivity to the Herwig++ tunes that differ in their settings for underlying event and color reconnection. The effect of Herwig 6 on m_t^{2D} is nearly completely driven by the shift in W mass which is larger than in all other setups.

Block 7 in Table 10.1 and Figure 10.10 show Sherpa samples created with a $t\bar{t} + 1$ jet matrix element, interfaced to the inbuilt parton shower model. For both the Lund String and the Cluster hadronization model, the W mass distribution is significantly wider and shifted to higher masses, while the m_t distribution is also wider but shifted to smaller values. As the jet mass shows decent agreement, the cause of the deviation is unknown. A peculiarity is the falloff in pseudorapidity of the untagged jets that is not visible for the bottom jets, and might contradict data. The reconstructed top-quark mass differs by 0.5 GeV between the hadronization models, while the W mass is stable. Investigation of the B hadron properties reveals that the Cluster model has a softer b fragmentation and a higher branching ratio to semileptonic decays. I.e., the energy fraction of undetected neutrinos is higher, leading to lower visible energy of the bottom jets. The branching ratio found in the Cluster model is outside the experimentally allowed range given by the Particle Data Group (PDG) [87]. The exact values are summarized in Table 10.2.

Summarizing this study, the different generator setups lead to sizable differences in the invariant masses of the reconstructed heavy objects. For the top-quark mass

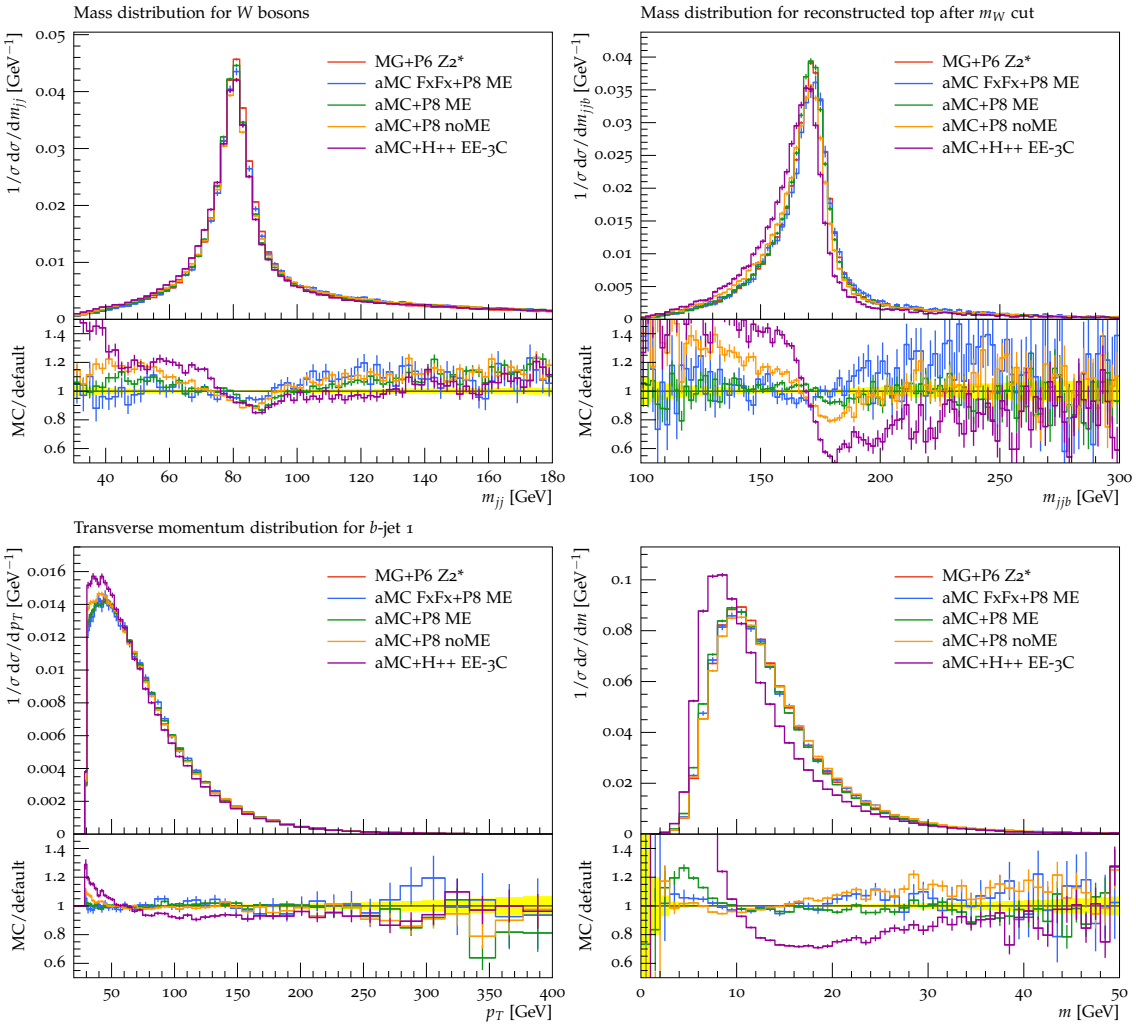

Figure 10.7.: Invariant mass of W-boson (upper left) and top-quark candidates (upper right), and transverse momentum (lower left) and invariant mass (lower right) of the leading bottom jet in simulated $t\bar{t}$ lepton+jets events, comparing the default setup and aMC@NLO. Each distribution is normalized to unity.

Table 10.2.: B hadron properties in Sherpa $t\bar{t}$ simulation, for both hadronization models. The jets are clustered including neutrinos. The first data column shows the mean of the b-fragmentation function, the second column shows the semi-leptonic branching ratio of B hadrons.

Model	$\left\langle p_T^{\mathrm{B}}/p_T^{\mathrm{jet}} \right\rangle$	BR($B \to \ell\nu X$)
Cluster	0.709	0.287
Lund String	0.726	0.247
PDG		0.239 – 0.268

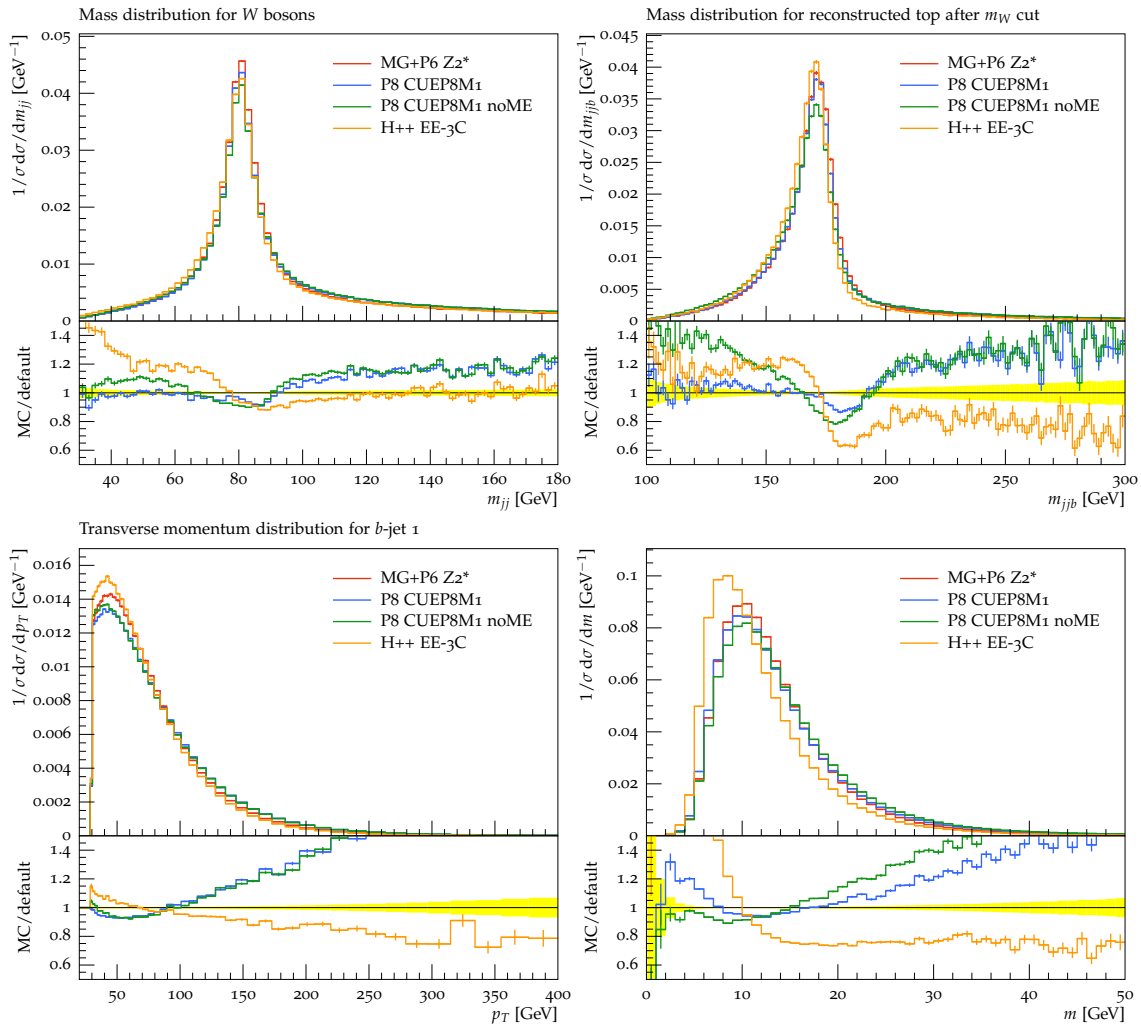

Figure 10.8.: Invariant mass of W-boson (upper left) and top-quark candidates (upper right), and transverse momentum (lower left) and invariant mass (lower right) of the leading bottom jet in simulated $t\bar{t}$ lepton+jets events, comparing the default setup and LO generators. Each distribution is normalized to unity.

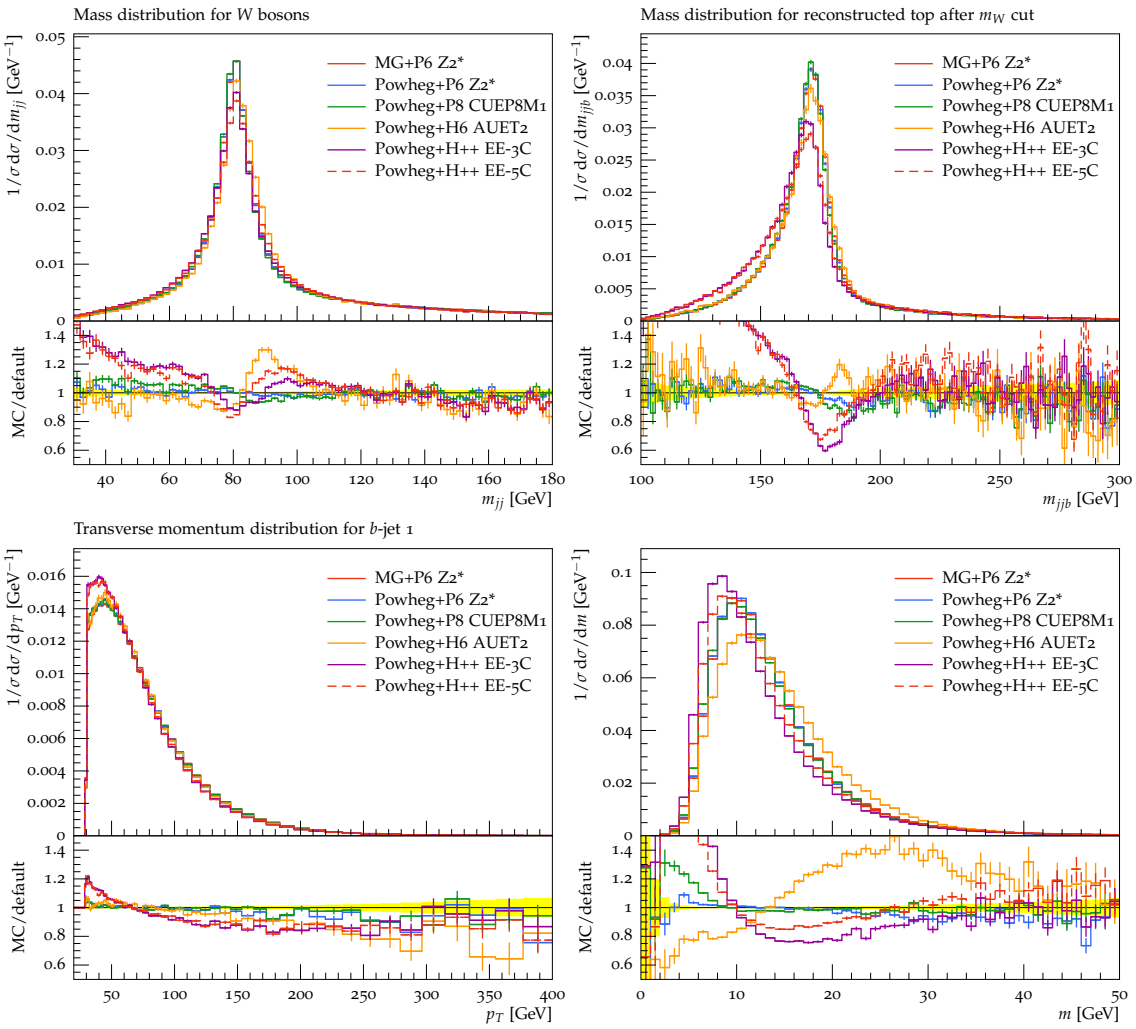

Figure 10.9.: Invariant mass of W-boson (upper left) and top-quark candidates (upper right), and transverse momentum (lower left) and invariant mass (lower right) of the leading bottom jet in simulated $t\bar{t}$ lepton+jets events, comparing the default setup and POWHEG. Each distribution is normalized to unity.

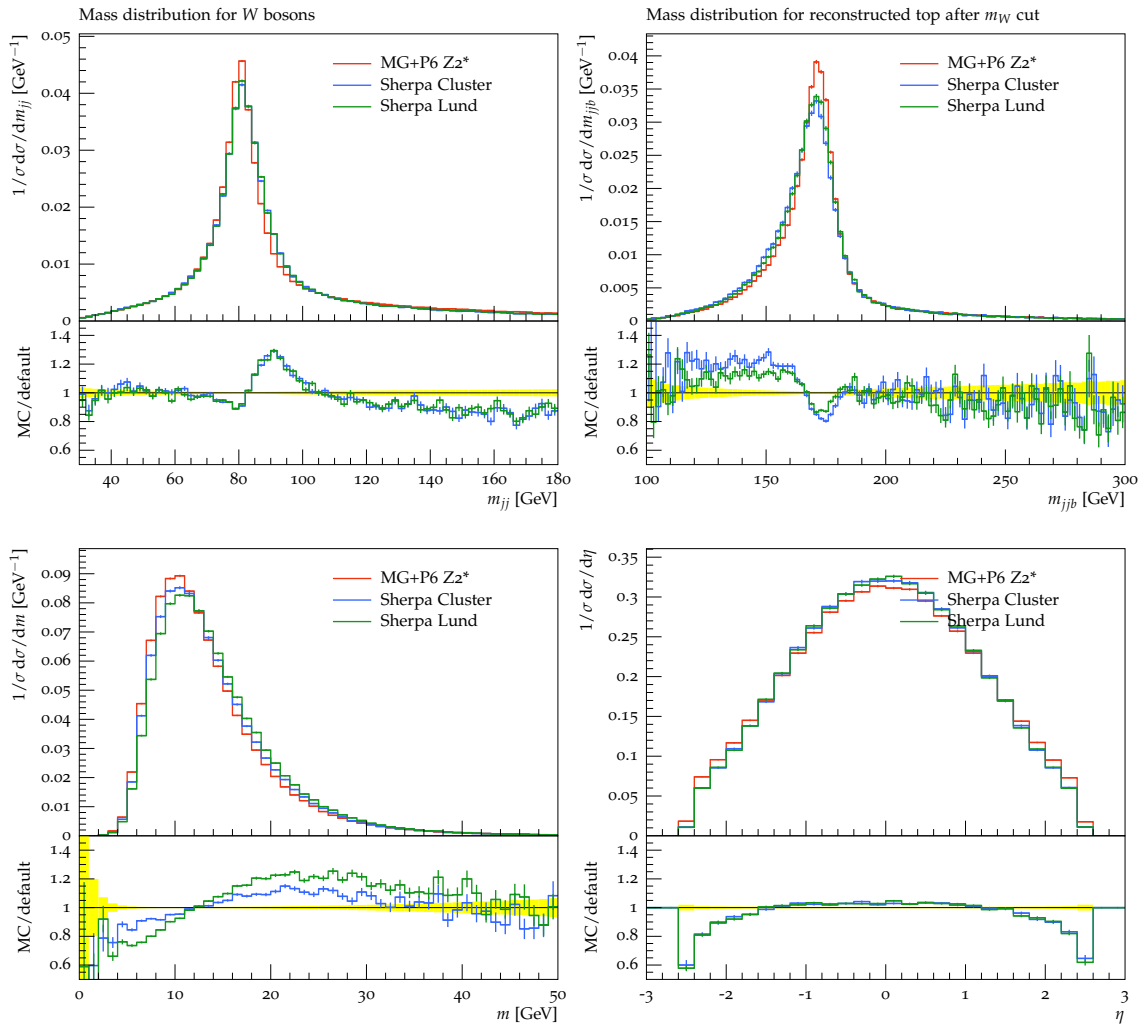

Figure 10.10.: Invariant mass of W-boson (upper left) and top-quark candidates (upper right), invariant mass of the leading bottom jet (lower left), and pseudorapidity of the leading untagged jet (right) in simulated $t\bar{t}$ lepton+jets events, comparing the default setup and SHERPA. Each distribution is normalized to unity.

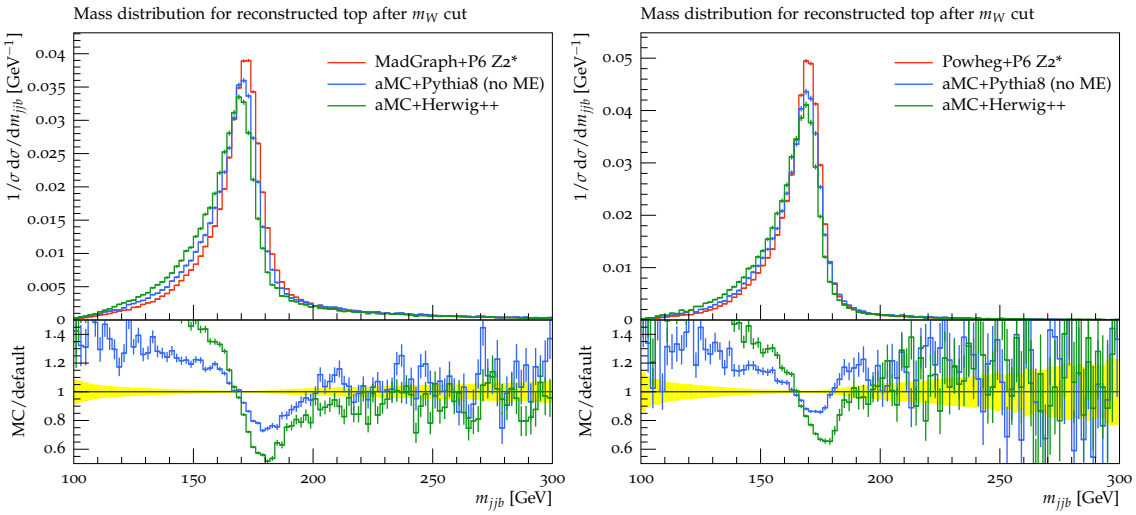

Figure 10.11.: Invariant mass of the reconstructed top-quark candidate in pp collisions at $\sqrt{s} = 7$ TeV (left) and in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV (right) using different generator setups.

in the 2D approach, a mean shift of 0.7 GeV with regard to the default setup is found, where the RMS is 0.9 GeV. Similar values are obtained for the directly reconstructed top-quark candidate mass m_{jjb}. The setups based on PYTHIA 6 and 8 using different tunes are found to be in excellent agreement with each other. A mean shift with respect to the default of 0.03 GeV with a RMS of 0.12 GeV is found.

The disagreement of the invariant mass distributions at particle level may be solved by particle-level measurements and advancements on the theory side in the future. In addition, the anticipated translation of the MC masses to well-defined field theory mass schemes (cf. Section 2.2) needs to be performed for each generator, possibly absorbing the observed differences. The top-quark mass result presented in this thesis is measured in the framework of MADGRAPH + PYTHIA 6, and an approximation of the top-quark mass for other generator setups can be obtained by subtracting the shifts documented in Table 10.1.

It should be noted that similar shifts in the particle-level masses are observed at $\sqrt{s} = 7$ TeV and in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, as shown in Figure 10.11. No MADGRAPH sample was available at $\sqrt{s} = 1.96$ TeV, therefore POWHEG is used as a reference there. The PYTHIA 8 configuration does not contain matrix-element corrections to the resonance decays.

10.2. Comparison of Cluster and Lund fragmentation

The impact of different hadronization models is further studied using SHERPA 2.1.0. Top-quark pair production with up to one additional jet with $p_{\mathrm{T}} > 20$ GeV is described by the LO matrix element calculation at $\sqrt{s} = 8$ TeV. After the p_T-ordered parton shower (CSSHOWER++), the events are passed either to the in-built Cluster fragmentation model (AHADIC++) and the SHERPA decay module (HADRONS++), or an interface to PYTHIA 6.4.18 for Lund String fragmentation and decays.

The decay products of the top quark and the additional jet are stored in the event record as they were calculated from the matrix element, from here on called "parton level". This serves as a common reference point for both hadronization models, as the subsequent parton shower is identical.

At particle level, jets (or "GenJets" in CMS jargon) are clustered using the anti-k_{T} algorithm with R=0.5 from all stable particles, including leptons and particles from the underlying event. In contrast to the Rivet analysis described in the previous section, neutrinos are explicitly included in the jet clustering. This removes the effects from the different decay tables found before (cf. Table 10.2). The matrix-element partons (hard quarks and gluons) are then matched to jets if they are within a distance in $\eta - \phi$ space $\Delta R < 0.5$. Using the matched pairs, a jet-parton response is defined as $p_{\mathrm{T}}\,(\mathrm{jet})\,/p_{\mathrm{T}}\,(\mathrm{parton})$. The ratio of the responses found in Cluster and String fragmentation is shown in Figure 10.12. The difference found for the quark/gluon mixture of $t\bar{t}$ events is below 0.3%. The differences relative to the $t\bar{t}$ mixture are at the level of 0.05% for gluon jets, and completely negligible for jets stemming from quarks.

The W-boson and top-quark masses reconstructed using the matching information are shown in Figure 10.13. All jets need to fulfill $p_{\mathrm{T}} > 30$ GeV. For the top-quark reconstruction, a W-boson mass between 70 and 90 GeV is required. The masses reconstructed at particle level agree within 10 MeV for both hadronization models when neutrinos are included in the jet clustering.

This study demonstrates that the difference between SHERPA with Cluster and Lund fragmentation found in Table 10.1 is mainly due to the different semi-leptonic branching ratios of B hadrons. No evidence for further effects due to hadronization model is found.

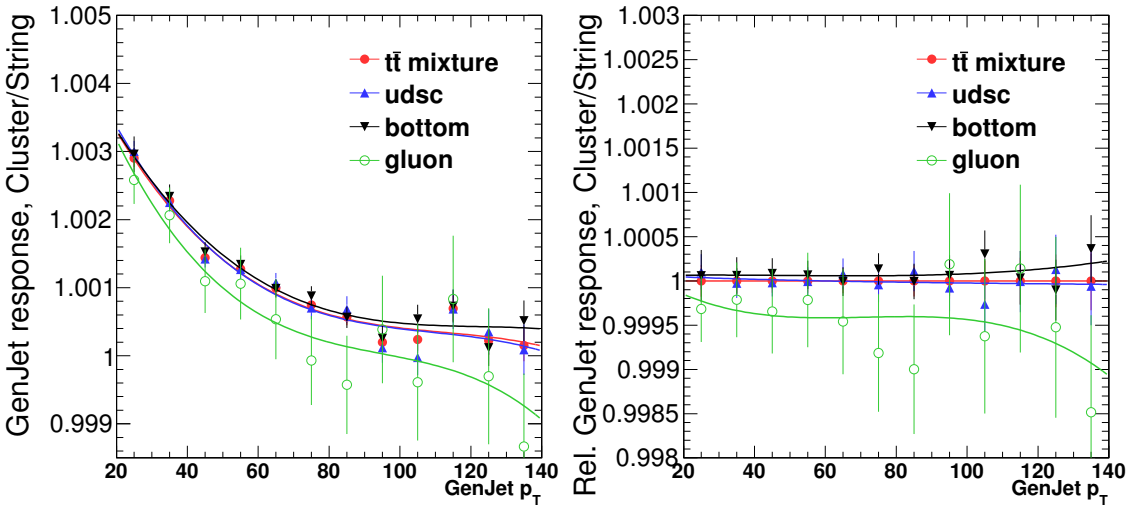

Figure 10.12.: The double-ratio of jet-parton response p_{T} (jet) /p_{T} (parton) between Cluster and String fragmentation (left) and relative to the $t\bar{t}$ mixture (right), as a function of the generated jet p_{T}.

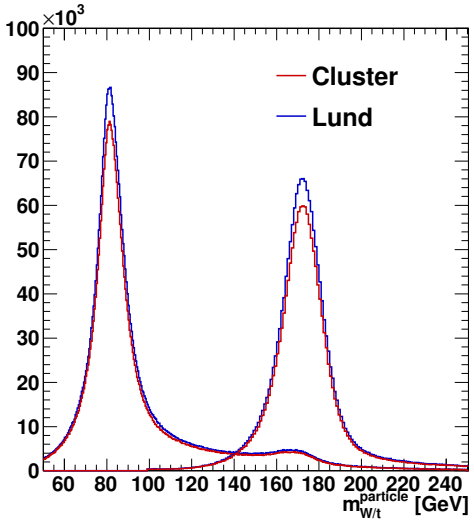

Figure 10.13.: Reconstructed mass distributions at particle level of W boson (left peak) and top-quark (right peak) candidates matched to parton information. The normalization is slightly different so that both distributions are visible.

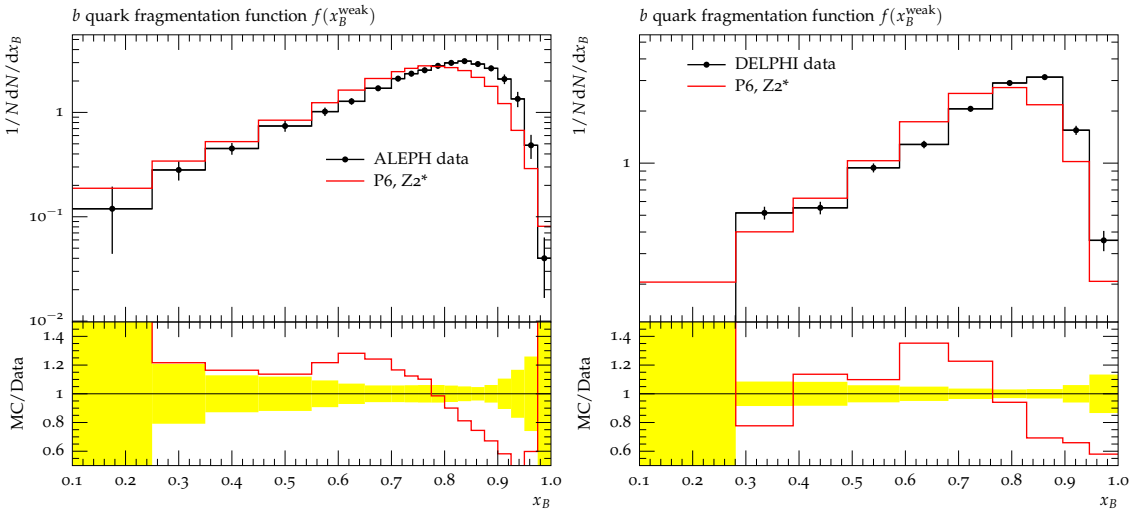

Figure 10.14.: Fragmentation functions measured by ALEPH (left) and DELPHI (right), compared to PYTHIA 6 Z2*.

10.3. Tuning of the b-fragmentation function

The following study tries to improve the tuning of the bottom fragmentation in PYTHIA 6 for estimating the systematic uncertainty on the top-quark mass. In order to avoid an impact on the light flavors, only the Bowler parameter for bottom fragmentation r_b is varied.

The fragmentation functions are tuned to $x_B = E_B/E_{beam}$ in $Z \to b\bar{b}$ events, where B denotes the weakly decaying B hadron. This distribution was measured by experiments at the LEP and SLC colliders at $\sqrt{s} = m_Z = 91$ GeV [152, 153, 193–195]. The PYTHIA 6 tune used in CMS is not able to correctly describe the x_B distributions. The b-fragmentation function is too soft, as shown in Figure 10.14.

For bringing the PYTHIA prediction into agreement with the existing x_B data, Professor [89] is used. Professor is based on the Rivet framework and is a tool for an automated tuning of event generators to data distributions. The parameter r_b (PARJ(47)) is expected to shift the peak of the x_B distributions. As a first step, samples with values of $0.25 < r_b < 1.5$ are created, and the envelopes are plotted in Figure 10.15. Most bins of the data are covered by the envelopes.

For each bin, the value is parametrized as a function of the input parameter and interpolated using a cubic function. Professor then fits the value of r_b that shows the best compatibility with the data, minimizing the $\chi^2 = |\text{MC} - \text{data}|/\sigma$ summed over all bins. As an estimate of the uncertainty, "Eigentunes" are provided by determination of the parameter values that double the χ^2.

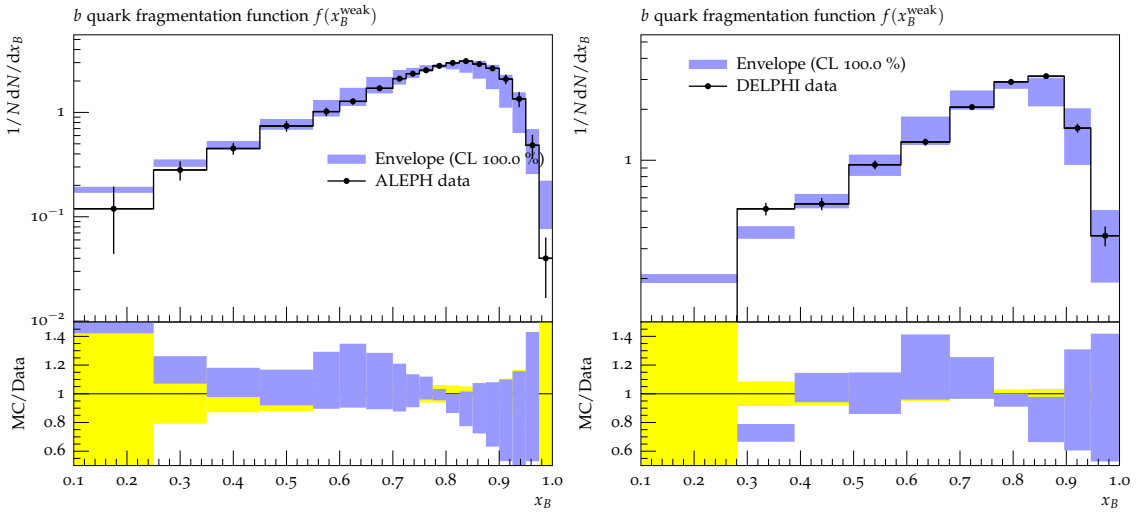

Figure 10.15.: Fragmentation functions measured by ALEPH (left) and DELPHI (right), compared to an envelope of predictions by Pythia.

Using the data from ALEPH and DELPHI, a new tune called Z2*rbLEP is fitted as shown in Table 10.3 and Figure 10.16. Its variations with softer and harder fragmentation cover the tunes obtained from the data of the single experiments (Z2*rbDELPHI, Z2*rbALEPH, and Z2*rbSLD) and a global fit to all of them (Z2*rbAll). The previous value for Z2* is outside this range, while there is good agreement with the tuning of Pythia 8. The best fit is possible with ALEPH data which yields the smallest χ^2/ndf and thus the lowest uncertainty. Inclusion of the DELPHI data increases the χ^2 substantially, mostly due to incompatible values at low x_B in the preliminary DELPHI analysis [193] implemented in Rivet. The published DELPHI data [153] seems to agree well with the ALEPH analysis and is expected to lead to an improved tuning with smaller uncertainties. Unfortunately, the updated DELPHI analysis is not available in the Rivet+Professor framework yet.

The Z2*rbLEP tune obtained in this section is used for evaluating the uncertainty from b fragmentation in all current CMS measurements of the top-quark mass by comparison with the default tune Z2*.

Table 10.3.: Overview of improved PYTHIA 6 tunes, their b-fragmentation parameter and goodness-of-fit. The default value of PYTHIA 6 is used in all common tunes, e.g., Z2* and Perugia 2011.

Tune	r_b	χ^2	Ndf
PYTHIA 6 default (Z2*, P11)	1.0		
PYTHIA 8 default	0.67		
Z2*rbLEP	$0.591^{+0.216}_{-0.275}$	69.0	28
– soft	0.807	138.0	28
– hard	0.316	138.0	28
Z2*rbDELPHI	$0.558^{+0.251}_{-0.326}$	47.9	9
Z2*rbALEPH	$0.626^{+0.166}_{-0.198}$	19.7	18
Z2*rbSLD	$0.716^{+0.182}_{-0.258}$	29.4	21
Z2*rbAll	$0.628^{+0.215}_{-0.286}$	104.5	50

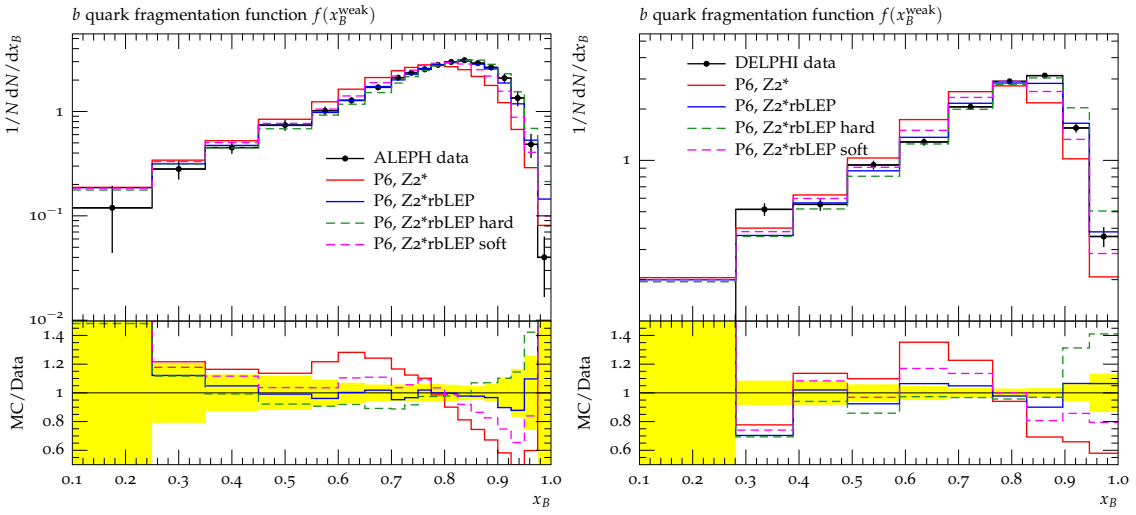

Figure 10.16.: Fragmentation functions measured by ALEPH (left) and DELPHI (right), compared to the new Z2*rbLEP tune and its systematic variations.

11. Phase-space dependence of the measurement

In this study, the dependence of the measured results in the 2D and 1D approaches on the selected phase space is investigated. The main idea is to investigate possible biases in the determination of the top-quark mass in particular regions of the phase space.* Therefore, the presented analysis method is applied to subsets of events or permutations, respectively, defined according to the value of a given kinematic observable after the kinematic fit. Using the large dataset at $\sqrt{s} = 8\,\text{TeV}$ the statistical precision is improved compared to our public 7 TeV analysis [14]. Compared to the preliminary analyses at 7 TeV and 8 TeV [11], every phase-space region is fully calibrated in this thesis.

11.1. Method

In each case, the sample is divided into 3 to 6 subcategories as a function of the value of the kinematic observable. For each bin, all permutations fulfilling the required kinematic property are used. As some observables depend on the jet-quark assignment that cannot be resolved unambiguously, like the p_{T} and η of a reconstructed top-quark, a single event is allowed to contribute to multiple bins. The outer bins are defined to contain no overflow.

Correlated shifts of the measured top-quark mass m_t are expected due to the nature of the systematic uncertainties of the inclusive measurement discussed so far and due to possibly different m_t in data and simulation. Therefore, the mean measured top-quark mass or JSF is subtracted, so that the results are of the form $m_t - \langle m_t \rangle$ and $\text{JSF} - \langle \text{JSF} \rangle$. The mean values $\langle m_t \rangle$ and $\langle \text{JSF} \rangle$ are given by the results of the inclusive measurement on the specific sample.

*The same framework could also be used to search for phase-space regions with particularly small uncertainties on the result. However, this strategy is usually limited by the size of the simulated samples used for calculating the systematic uncertainties.

The statistical uncertainties of data and simulation are corrected for the correlations induced by the subtraction of the mean values. A reduced set of systematic uncertainties is added to the data uncertainties in quadrature. These include the total JEC uncertainty, the flavor-dependent JEC uncertainties, as well as the uncertainties on jet energy resolution, renormalization/factorization scale, and ME-PS matching threshold. The total uncertainties (relative to the inclusive measurement) are usually in the order of $0.5 - 1\,\text{GeV}$.

Two plots are shown for the top-quark mass in the 2D approach in each bin of the kinematic observable under study:

- the kinematic biases $\left(m_t^{\text{2D}} - \left\langle m_t^{\text{2D}} \right\rangle\right)$ using the *default calibration* described in Section 7.2 (left figures)
- the residual biases $\left(m_{t,\text{cal}}^{\text{2D}} - \left\langle m_t^{\text{2D}} \right\rangle\right)$ after the full recalibration of each bin, denoted *bin-by-bin calibration* in the following (right figures)

The corresponding results for the JSF and the mass from the 1D approach, as well as the number of event permutations per category can be found in Appendix A.

Default calibration

Each bin shows the difference between the value of the top-quark mass or JSF in that bin and the result obtained from the full event sample. The observed biases of up to several GeV are caused by shifts of the mean m_t^{fit} and m_W^{reco} distributions in certain phase-space regions with respect to the inclusive sample that was used for derivation of the ideogram templates and the default calibration.

In each plot using the default calibration, the data are compared to the corresponding predictions obtained from simulations using MadGraph + Pythia 6 with the Z2*, P11, and P11noCR tunes, Powheg + Pythia 6 with the Z2* tune, Powheg + Herwig 6 with the AUET2 tune, and MC@NLO + Herwig 6 with the default tune. The double-difference plots in the bottom compare data and the default generator setup MadGraph + Pythia 6 Z2*.

Bin-by-bin calibration

A different sensitivity to the input m_t and JSF is expected when the measurement is not performed on the inclusive sample. Therefore, a calibration to the MadGraph samples with different input m_t and JSF similar to the one described in Sec. 7.2 is

employed for each individual bin. While the inclusive calibration is applied on likelihood level, this bin-by-bin calibration is applied directly to the values extracted from the pseudo-experiments. Therefore, no additional pseudo-experiments are needed, and multiple automated iterations can be performed to ensure that no significant biases between the calibrated and the input values of m_t and JSF remain.

In the bin-by-bin calibrated plots, each data point can be interpreted as the shift of the measured result with regard to the inclusive result presented in Section 9.1. The data points are fitted with linear functions, where the uncertainties are indicated by dashed curves. The linear slope then can be used to quantify the dependence of the measured mass and JSF on the phase space.

11.2. Results

In this section, the differential results are grouped according to their anticipated sensitivity to the systematic uncertainties in the measured top-quark mass. The effects of color connection, initial and final state radiation, and the sensitivity to the kinematics of the jets from the top decay are of specific interest.

Reconstructed top quarks

The measured top-quark mass might depend on the kinematic properties of the top quark as a result of a possibly ill-defined top-quark mass in the simulation or shortcomings in the modeling of color reconnection [29, 196].

The masses measured from top quarks with different transverse momentum and pseudorapidity, shown in Figures 11.1 and 11.2, are expected to have different sensitivity to the color field between the top quarks and color-connected spectator partons. The invariant mass of the $t\bar{t}$ system, shown in Figure 11.3, is highly correlated with the top-quark p_{T} but removes the dependency on the initial transverse boost of the $t\bar{t}$ system.

There is a residual dependence on the transverse momentum of the top quark visible after the bin-by-bin calibration with a slope of $-4.1 \times 10^{-3} \pm 2.4 \times 10^{-3}$, corresponding to a significance of 1.7σ. Although the difference $\left(m_t^{\mathrm{2D}} - \left\langle m_t^{\mathrm{2D}} \right\rangle\right)$ acquires absolute values of up to 10 GeV in the case of $m_{t\bar{t}}$ (Figure 11.3), the simulation follows the general trend in data.

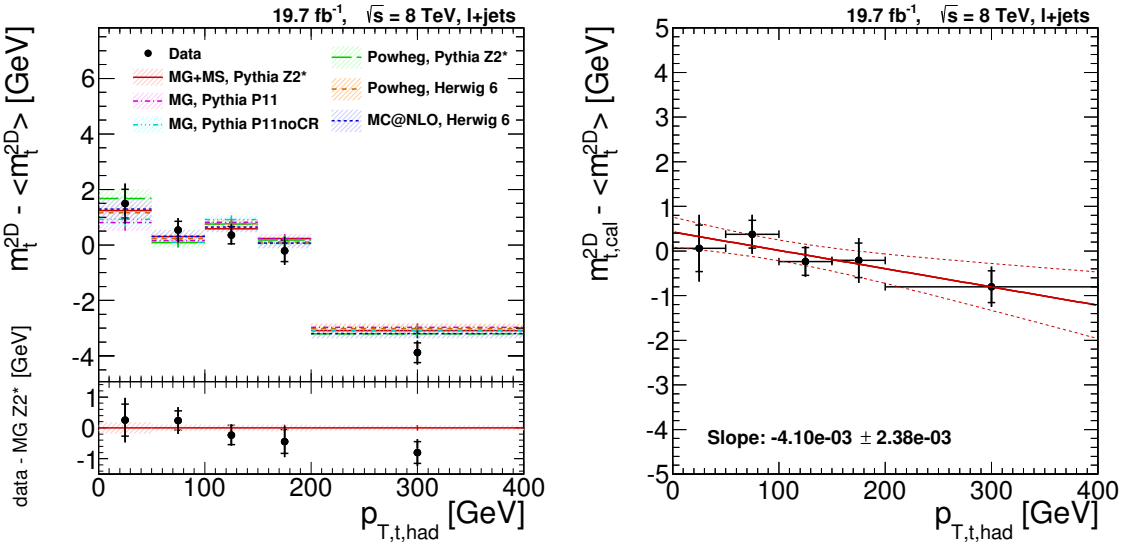

Figure 11.1.: Measurement of m_t^{2D} as a function of the transverse momentum of the hadronically decaying top quark, $p_{\mathrm{T,t,had}}$, before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $p_{\mathrm{T,t,had}}$ is shown in Figure A.2 of the appendix.

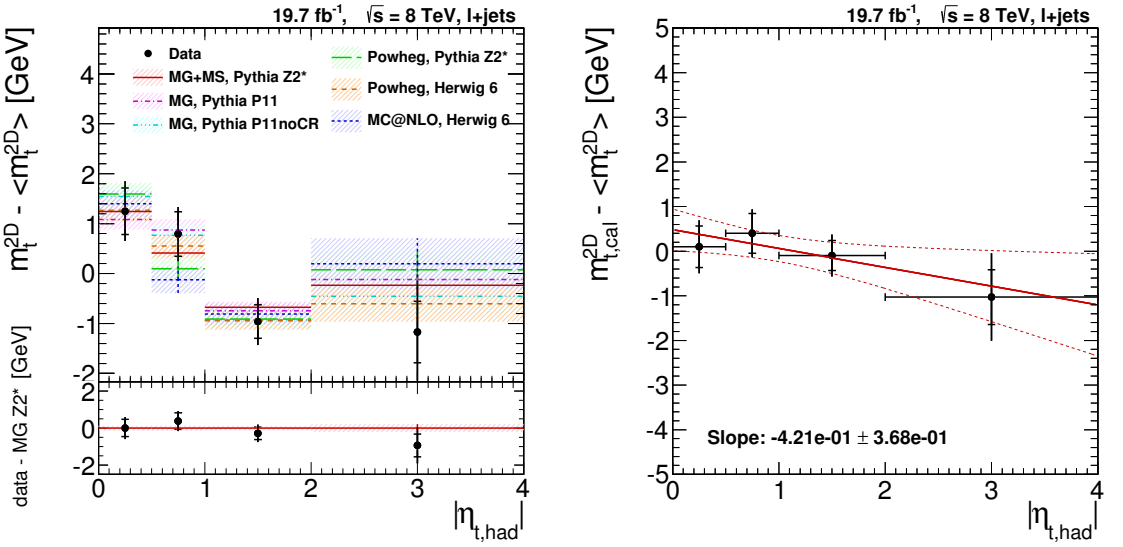

Figure 11.2.: Measurement of m_t^{2D} as a function of the pseudo-rapidity of the hadronically decaying top quark, $|\eta_{\mathrm{t,had}}|$, before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $|\eta_{\mathrm{t,had}}|$ is shown in Figure A.3 of the appendix.

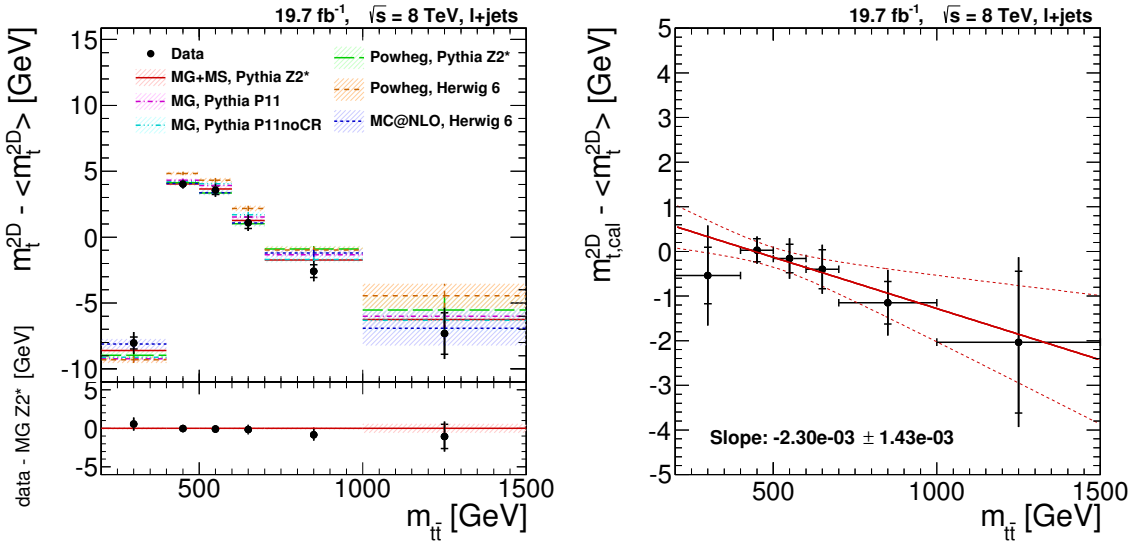

Figure 11.3.: Measurement of m_t^{2D} as a function of the invariant mass of the $t\bar{t}$ system $m_{t\bar{t}}$ before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $m_{t\bar{t}}$ is shown in Figure A.5 of the appendix.

Bottom quarks

The bottom quarks carry the color charge of their parent top quark and are thus color-connected to either initial state radiation or the beam remnants. To test the sensitivity to the b-quark kinematics, the transverse momentum and pseudo-rapidity of the b jet from the hadronic top-quark decay is studied, shown in Figures 11.4 and 11.5.

The p_{T} of the b jet directly enters the invariant mass of the top quark, thus leading to a turn-on behavior that is well described by simulation. The residual dependency has a significance of 1.8σ and likely has the same origin as the one on the top-quark p_{T}. Each bin in Figure 11.5 has a statistical precision similar to twice the complete Tevatron dataset. The Perugia 11 tune shows some deviation at high η but larger simulated samples are needed for clarification.

In addition, the spatial separation of the bottom jets originating from the two top quarks, $\Delta R_{b\bar{b}}$, is shown in Figure 11.6.

Reconstructed W bosons and light quarks

Effects of color reconnection involving the color field between the quarks from the hadronic W decay may show some sensitivity of the top-quark mass against the

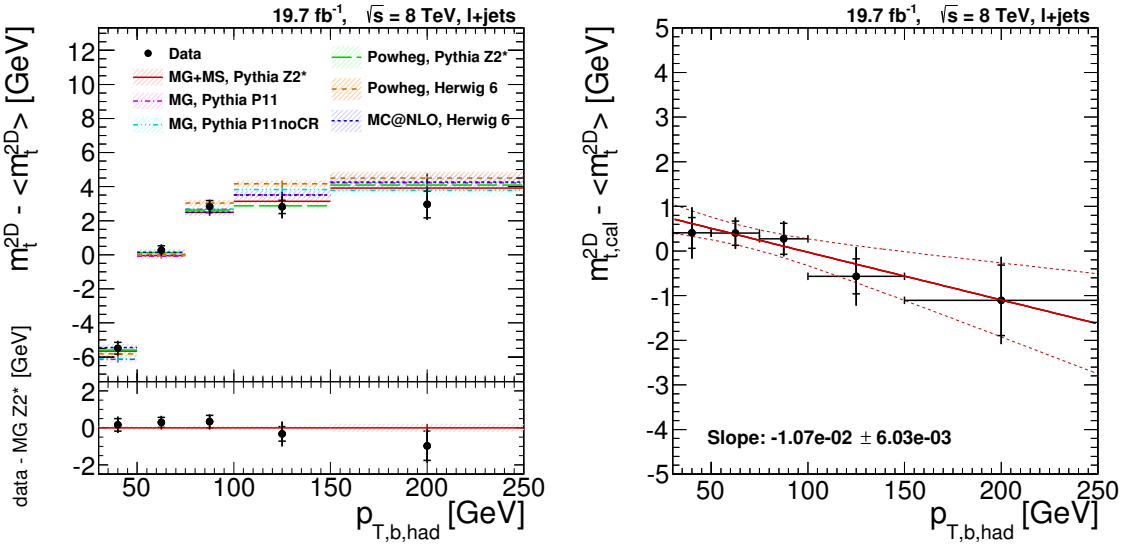

Figure 11.4.: Measurement of m_t^{2D} as a function of transverse momentum of the b jet assigned to the hadronic decay branch, $p_{\mathrm{T,b,had}}$, before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $p_{\mathrm{T,b,had}}$ is shown in Figure A.8 of the appendix.

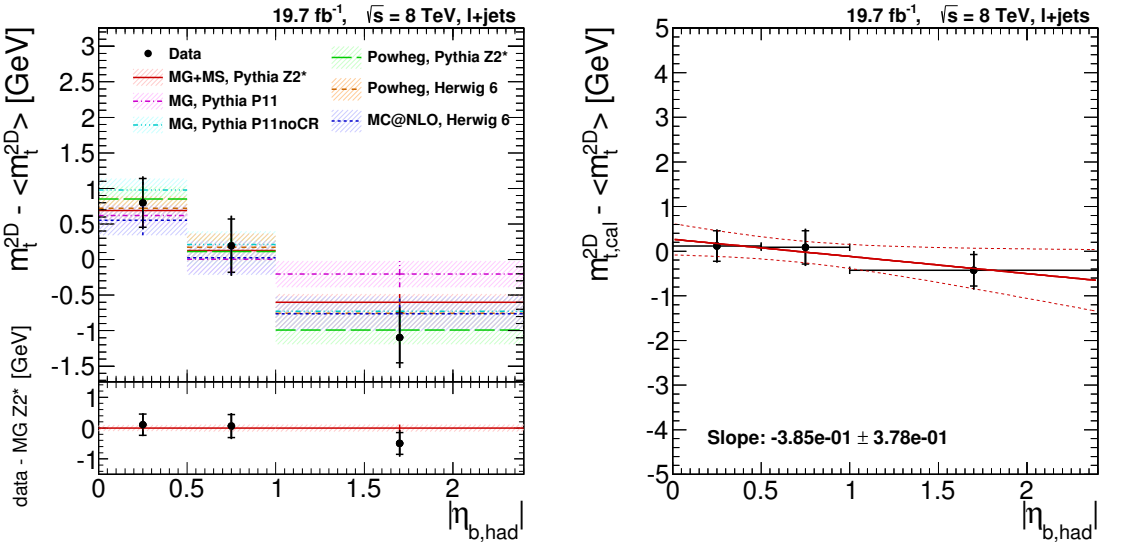

Figure 11.5.: Measurement of m_t^{2D} as a function of the pseudo-rapidity of the b jet assigned to the hadronic decay branch, $|\eta_{\mathrm{b,had}}|$, before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $|\eta_{\mathrm{b,had}}|$ is shown in Figure A.9 of the appendix.

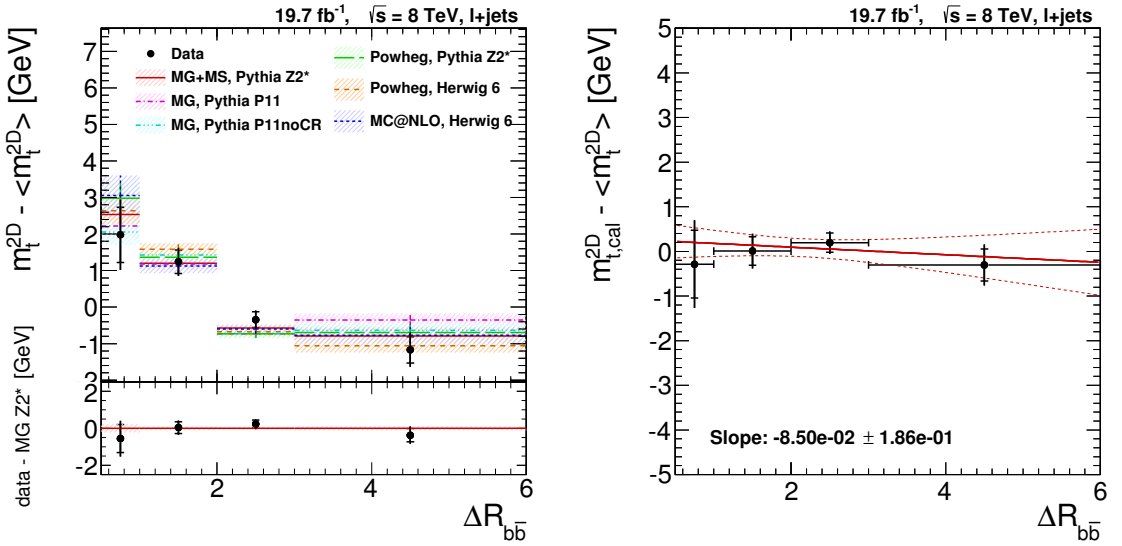

Figure 11.6.: Measurement of m_t^{2D} as a function of the distance between the b jets in η–ϕ–space, $\Delta R_{b\bar{b}}$, before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $\Delta R_{b\bar{b}}$ is shown in Figure A.10 of the appendix.

opening angles between the light jets in terms of the separation in $\eta - \phi$ space, $\Delta R_{q\bar{q}}$, shown in Figure 11.7. The biases cover a range of 15 GeV but are generally well described by the simulation, with the exception of the POWHEG + HERWIG 6 setup that shows a higher (lower) top-quark mass at lower (higher) values of $\Delta R_{q\bar{q}}$, leading to a χ^2/ndf of 10.2/3.

In order to further test the sensitivity to the light-quark kinematics, the transverse momentum and pseudo-rapidity of the leading light jet and the hadronically decaying W boson itself are studied. These are shown in Figures 11.8, 11.9, 11.10 and 11.11, respectively. The dependencies found in data are well described by most simulations. As the transverse momentum of the W boson is highly correlated with the opening angle of its decay products, there are similar deviations found between data and the POWHEG + HERWIG 6 simulation.

Additional radiation

In order to look for effects due to initial and final state radiation (ISR/FSR), the following observables are studied: the multiplicity of jets with $p_{\mathrm{T}} > 30$ GeV, the transverse hadronic energy H_{T}^4, defined as the scalar sum of the p_{T} of the four leading jets, and the transverse momentum of the $t\bar{t}$ system. The results are shown

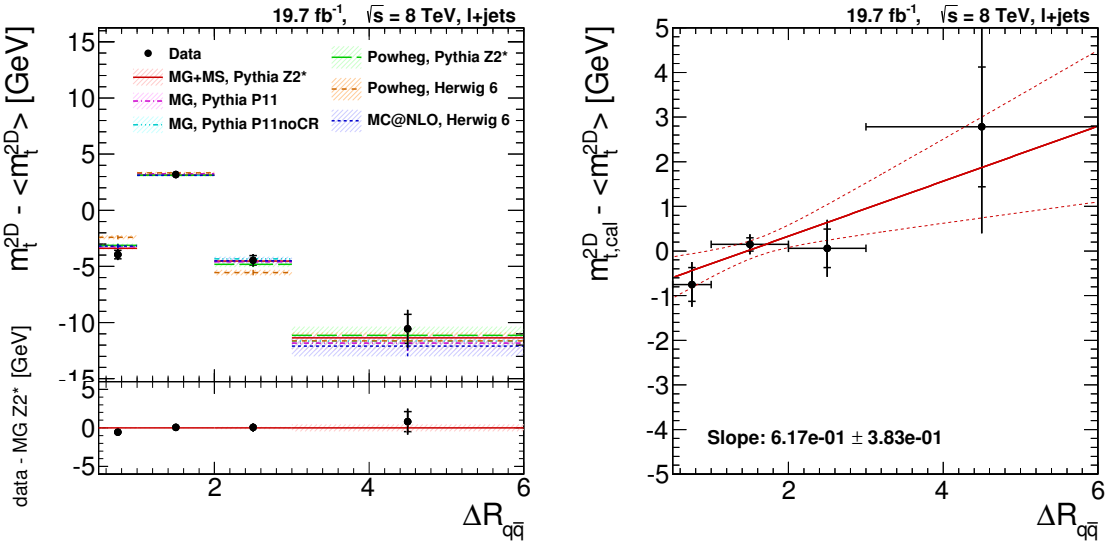

Figure 11.7.: Measurement of m_t^{2D} as a function of the distance between the light-quark jets in η–ϕ–space, $\Delta R_{q\bar{q}}$ before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $\Delta R_{q\bar{q}}$ is shown in Figure A.1 of the appendix.

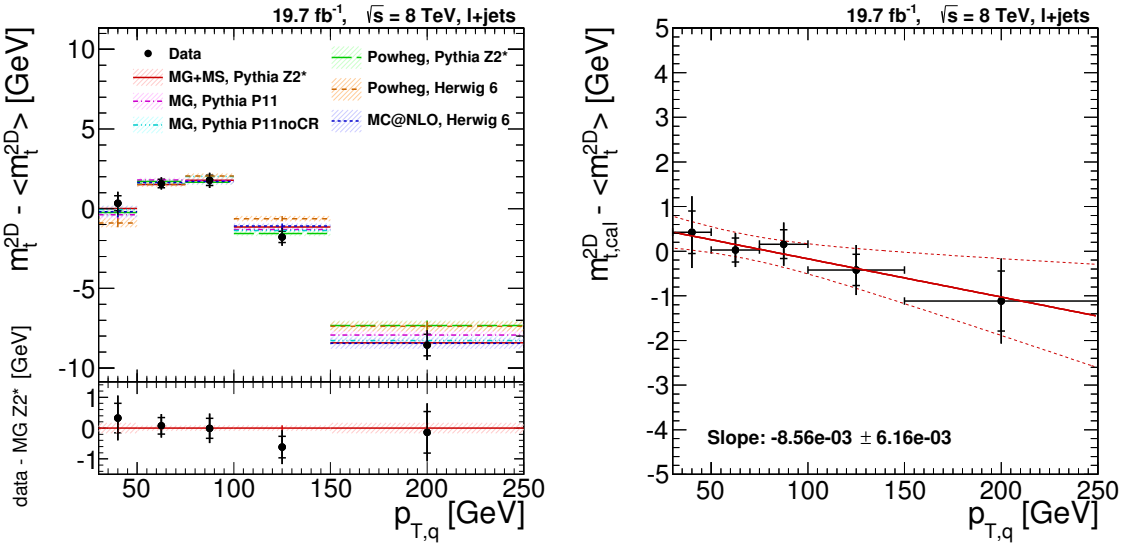

Figure 11.8.: Measurement of m_t^{2D} as a function of the transverse momentum of the leading light jet assigned to the hadronic decay of a W boson, $p_{\mathrm{T,q}}$, before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $p_{\mathrm{T,q}}$ is shown in Figure A.11 of the appendix.

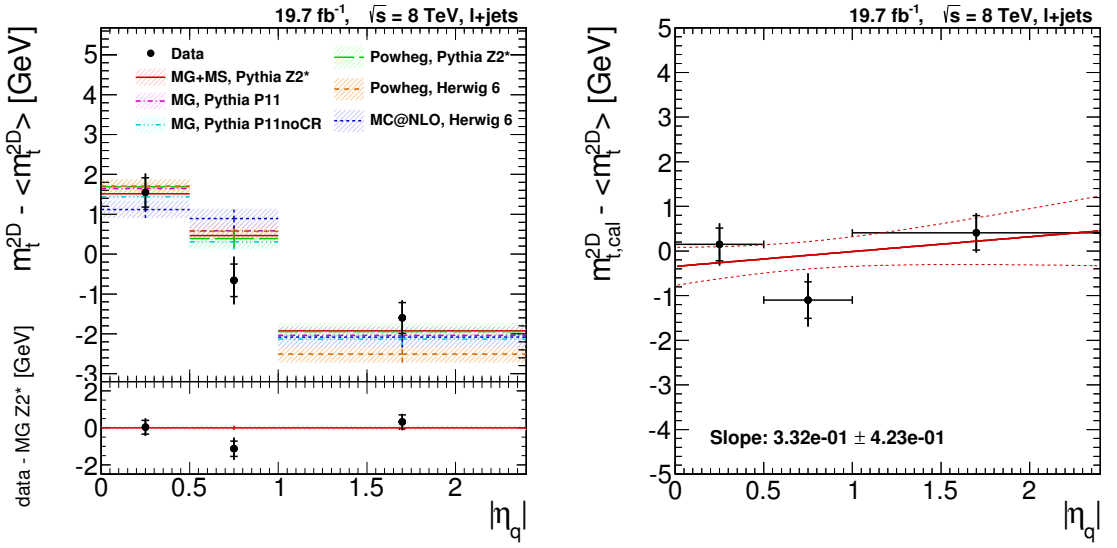

Figure 11.9.: Measurement of m_t^{2D} as a function of the pseudo-rapidity of the leading light jet assigned to the hadronic decay of a W boson, $|\eta_q|$, before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $|\eta_q|$ is shown in Figure A.12 of the appendix.

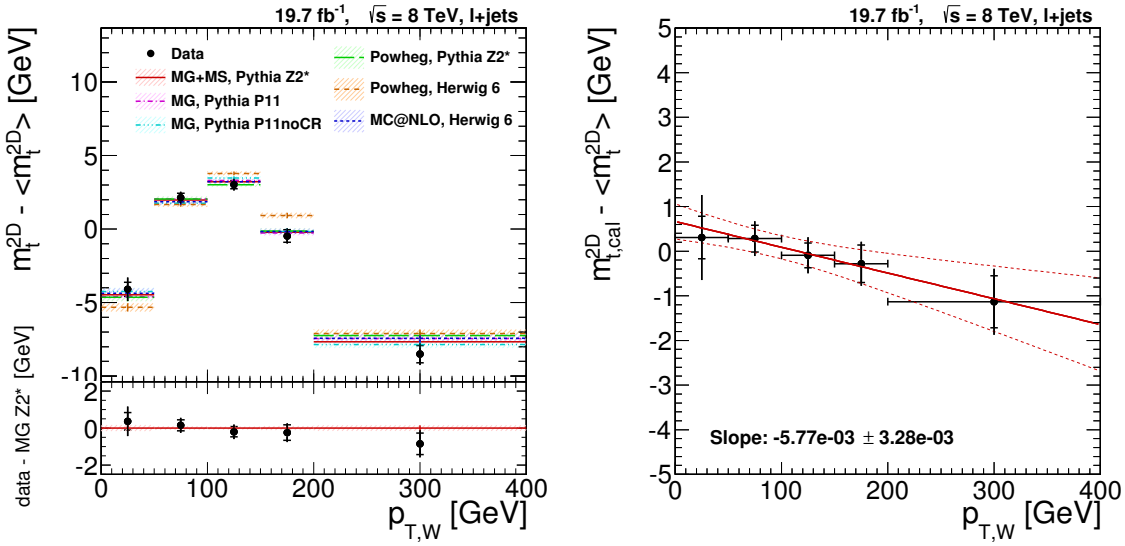

Figure 11.10.: Measurement of m_t^{2D} as a function of the transverse momentum of the hadronically decaying W boson, $p_{T,W}$, before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $p_{T,W}$ is shown in Figure A.13 of the appendix.

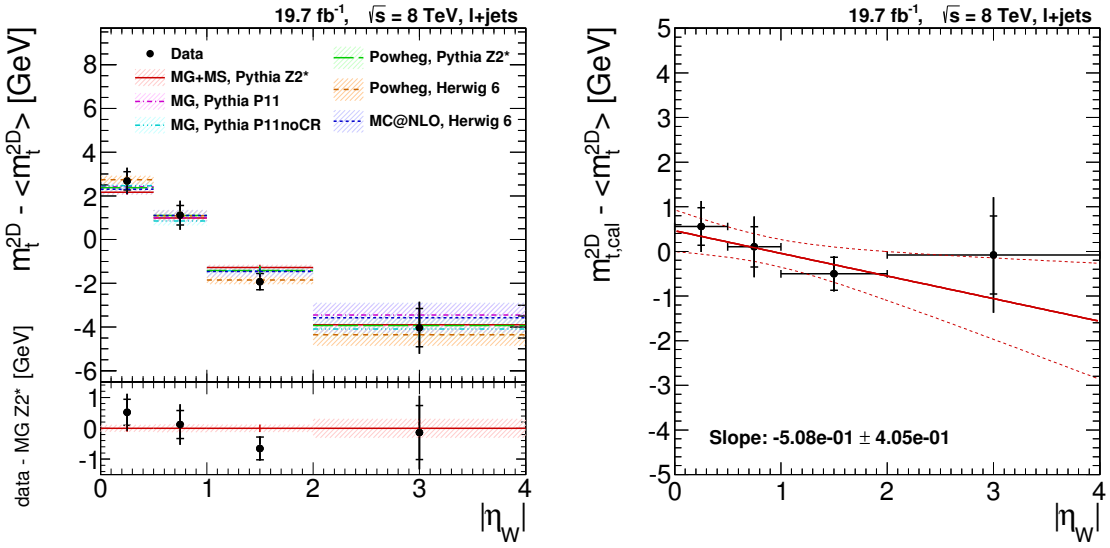

Figure 11.11.: Measurement of m_t^{2D} as a function of the pseudo-rapidity of the hadronically decaying W boson, $|\eta_{\mathrm{W}}|$, before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $|\eta_{\mathrm{W}}|$ is shown in Figure A.14 of the appendix.

in Figures 11.12, 11.13, and 11.14, respectively. For the jet multiplicity both samples showered with HERWIG 6 yield a larger top-quark mass when more jets are present but the significance is low. There is a large dependency of the extracted top-quark mass on H_{T}^4. After bin-by-bin calibration, an effect of about 3.5σ with respect to the default simulation remains, suggesting that the measured top-quark mass drops by 1 GeV with every 100 GeV in H_{T}^4. However, the effect seems to arise from a residual dependency in the JSF measurement from the reconstructed W-boson mass, while the m_t^{1D} result remains stable, as shown in Figure A.4.

Pileup

The same framework is used to test the sensitivity of the measurement on pileup, as shown in Fig. 11.15. There is a visible effect on the measured top-quark mass of about 0.2 GeV per reconstructed primary vertex, yielding the pileup uncertainty found for the inclusive measurement. After calibration, the data seem to be slightly more sensitive to pileup than the simulation.

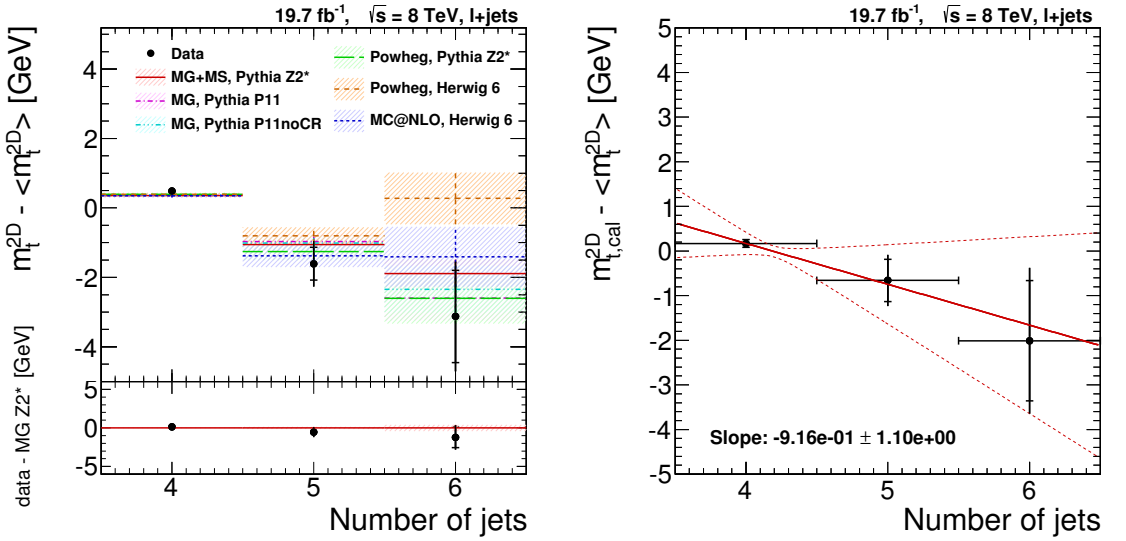

Figure 11.12.: Measurement of m_t^{2D} as a function of number of jets with $p_{\mathrm{T}} > 30$ GeV before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on jet multiplicity is shown in Figure A.7 of the appendix.

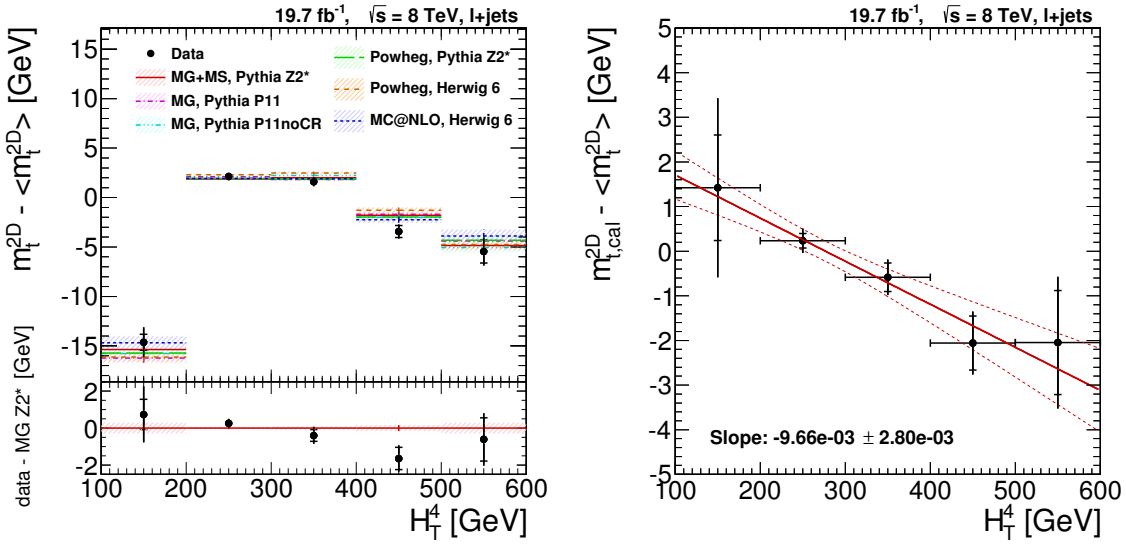

Figure 11.13.: Measurement of m_t^{2D} as a function of H_{T}^4, defined as the scalar sum of the p_{T} of the four leading jets, before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on H_{T}^4 is shown in Figure A.4 of the appendix.

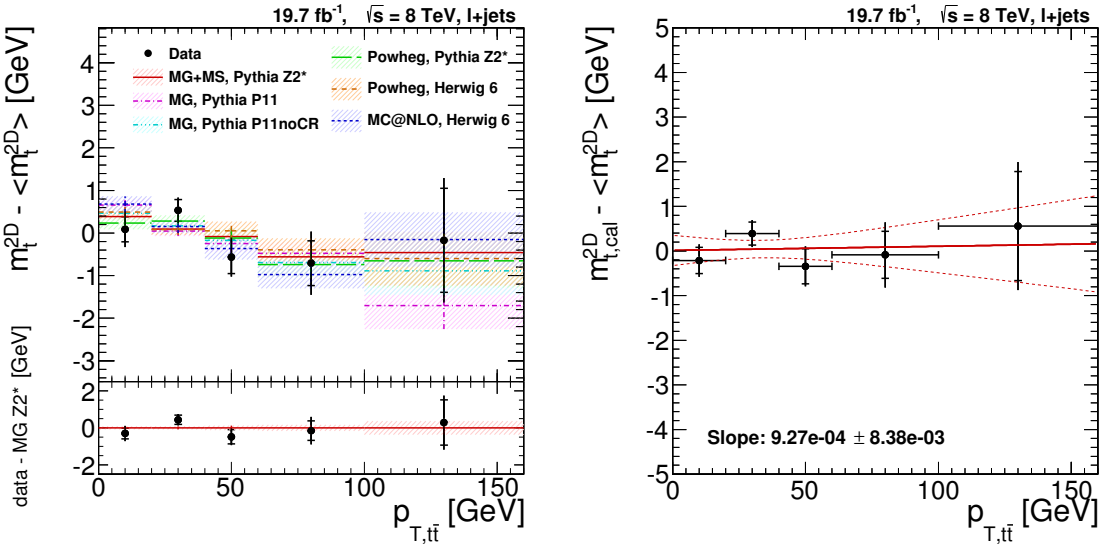

Figure 11.14.: Measurement of m_t^{2D} as a function of the transverse momentum of the $t\bar{t}$ system, $p_{\mathrm{T,t\bar{t}}}$, before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on $p_{\mathrm{T,t\bar{t}}}$ is shown in Figure A.6 of the appendix.

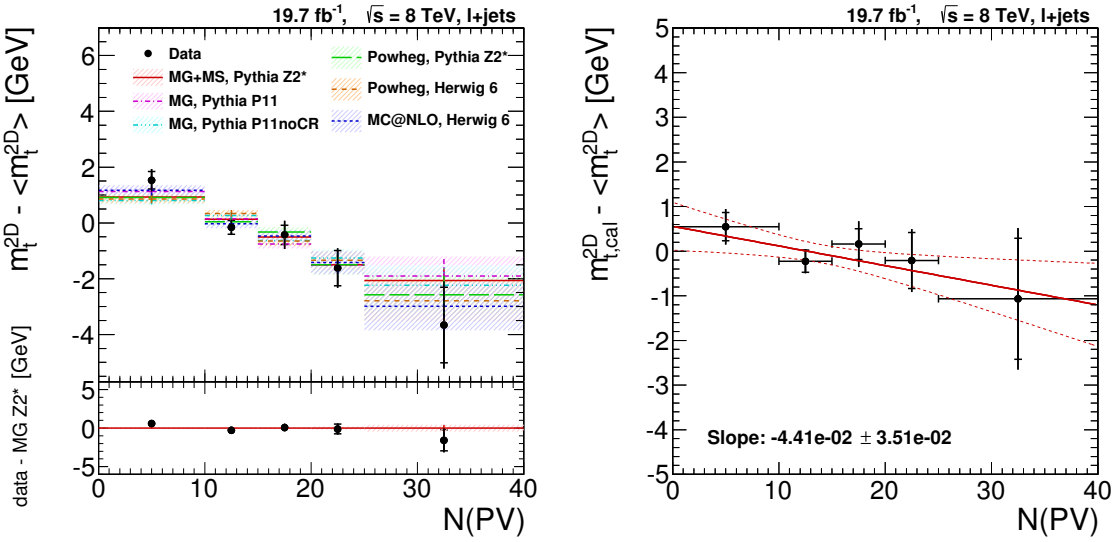

Figure 11.15.: Measurement of m_t^{2D} as a function of the number of reconstructed primary vertices N(PV) before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on N(PV) is shown in Figure A.15 of the appendix.

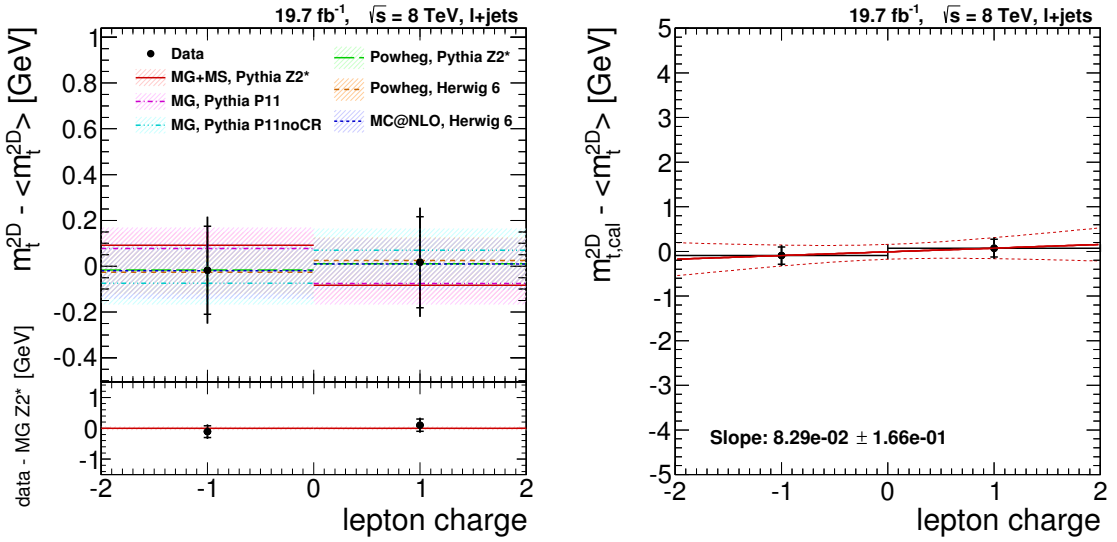

Figure 11.16.: Measurement of m_t^{2D} as a function of the charge of the hard lepton before (left) and after (right) bin-by-bin calibration. The dependence of m_t^{1D} and JSF on the lepton charge is shown in Figure A.16 of the appendix.

Lepton charge

Splitting the sample according to the charge of the hard lepton, as shown in Fig. 11.16, allows for a check of CPT invariance that predicts identical masses for particles and anti-particles. If the selected lepton is negatively (positively) charged, the hadronic decay branch originated from a top quark (anti-top quark). The kinematic fit requires equal top masses for both decay branches but the mass is mainly reconstructed from the hadronic decay branch that has a better resolution than the leptonic one. Therefore, a difference between the mass of the top quark and the anti-top quark should still be visible. The masses agree very well within their statistical uncertainties of 0.2 GeV. The dedicated analyses from ATLAS [197] and CMS [198] also confirm the Standard Model expectation of equal masses.

11.3. Summary

For each measurement of m_t and JSF as a function of a kinematic observable, the level of agreement between data and the $t\bar{t}$ simulation is quantified by a χ^2 defined as $\sum (\text{data - sim})^2 \,/\, (\sigma^2_{\text{data}} + \sigma^2_{\text{sim}})$, as shown in Table 11.1 for the default setup, MadGraph with Pythia 6 Tune Z2*. Within the precision of the current data, the

Observable	m_t^{1D} χ^2	JSF χ^2	m_t^{2D} χ^2	Ndf
$p_{\mathrm{T,t,had}}$	2.99	16.30	4.86	4
$\lvert\eta_{\mathrm{t,had}}\rvert$	7.91	1.02	1.68	3
$p_{\mathrm{T,b,had}}$	3.14	7.30	2.83	4
$\lvert\eta_{\mathrm{b,had}}\rvert$	2.26	0.19	1.35	2
$m_{t\bar{t}}$	1.30	6.29	1.99	5
$p_{\mathrm{T,t\bar{t}}}$	0.52	3.29	3.60	4
$\Delta R_{q\bar{q}}$	2.91	8.93	1.48	3
$\Delta R_{b\bar{b}}$	0.68	1.20	1.69	3
H_{T}^4	0.87	11.12	7.23	4
Number of jets	4.28	0.58	2.73	2
$p_{\mathrm{T,q,had}}^1$	7.99	11.26	1.45	4
$\lvert\eta_{\mathrm{q,had}}^1\rvert$	1.98	3.37	3.89	2
$p_{\mathrm{T,W,had}}$	2.73	13.03	2.05	4
$\lvert\eta_{\mathrm{W,had}}\rvert$	2.43	1.42	3.33	3
Total	41.98	85.30	40.17	47

Table 11.1.: The χ^2 value quantifying the agreement between data and the default simulation for each measurement of m_t and JSF as a function of a kinematic observable.

phase-space dependencies are well described for most observables by the predictions of the simulation.

The summed χ^2 values for different generator setups are compared in Table 11.2. For each sample, the discrepancy to the χ^2 for m_t^{1D}, JSF, and m_t^{2D} are given. Due to the negligible correlation of m_t^{fit} and m_W^{reco}, the measurements of m_t^{1D} and JSF are independent of each other. Therefore, their χ^2 values can be added for the evaluation of the discrepancy between data and simulation in standard deviations. The best overall agreement is found for the MadGraph + Pythia 6 P11noCR setup, while both Powheg simulations deviate by more than 4σ. For for m_t^{2D}, the default setup MadGraph + Pythia 6 Z2* performs best.

It should be noted that jet energy corrections and b-tag scale factors were not derived for Herwig 6, so that the discrepancies seen in the Powheg + Herwig 6 setup may originate either from a different modeling of the top-quark mass distribution (cf. Section 10.1) or from additional experimental biases due to the fact that the calibrations are determined with respect to Pythia 6. In this respect, the relatively good agreement between data and the MC@NLO + Herwig 6 setup is surprising.

Sample	χ^2			Standard deviations	
	m_t^{1D}	JSF	m_t^{2D}	m_t^{1D}+JSF	m_t^{2D}
MG + Pythia 6 Z2*	41.98	85.30	40.17	2.49	0.32
MG + Pythia 6 P11	43.60	67.90	58.14	1.62	1.52
MG + Pythia 6 P11noCR	40.90	55.18	45.14	0.80	0.60
Powheg + Pythia 6 Z2*	53.20	105.96	41.83	4.16	0.40
Powheg + Herwig 6	100.06	63.66	102.26	4.39	4.54
MC@NLO + Herwig 6	46.96	75.94	43.05	2.25	0.47

Table 11.2.: The summed χ^2 values for different generator setups. The number of degrees of freedom is 47, correlations among the kinematic observables are not taken into account.

The good overall agreement found for the sample based on MadGraph + Pythia 6 P11noCR simulation poses no fundamental issue for the inclusive measurement presented in this thesis, as the inclusive masses extracted from the P11noCR sample agree with the default calibration (using Z2*) within $0.10 \pm 0.10\,\mathrm{GeV}$ in the 2D approach, and within $0.05 \pm 0.06\,\mathrm{GeV}$ when the 1D approach is used.

12. Summary

This thesis described the measurement of the top-quark mass in the lepton+jets channel using the ideogram method in a hybrid approach. This approach combines the information from the external jet energy calibration obtained in γ/Z+jet events with a simultaneous (*in-situ*) measurement of a jet scale factor (JSF) from the decay of W bosons to jets. The analysis results in a measured top-quark mass of

$$m_t^{\text{hyb}} = 172.39 \pm 0.16\,(\text{stat.+JSF}) \pm 0.49\,(\text{syst.})\,\text{GeV}$$

in the framework of the PYTHIA parton shower used for the calibration. With an overall uncertainty of 0.52 GeV, the analysis presented in this thesis is the most precise single measurement of the top-quark mass to date.

The results relying completely on either the simultaneous JSF measurement (2D approach) or the external calibration (1D approach) are found to be in good agreement:

$$\begin{aligned} m_t^{\text{2D}} &= 172.15 \pm 0.19\,(\text{stat.+JSF}) \pm 0.61\,(\text{syst.})\,\text{GeV}, \\ m_t^{\text{1D}} &= 172.63 \pm 0.12\,(\text{stat.}) \pm 0.65\,(\text{syst.})\,\text{GeV}. \end{aligned}$$

The precision is achieved by the selection of an extremely pure sample of $t\bar{t}$ events from the large dataset of proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ that was recorded by the CMS experiment in 2012 and corresponds to an integrated luminosity of $20\,\text{fb}^{-1}$. The events are reconstructed using the particle-flow algorithm that combines all available detector information in an optimal way and drastically improves the jet resolution achieved at CMS. Background processes are suppressed by requiring two jets to be tagged as stemming from bottom quarks. A kinematic fit is employed to improve the full reconstruction of the $t\bar{t}$ system.

From the reconstructed events, the ideogram method implemented here extracts the top-quark mass using either the simultaneous determination of a jet scale factor from the reconstructed W-boson mass (*2D approach*), the very precise jet energy

calibration from γ/Z+jet events (*1D approach*), or a combination of them (*hybrid approach*). The results using all three approaches are in agreement with each other. The systematic uncertainties are factorized as much as possible in order to probe the sensitivity of the measurement to single systematic effects and to avoid double-counting.

The high statistical precision obtained in this analysis allows for an evaluation in different phase-space regions, showing that the measurement result is stable against various kinematic observables. The origin of the small residual deviations is not fully understood, both experimental and QCD modeling effects are possible.

Interpretation and outlook

The top-quark mass measured here is the value of the mass parameter in MC simulation needed to recover the observed invariant mass distribution in data. As inferred from the study of top-quark reconstruction on particle level, the invariant mass distribution predicted for the same input top-quark mass differs severely for different generator setups. This implies that the measured top-quark mass is – strictly speaking – only valid in the framework of the Pythia generator.

The differences found between the Pythia, Herwig, and Sherpa simulations at particle level need to be understood in the future. Comparing the different fragmentation models in Sherpa shows good agreement for "perfectly measured" particle-level jets, so that only variations of the jet composition and the momenta of the constituents need to be accounted for as "hadronization" uncertainties. In CMS analyses, this is covered by the flavor-dependent JEC uncertainties based on the different detector responses for jets generated with Pythia 6 and Herwig++, and dedicated variations in the b-quark fragmentation process.

The origin of the differences in reconstructed top-quark or W-boson masses is therefore likely to be found in the radiation modeling of the parton shower. Still, most generator setups are able to correctly describe the number of jets [158,161], gap fractions [161, 191], and jet shapes [166] within the uncertainties of these dedicated measurements. This is a clear motivation for the ATLAS and CMS collaborations to increase the precision of these analyses and to provide additional measurements of different observables in top-quark events. These include, e.g., underlying-event observables [199], jet masses, and properties of fully reconstructed top-quark candidates and the $t\bar{t}$ system at particle level, including the top-quark mass distribution itself. Adding such a complete set of analyses to the Rivet framework allows for

a convenient comparison of data and simulation, and permanently preserves the experimental data.

There are theory efforts to improve the understanding of the PYTHIA mass parameter in the framework of the MSR scheme, requiring a challenging calculation at hadron level [30] using soft-collinear effective theories. Once succeeded, the top-quark mass presented in this thesis can be translated into a well-defined mass scheme with a small uncertainty. It is likely that different conversion parameters will be obtained for the various parton showers, ultimately compensating for the deviations found in the invariant mass distributions at particle level.

In the meanwhile, extractions of the top-quark mass using either the total $t\bar{t}$ production cross-section [35, 38] or alternative observables like the invariant mass of the $t\bar{t}+1$ parton system [200, 201] demonstrate agreement with the mass obtained from top-quark reconstruction. Their precision is typically in the order of 3 GeV but the results have less dependence on the mass scheme implemented in the MC generator.

To conclude, 20 years after the discovery of the top-quark, the standard methods for measuring its mass have reached an uncertainty in the order of 0.5 GeV, as demonstrated in this thesis. This precision raises questions about the exact nature of the top-quark mass measured at hadron colliders, and the relations between well-defined field-theory masses and those implemented in the various MC generators. It is anticipated that these relations will be known in the future, so that the result presented here can be translated into any mass scheme, possibly under consideration of its phase-space dependence. In addition, the experimental collaborations may preserve their data as unfolded mass distributions.

A. Appendix to Chapter 11

This appendix includes the results of m_t^{1D}, JSF and m_t^{2D} as functions of kinematic observables, as described in Chapter 11. In addition, the distributions of the observables under study are displayed.

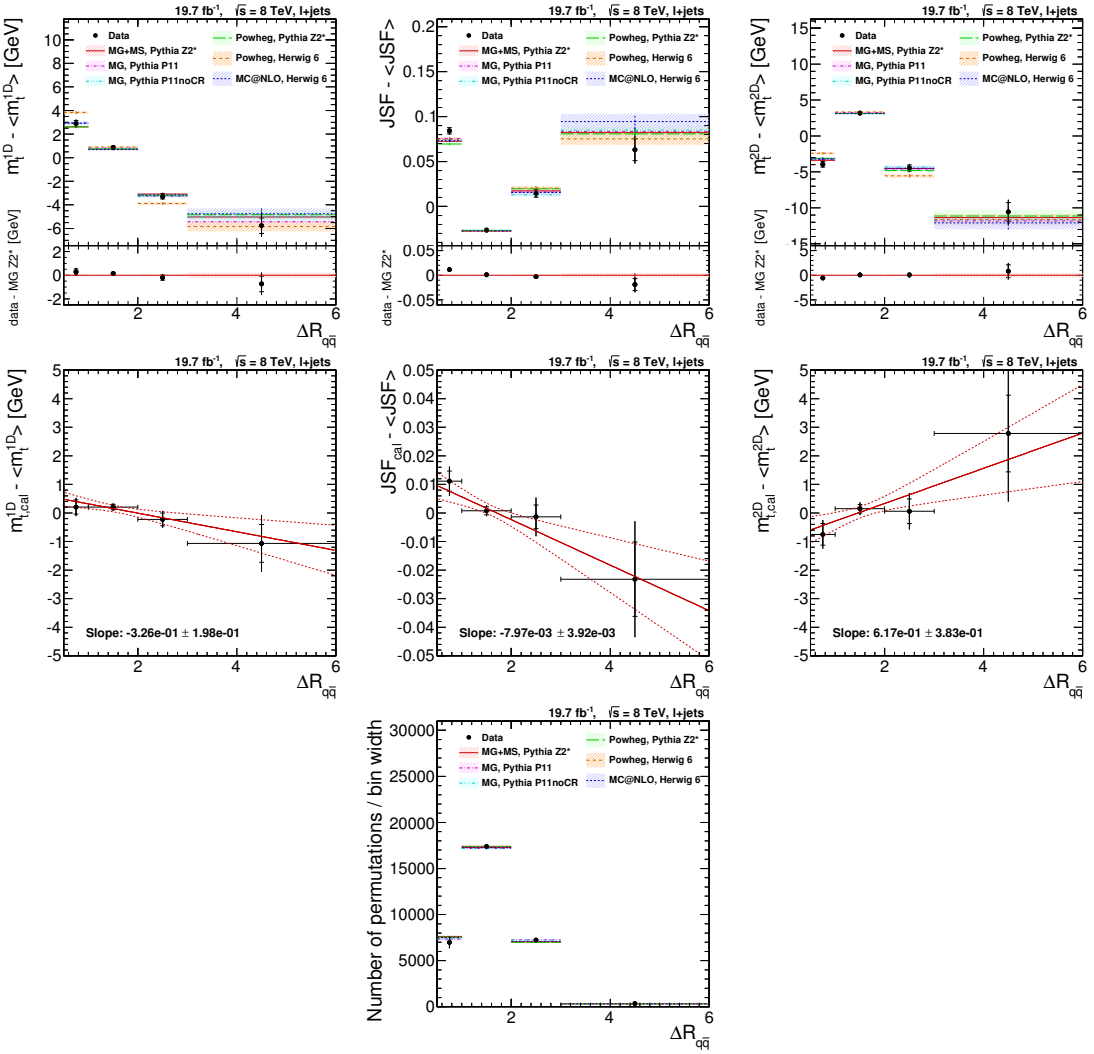

Figure A.1.: Measurement as a function of the ΔR of the light-quark jets: $m_t^{1\mathrm{D}}$ (left), JSF (center) and $m_t^{2\mathrm{D}}$ (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $\Delta R_{q\bar{q}}$ bin, divided by the bin width.

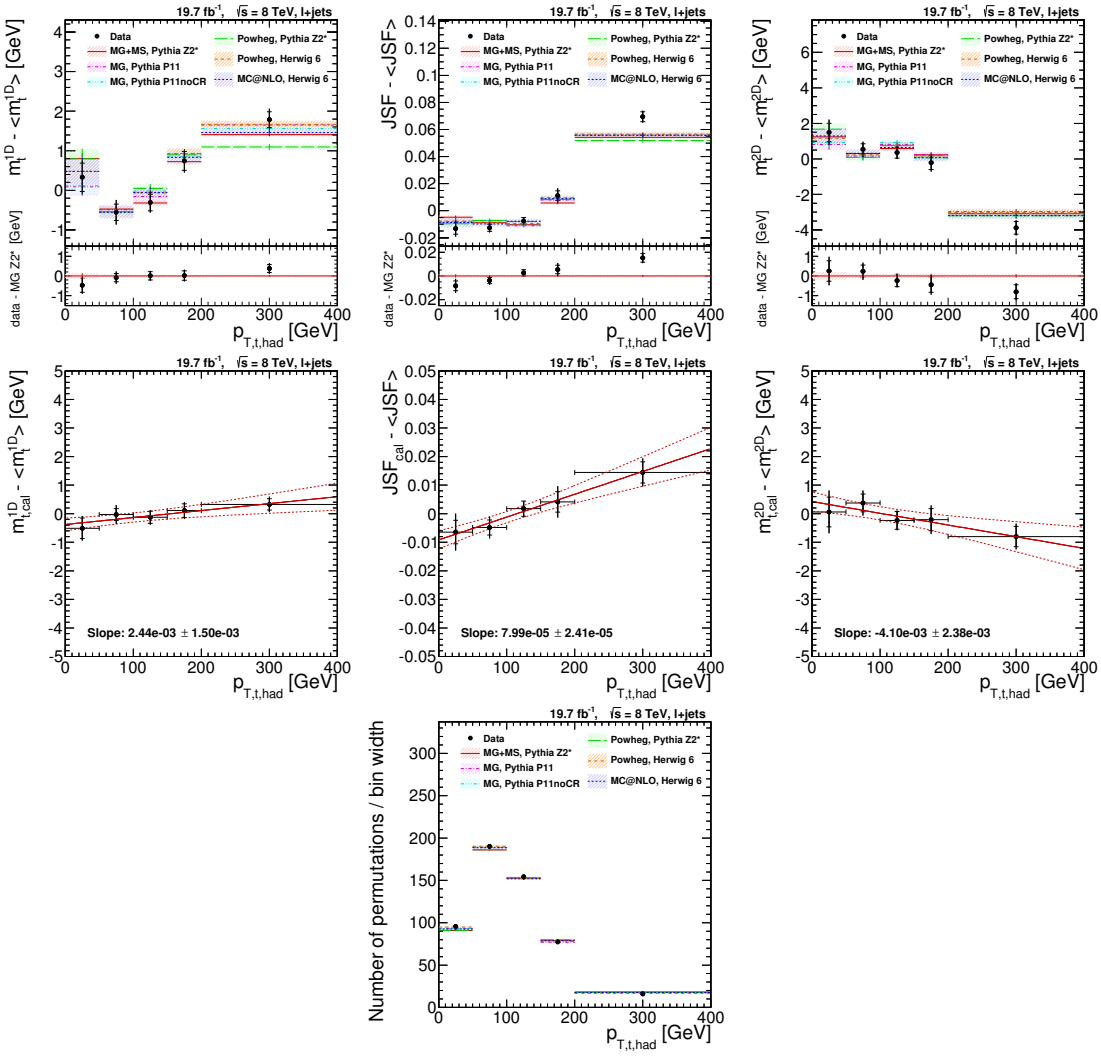

Figure A.2.: Measurement as a function of the transverse momentum of the hadronically decaying top quark: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $p_{\mathrm{T,t,had}}$ bin, divided by the bin width.

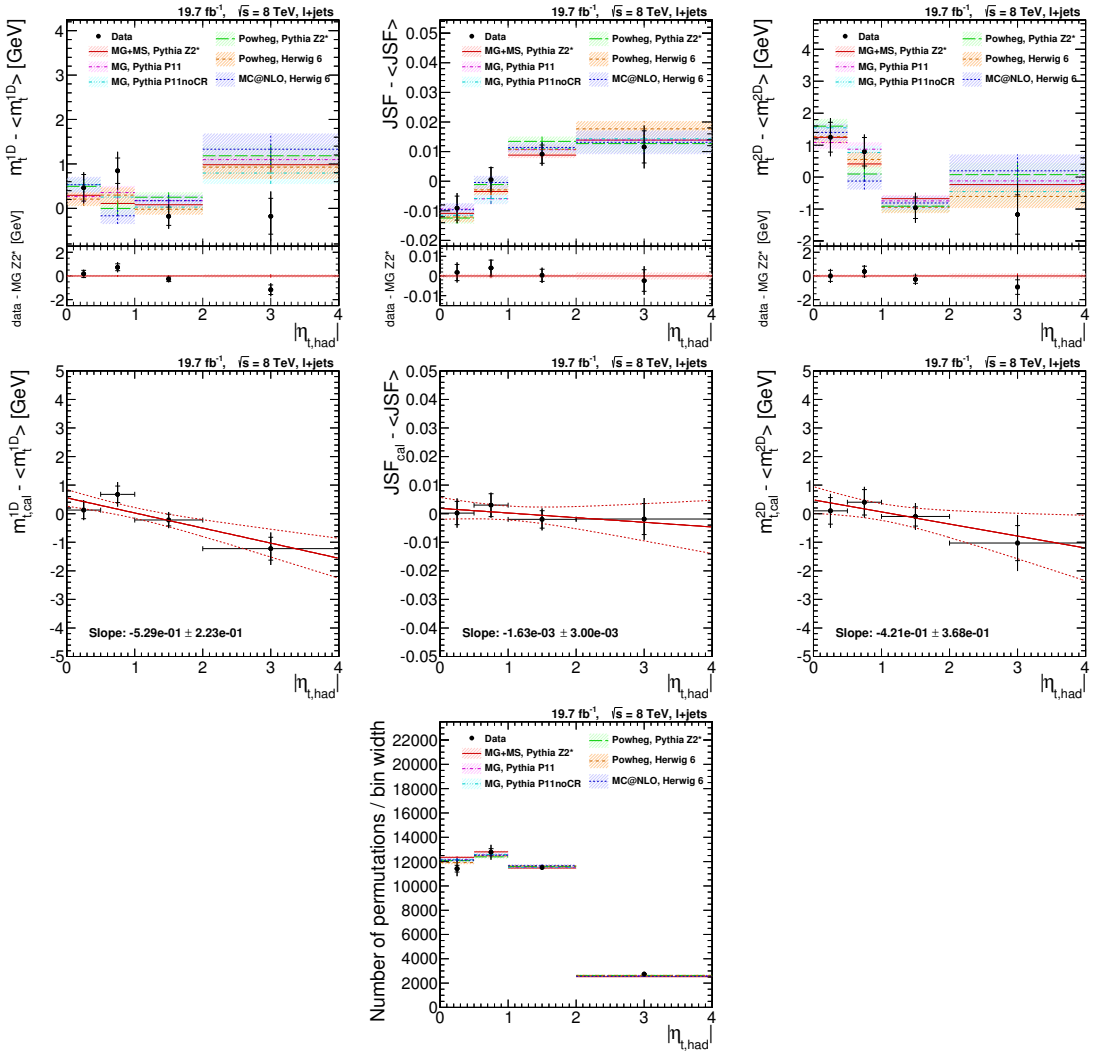

Figure A.3.: Measurement as a function of the pseudo-rapidity of the hadronically decaying top quark: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $|\eta_{\mathrm{t,had}}|$ bin, divided by the bin width.

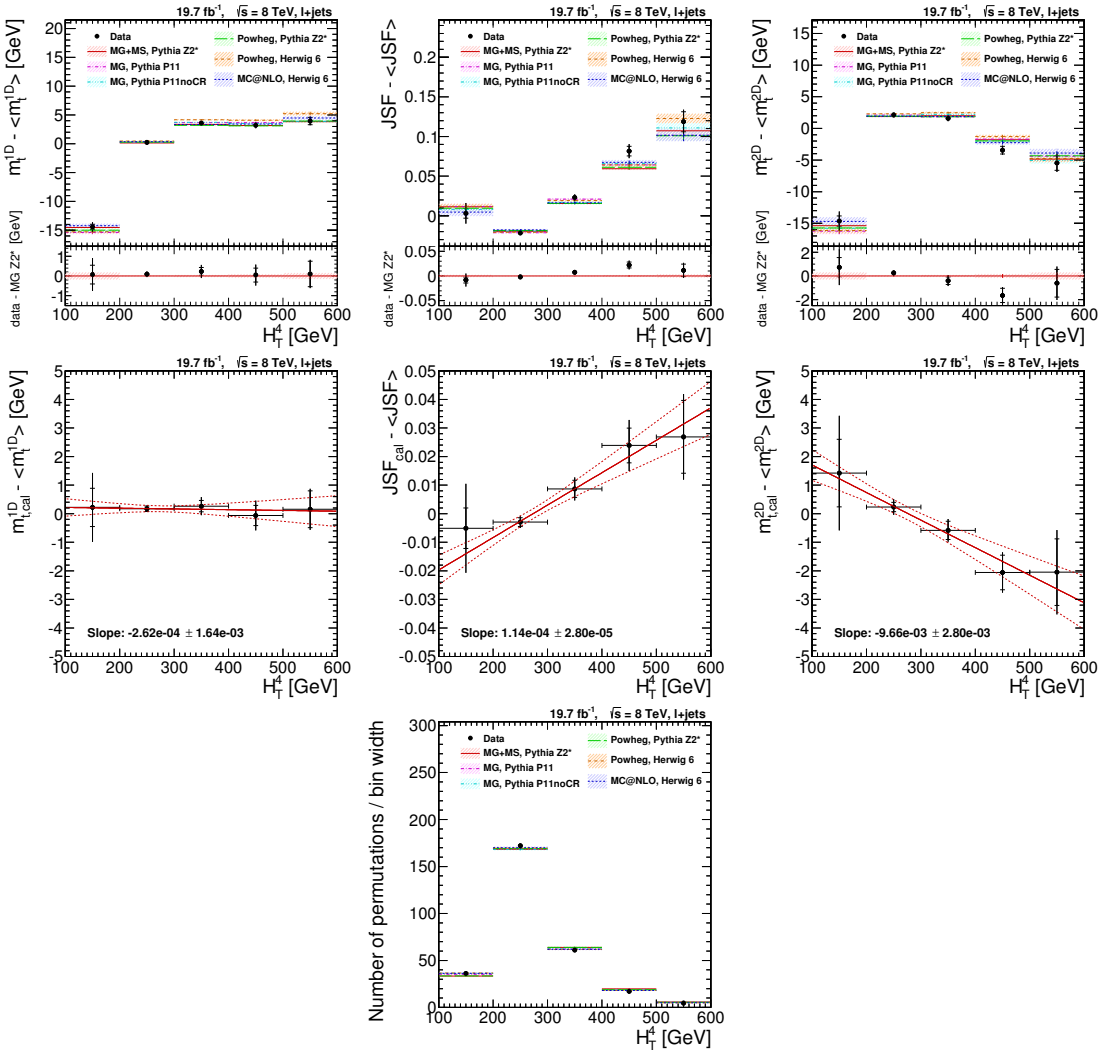

Figure A.4.: Measurement as a function of H_{T}, defined as the scalar sum of the p_{T} of the four leading jets: $m_t^{1\mathrm{D}}$ (left), JSF (center) and $m_t^{2\mathrm{D}}$ (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per H_{T} bin, divided by the bin width.

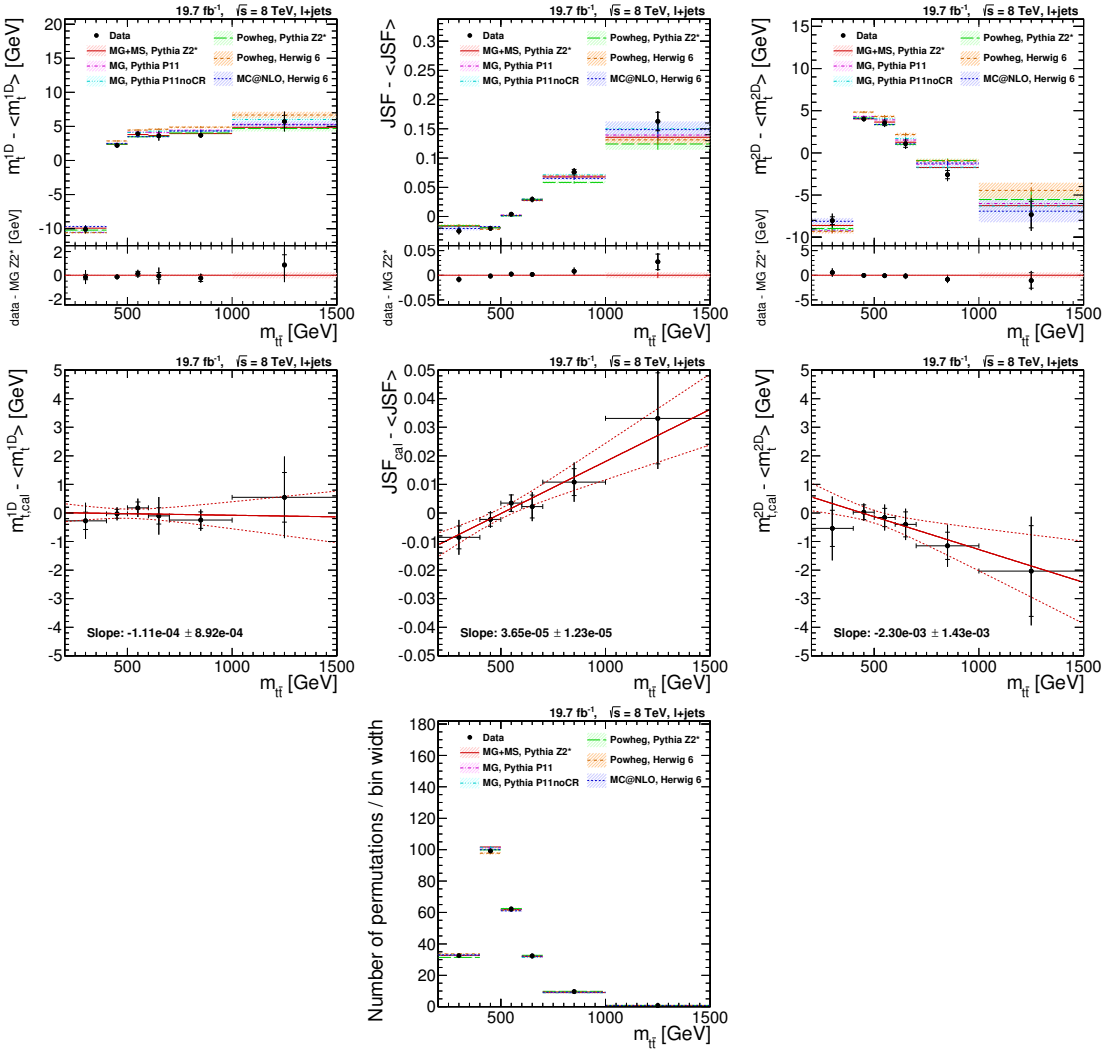

Figure A.5.: Measurement as a function of the invariant mass of the $t\bar{t}$ system: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $m_{t\bar{t}}$ bin, divided by the bin width.

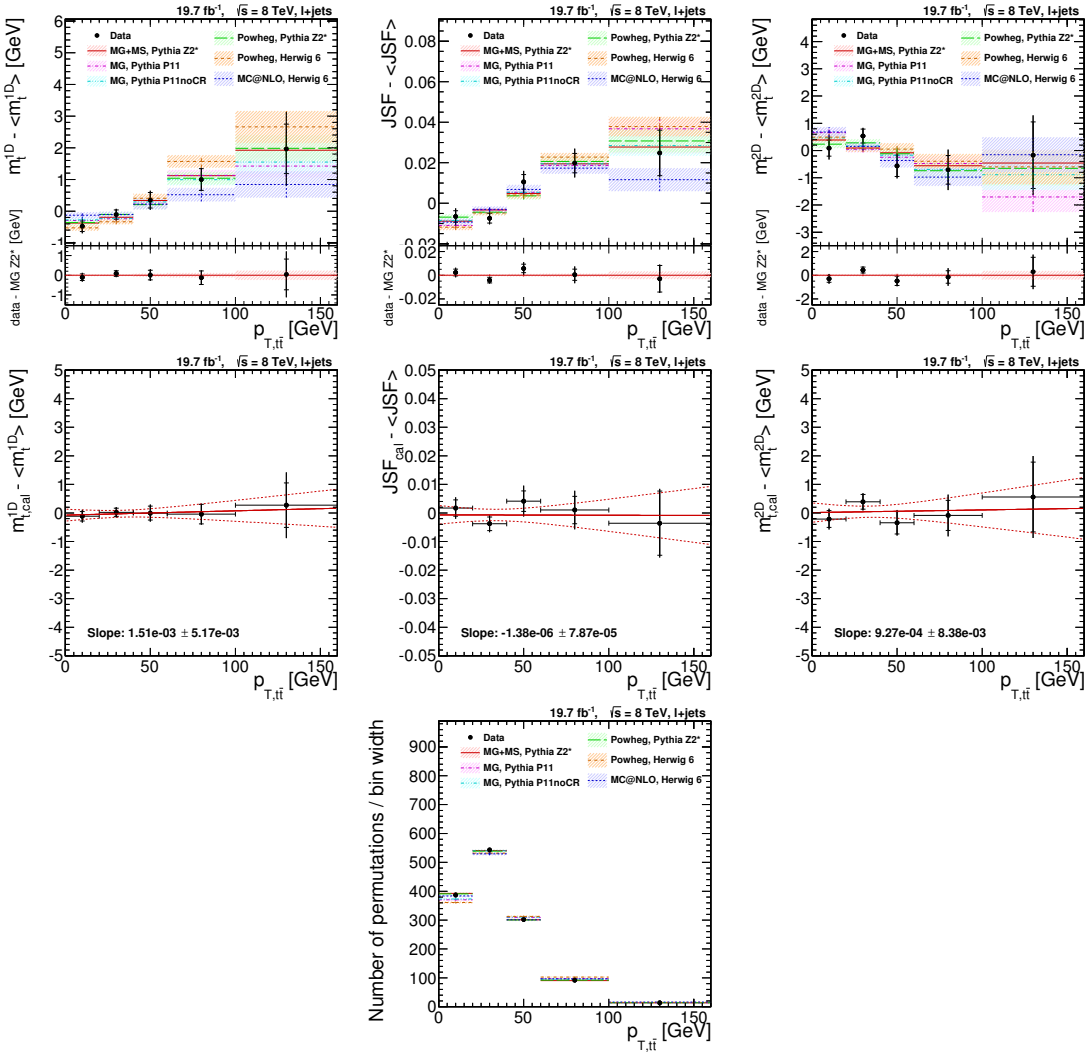

Figure A.6.: Measurement as a function of the transverse momentum of the $t\bar{t}$ system: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $p_{T,t\bar{t}}$ bin, divided by the bin width.

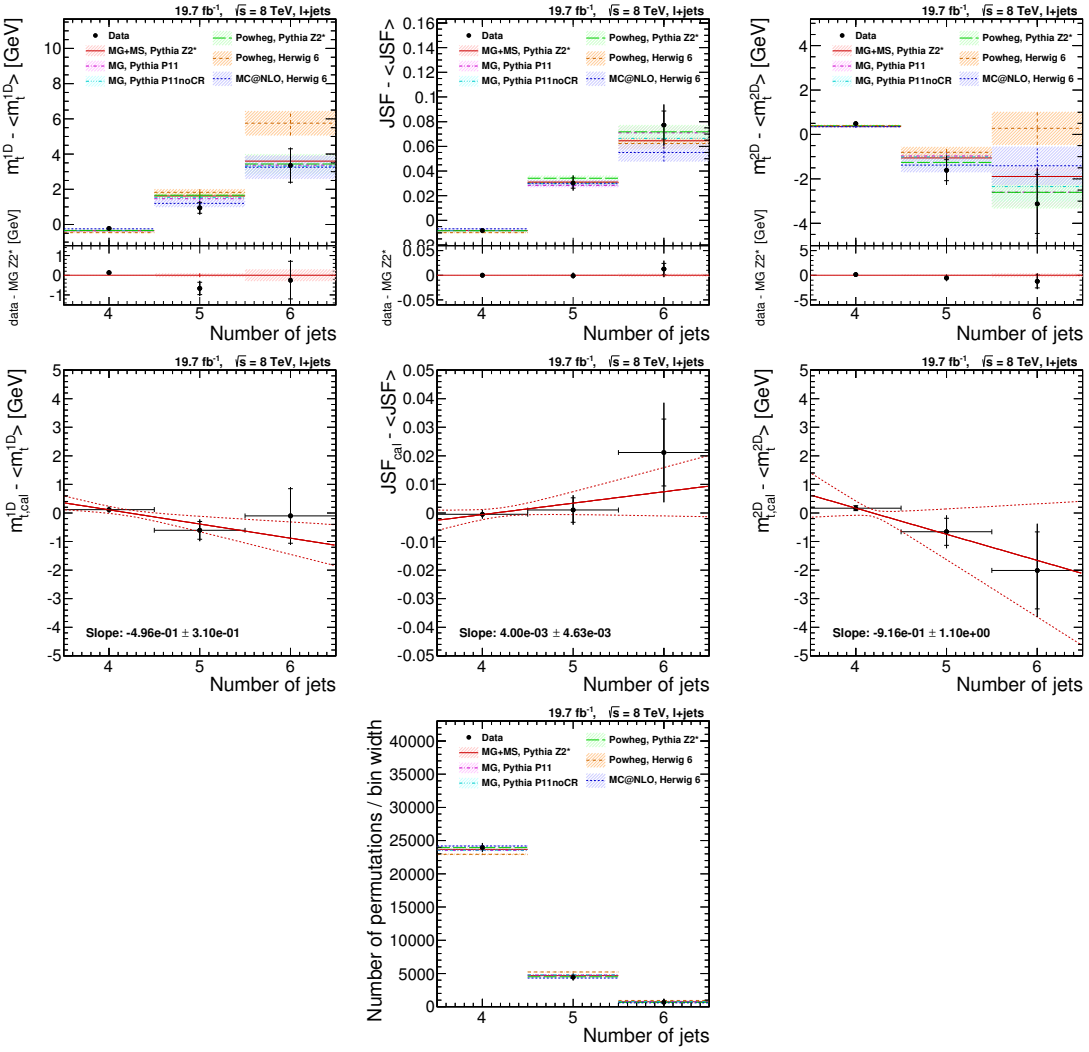

Figure A.7.: Measurement as a function of number of jets with $p_{\mathrm{T}} > 30$ GeV: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per jet multiplicity bin, divided by the bin width.

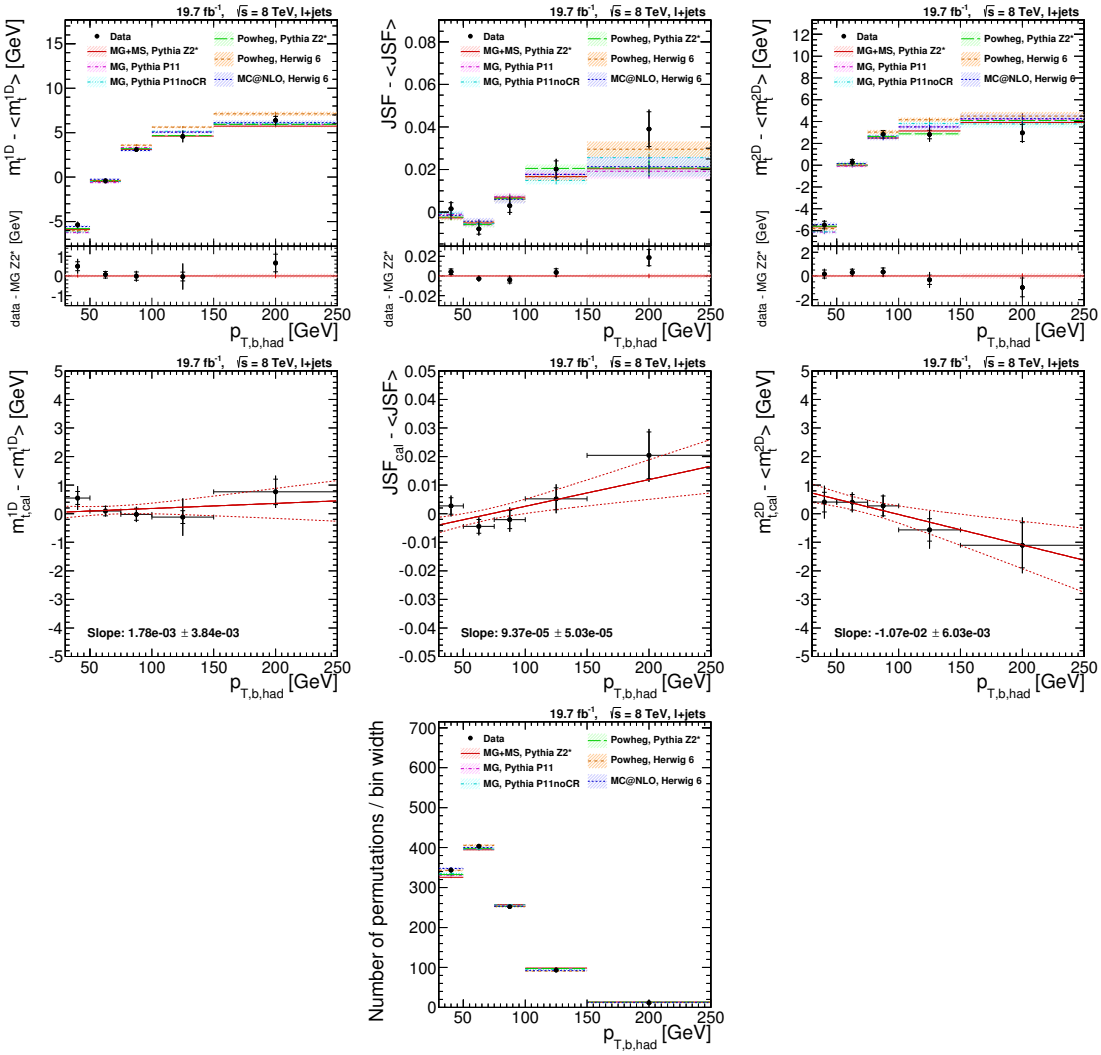

Figure A.8.: Measurement as a function of the p_{T} of the b-jet assigned to the hadronic decay branch: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $p_{\mathrm{T,b,had}}$ bin, divided by the bin width.

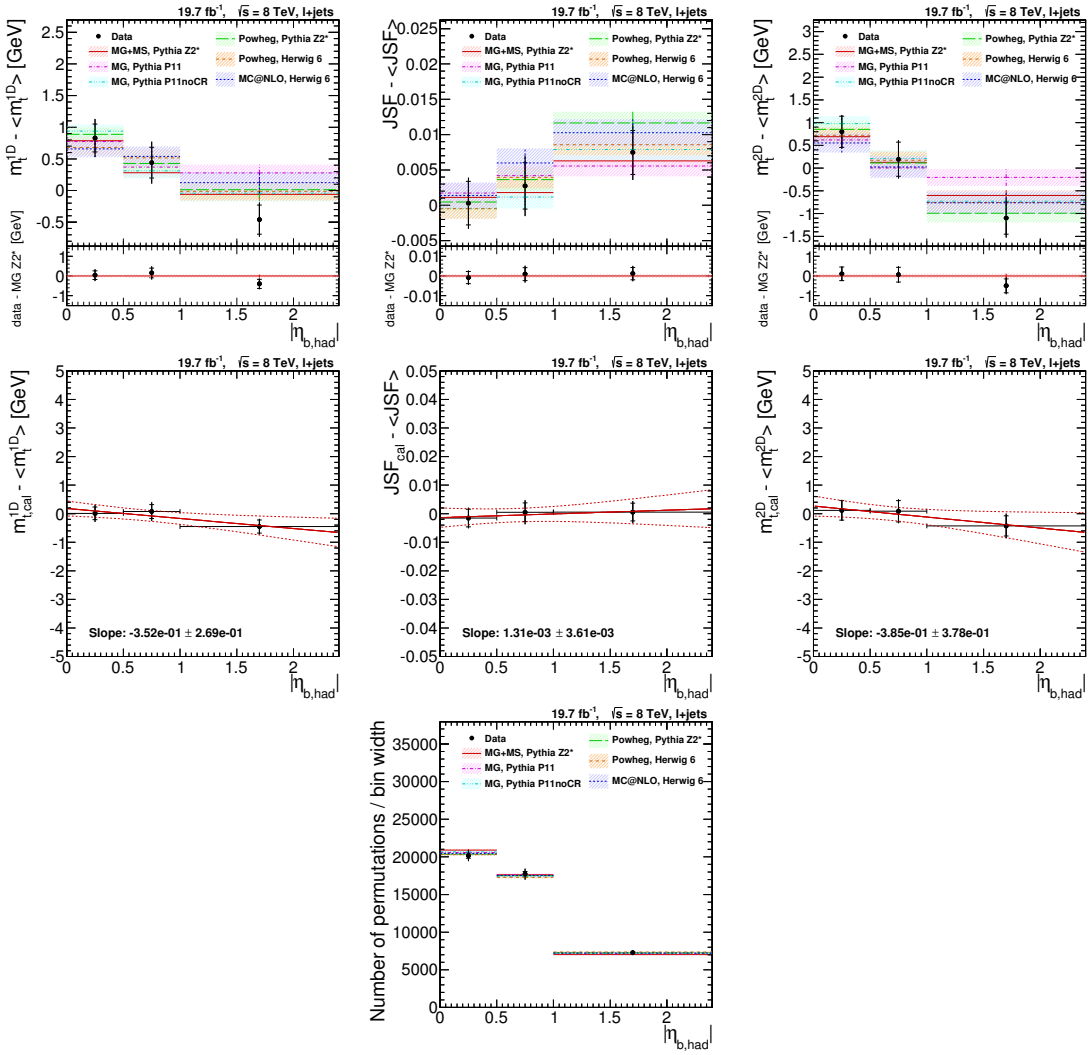

Figure A.9.: Measurement as a function of the pseudo-rapidity of the b-jet assigned to the hadronic decay branch: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $|\eta_{\mathrm{b,had}}|$ bin, divided by the bin width.

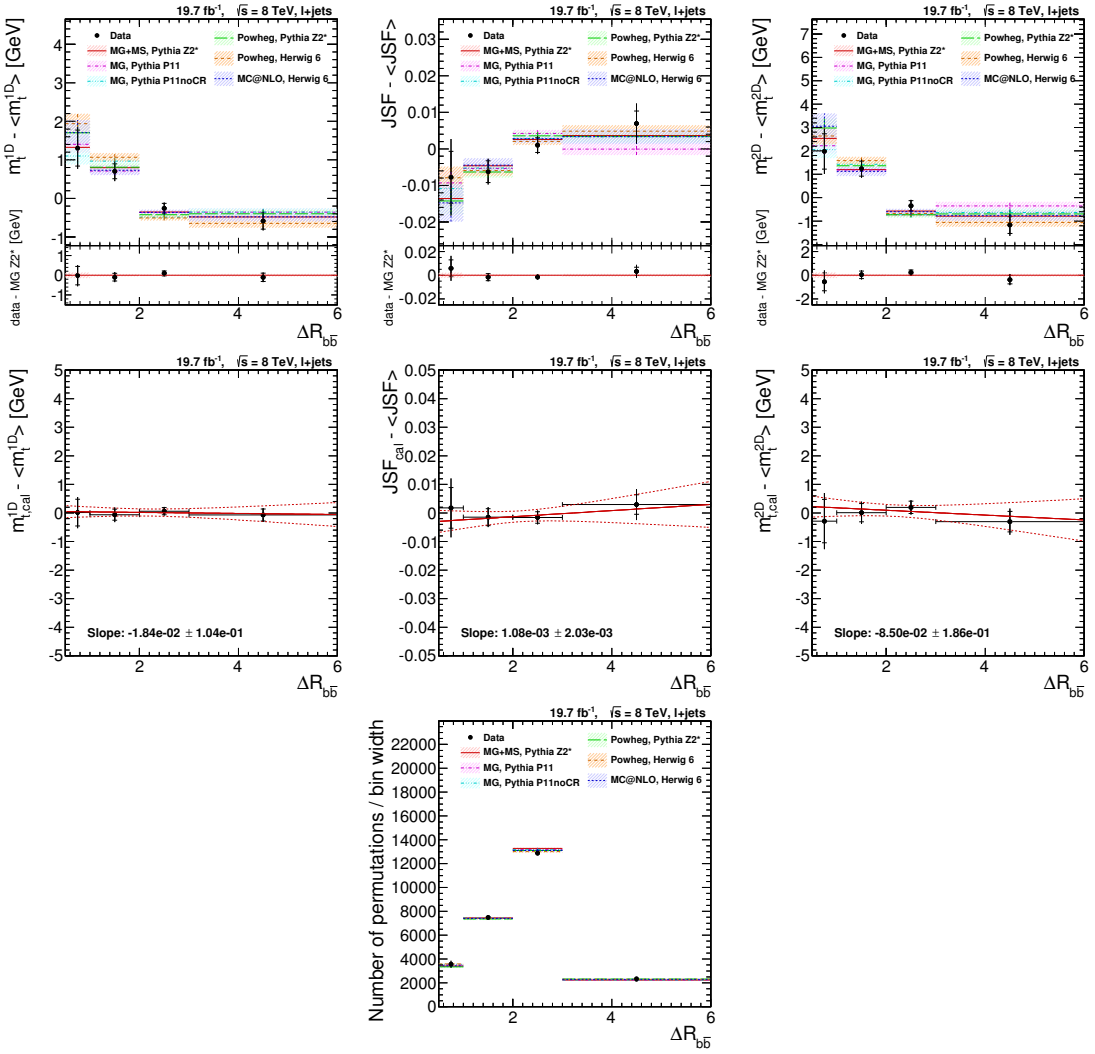

Figure A.10.: Measurement as a function of the ΔR of the b-jets: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $\Delta R_{b\bar{b}}$ bin, divided by the bin width.

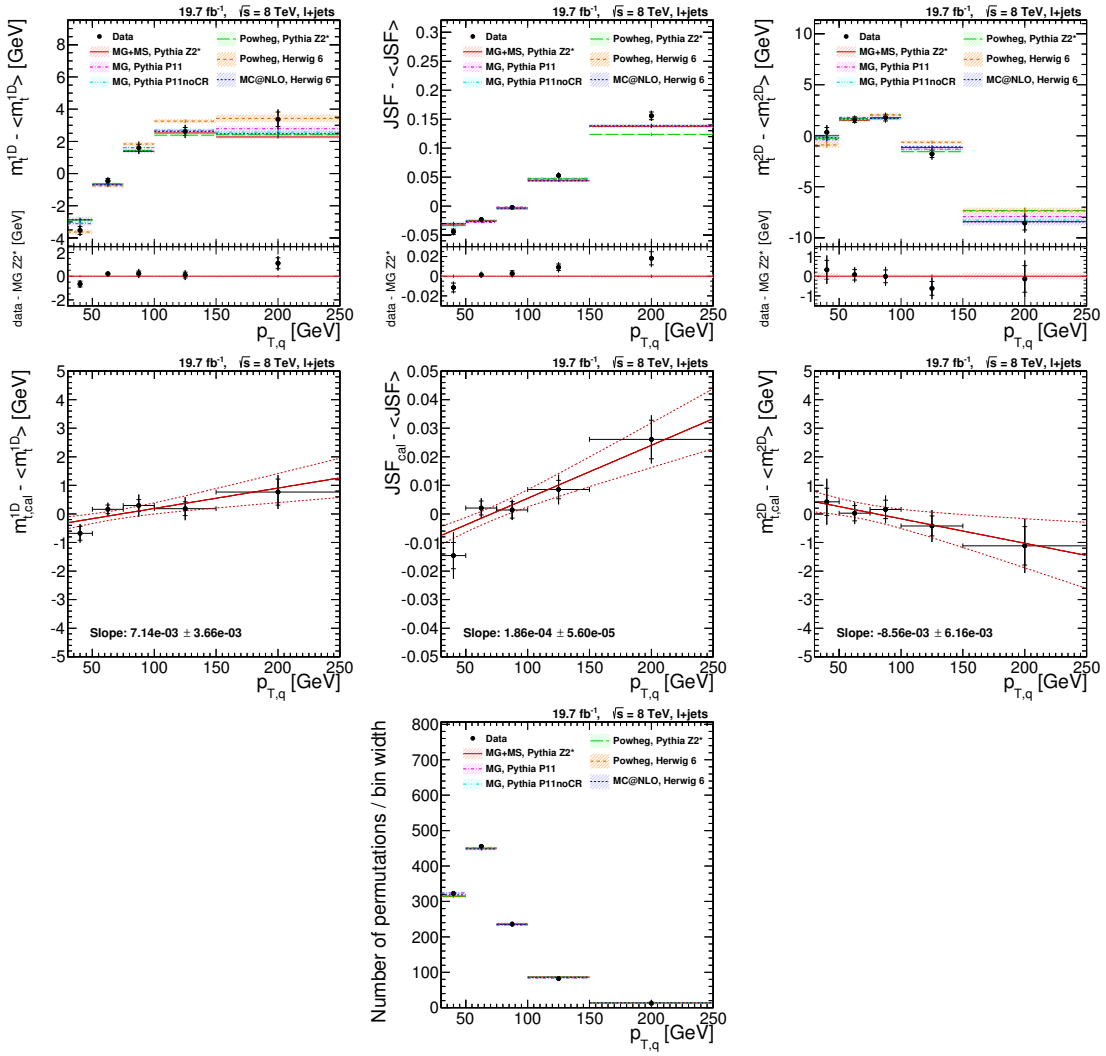

Figure A.11.: Measurement as a function of the p_{T} of the leading light jet assigned to the hadronic decay of a W boson: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $p_{\mathrm{T,q,had}}^{1}$ bin, divided by the bin width.

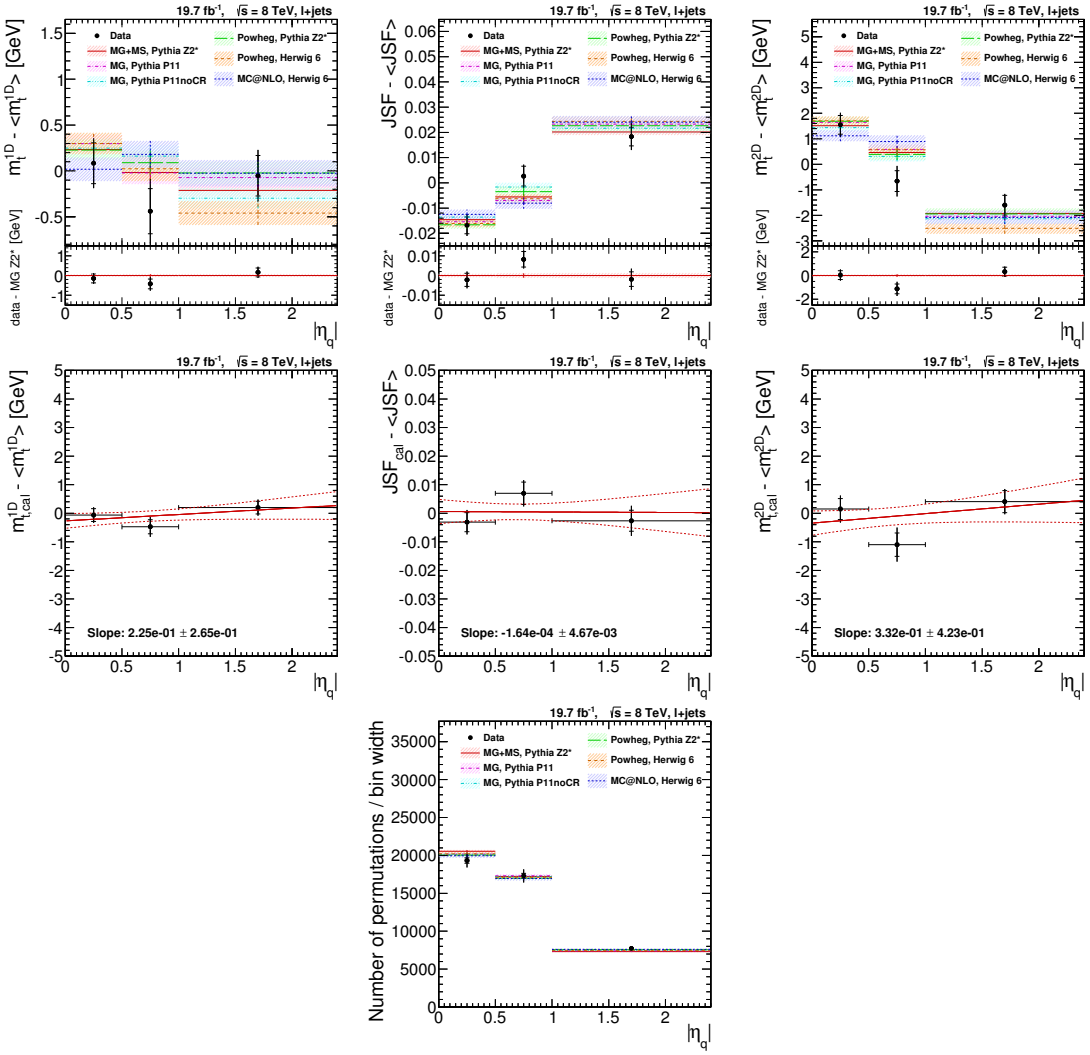

Figure A.12.: Measurement as a function of the pseudo-rapidity of the leading light jet assigned to the hadronic decay of a W boson: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $\left|\eta^1_{\mathrm{q,had}}\right|$ bin, divided by the bin width.

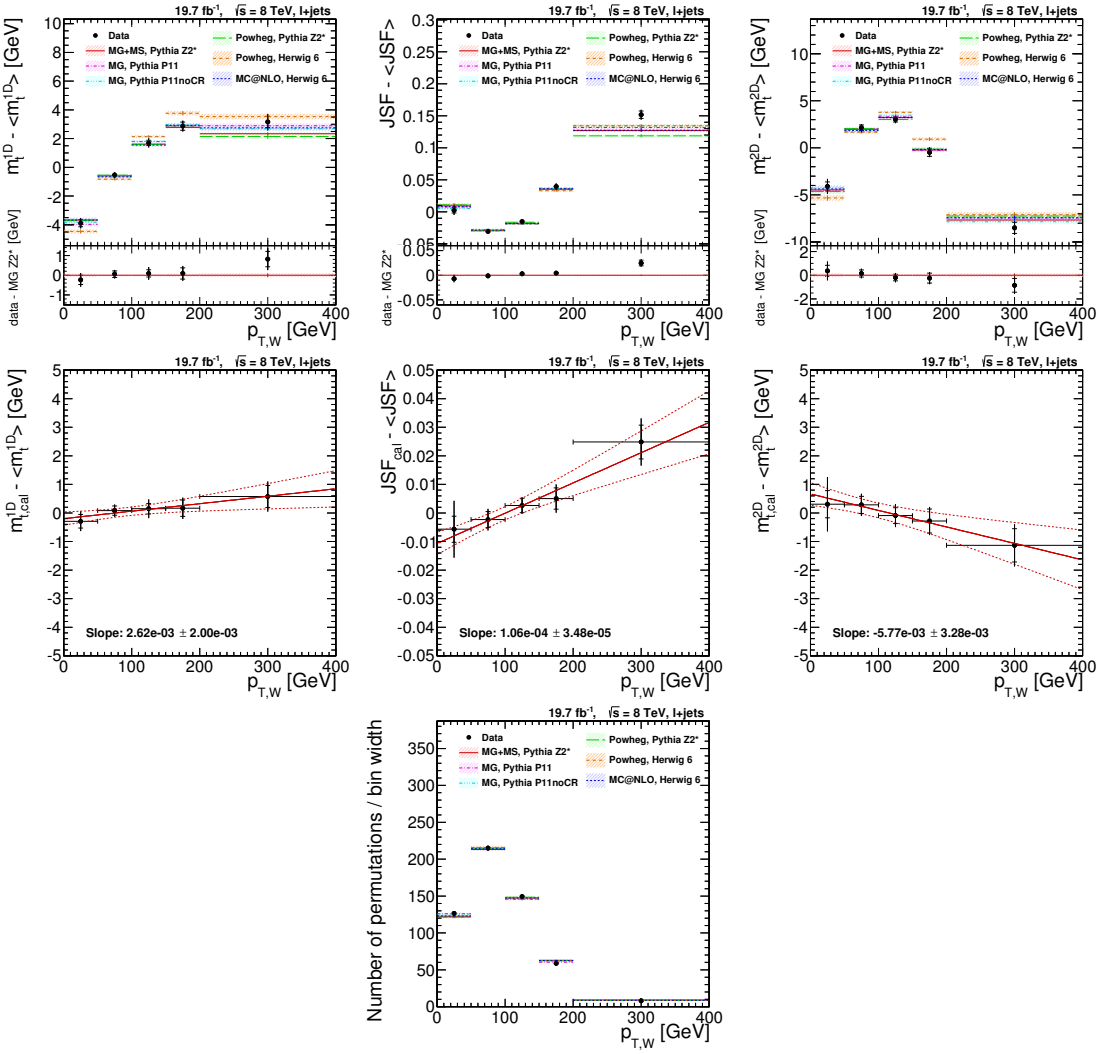

Figure A.13.: Measurement as a function of the p_{T} of the hadronically decaying W boson: $m_t^{1\mathrm{D}}$ (left), JSF (center) and $m_t^{2\mathrm{D}}$ (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $p_{\mathrm{T,W,had}}$ bin, divided by the bin width.

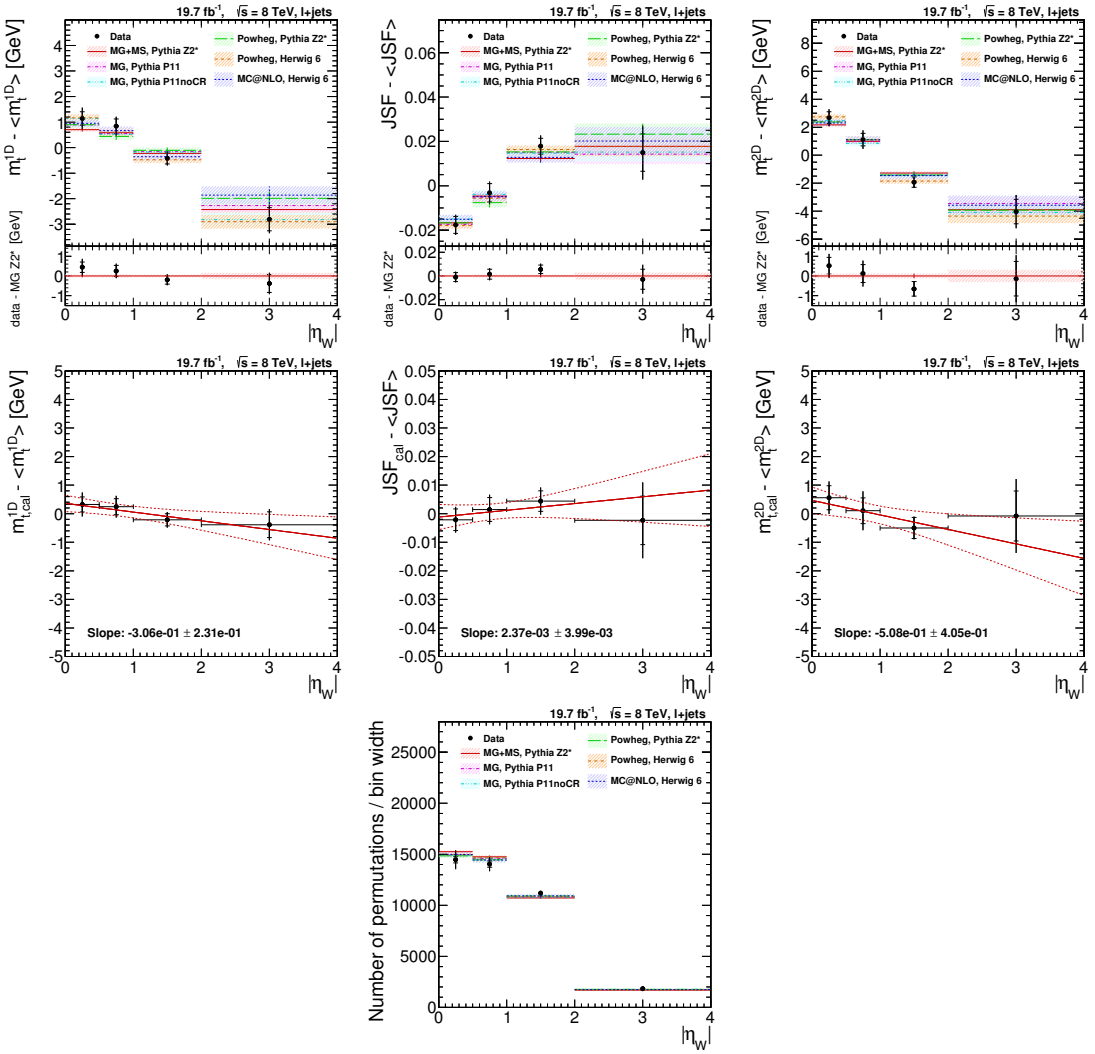

Figure A.14.: Measurement as a function of the pseudo-rapidity of the hadronically decaying W boson: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per $|\eta_{\mathrm{W,had}}|$ bin, divided by the bin width.

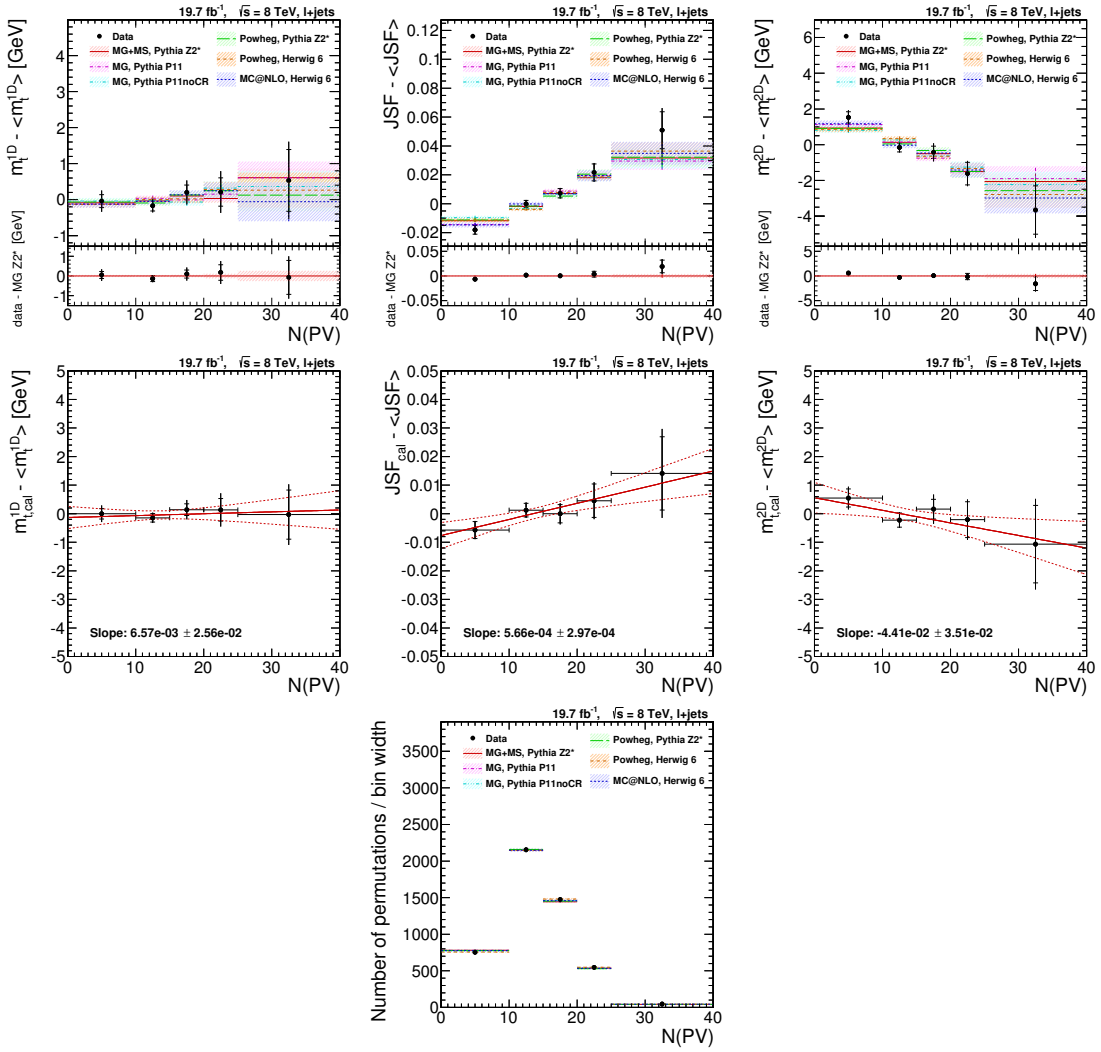

Figure A.15.: Measurement as a function of the number of reconstructed primary vertices: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per N(PV) bin, divided by the bin width.

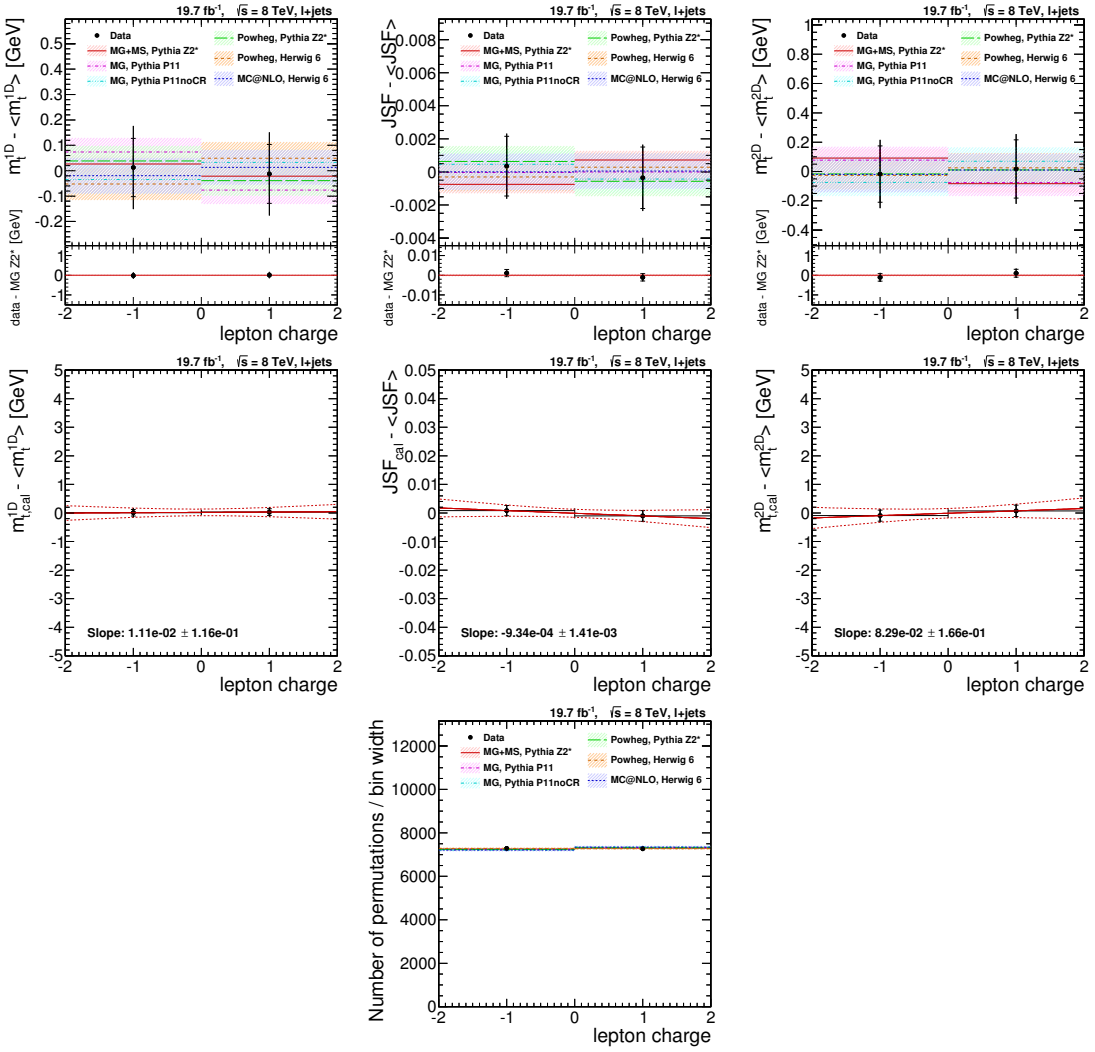

Figure A.16.: Measurement as a function of the charge of the hard lepton: m_t^{1D} (left), JSF (center) and m_t^{2D} (right), before (first row) and after (second row) bin-by-bin calibration. The lower plot shows the number of permutations per lepton charge bin, divided by the bin width.

Bibliography

[1] Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.*, 74:2626–2631, Apr 1995.

[2] Search for High Mass Top Quark Production in $p\overline{p}$ Collisions at $\sqrt{s} =$ 1.8 TeV. *Phys. Rev. Lett.*, 74:2422–2426, Mar 1995.

[3] ATLAS, CDF, CMS, and D0 Collaborations. First combination of Tevatron and LHC measurements of the top-quark mass. 2014. arXiv:1403.4427.

[4] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

[5] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.

[6] Bernd A. Kniehl. Dependence of electroweak parameters on the definition of the top quark mass. *Z. Phys.*, C72:437–447, 1996.

[7] The ALEPH, DELPHI, L3, OPAL Collaborations, the LEP Electroweak Working Group. Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP. *Phys. Rept.*, 532:119, 2013.

[8] M. Baak et al. The global electroweak fit at NNLO and prospects for the LHC and ILC. *Eur. Phys. J.*, C74(9):3046, 2014.

[9] Giuseppe Degrassi, Stefano Di Vita, Joan Elias-Miro, Jose R. Espinosa, Gian F. Giudice, et al. Higgs mass and vacuum stability in the Standard Model at NNLO. *JHEP*, 1208:098, 2012.

[10] Giuseppe Degrassi. The role of the top quark in the stability of the SM Higgs potential. *Nuovo Cim.*, C037(02):47–53, 2014.

[11] Measurement of the top-quark mass in t t-bar events with lepton+jets final states in pp collisions at $\sqrt{s}$=8 TeV. Technical Report CMS-PAS-TOP-14-001, CERN, Geneva, 2014.

[12] Markus Seidel, Peter Schleper, and Hartmut Stadie. Measurement of Top Quark Mass and Jet Energy Scale at the CMS Experiment using the Ideogram Method, 2011. CERN-THESIS-2011-096.

[13] Serguei Chatrchyan et al. Measurement of the top-quark mass in $t\bar{t}$ events with lepton+jets final states in *pp* collisions at $\sqrt{s}$ = 7 TeV. *JHEP*, 12:105, 2012.

[14] Study of the dependence of the top-quark mass measurement on event kinematics. CMS Physics Analysis Summary CMS-PAS-TOP-12-029, CERN, 2013.

[15] Serguei Chatrchyan et al. Measurement of the top-quark mass in all-jets $t\bar{t}$ events in pp collisions at $\sqrt{s}$=7 TeV. *Eur. Phys. J.*, C74(4):2758, 2014.

[16] Measurement of the top-quark mass in t t-bar events with all-jets final states in pp collisions at $\sqrt{s}$=8 TeV. Technical Report CMS-PAS-TOP-14-002, CERN, Geneva, 2014.

[17] Eike Steffen Schlieckau. *Measurement of Top-Quark Mass and Inclusive Top-Quark-Pair Production Cross Section in pp Collisions at $\sqrt{s}$ = 7/8 TeV with CMS*. PhD thesis, 2014. DESY-THESIS-2014-024, CERN-THESIS-2014-109.

[18] Henning Kirschenmann. *Jet Energy Scale Corrections and their Impact on Measurements of the Top-Quark Mass at CMS*. PhD thesis, 2014. DESY-THESIS-2014-031, CERN-THESIS-2014-170.

[19] Matthew Robinson. *Symmetry and the standard model: Mathematics and particle physics*. Springer, New York, USA, 2011. ISBN 9781441982667.

[20] Michele Maggiore. *A Modern introduction to quantum field theory*. Oxford University Press, 2005. ISBN 0198520735.

[21] Andrew Purcell. Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules. (BUL-NA-2012-269. 35/2012):10, Aug 2012.

[22] Vardan Khachatryan et al. Measurement of the inclusive 3-jet production differential cross section in proton-proton collisions at 7 TeV and determination of the strong coupling constant in the TeV range. *Eur. Phys. J.*, C75(5):186, 2015.

[23] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, 1963.

[24] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[25] M. Beneke and Vladimir M. Braun. Heavy quark effective theory beyond perturbation theory: Renormalons, the pole mass and the residual mass term. *Nucl. Phys. B*, 426:301, 1994.

[26] Fred Jegerlehner, Mikhail Yu. Kalmykov, and Bernd A. Kniehl. On the difference between the pole and the $\overline{\rm MS}$ masses of the top quark at the electroweak scale. *Phys. Lett.*, B722:123–129, 2013.

[27] J. Fleischer and F. Jegerlehner. Radiative corrections to Higgs-boson decays in the Weinberg-Salam model. *Phys. Rev. D*, 23:2001–2026, May 1981.

[28] Fred Jegerlehner, Mikhail Yu. Kalmykov, and Bernd A. Kniehl. Self-consistence of the Standard Model via the renormalization group analysis. *J. Phys. Conf. Ser.*, 608(1):012074, 2015.

[29] Andre H. Hoang and Iain W. Stewart. Top Mass Measurements from Jets and the Tevatron Top-Quark Mass. *Nucl. Phys. Proc. Suppl.*, 185:220–226, 2008.

[30] Andre H. Hoang. The Top Mass: Interpretation and Theoretical Uncertainties. In *7th International Workshop on Top Quark Physics (TOP2014) Cannes, France, September 28-October 3, 2014*, 2014.

[31] S. Moch. Precision determination of the top-quark mass. *PoS*, LL2014:054, 2014.

[32] H. Albrecht et al. Observation of $\mathrm{B}^0 - \bar{\mathrm{B}}^0$ Mixing. *Phys. Lett.*, B192:245–252, 1987.

[33] S. Schael et al. Precision electroweak measurements on the Z resonance. *Phys. Rept.*, 427:257–454, 2006.

[34] Chris Quigg. Top-ology. *Phys. Today*, 50N5:20–26, 1997.

[35] S. Alekhin, A. Djouadi, and S. Moch. The top quark and Higgs boson masses and the stability of the electroweak vacuum. *Phys. Lett.*, B716:214–219, 2012.

[36] Isabella Masina. Higgs boson and top quark masses as tests of electroweak vacuum stability. *Phys. Rev.*, D87(5):053001, 2013.

[37] Michal Czakon, Paul Fiedler, and Alexander Mitov. Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$. *Phys. Rev. Lett.*, 110:252004, 2013.

[38] Serguei Chatrchyan et al. Determination of the top-quark pole mass and strong coupling constant from the t t-bar production cross section in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett.*, B728:496–517, 2014.

[39] Top LHC working group. ATLAS+CMS Top Quark Physics Summary and Combination Figures. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOPSummaryFigures` (2015-2-20).

[40] Werner Bernreuther. Top quark physics at the LHC. *J. Phys. G*, G35:083001, 2008.

[41] Vardan Khachatryan et al. Measurement of the ratio $B(t \to Wb)/B(t \to Wq)$ in pp collisions at $\sqrt{s} = 8$ TeV. *Phys. Lett.*, B736:33, 2014.

[42] M. Jezabek and Johann H. Kuhn. QCD Corrections to Semileptonic Decays of Heavy Quarks. *Nucl. Phys.*, B314:1, 1989.

[43] Serguei Chatrchyan et al. Measurements of $t\bar{t}$ spin correlations and top-quark polarization using dilepton final states in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. Lett.*, 112:182001, 2014.

[44] Georges Aad et al. Measurement of Spin Correlation in Top-Antitop Quark Events and Search for Top Squark Pair Production in pp Collisions at $\sqrt{s} = 8$ TeV Using the ATLAS Detector. *Phys. Rev. Lett.*, 114(14):142001, 2015.

[45] Zhenyu Han, Andrey Katz, David Krohn, and Matthew Reece. (Light) Stop Signs. *JHEP*, 1208:083, 2012.

[46] Andy Buckley, Jonathan Butterworth, Stefan Gieseke, David Grellscheid, Stefan Hoche, et al. General-purpose event generators for LHC physics. *Phys. Rept.*, 504:145–233, 2011.

[47] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 1407:079, 2014.

[48] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[49] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 0206:029, 2002.

[50] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 0307:001, 2003.

[51] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al. Event generation with SHERPA 1.1. *JHEP*, 0902:007, 2009.

[52] E. Boos et al. CompHEP 4.4: Automatic computations from Lagrangians to events. *Nucl. Instrum. Meth.*, A534:250–259, 2004.

[53] Borut Paul Kersevan and Elzbieta Richter-Was. The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1. *Comput. Phys. Commun.*, 184:919–985, 2013.

[54] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, 0101:010, 2001.

[55] M. Bahr, S. Gieseke, M.A. Gigg, D. Grellscheid, K. Hamilton, et al. Herwig++ Physics and Manual. *Eur. Phys. J.*, C58:639–707, 2008.

[56] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.

[57] Torbjörn Sjöstrand et al. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[58] D.J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.

[59] Peter Skands. Introduction to QCD. In *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales (TASI 2012)*, pages 341–420, 2013.

[60] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys.*, B126:298, 1977.

[61] V.N. Gribov and L.N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.

[62] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977.

[63] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J. C*, 63:189–285, 2009.

[64] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : going beyond. *JHEP*, 06:128, 2011.

[65] T. Kinoshita. Mass singularities of Feynman amplitudes. *J. Math. Phys.*, 3:650–677, 1962.

[66] T.D. Lee and M. Nauenberg. Degenerate Systems and Mass Singularities. *Phys. Rev.*, 133:B1549–B1562, 1964.

[67] S. Catani and M.H. Seymour. A General algorithm for calculating jet cross-sections in NLO QCD. *Nucl. Phys.*, B485:291–419, 1997.

[68] David A. Kosower. Antenna factorization of gauge theory amplitudes. *Phys. Rev.*, D57:5410–5416, 1998.

[69] Stefan Hoeche, Frank Krauss, Marek Schonherr, and Frank Siegert. QCD matrix elements + parton showers: The NLO case. *JHEP*, 1304:027, 2013.

[70] Fabio Cascioli, Philipp Maierhofer, and Stefano Pozzorini. Scattering Amplitudes with Open Loops. *Phys. Rev. Lett.*, 108:111601, 2012.

[71] C.F. Berger, Z. Bern, L.J. Dixon, F. Febres Cordero, D. Forde, et al. An Automated Implementation of On-Shell Methods for One-Loop Amplitudes. *Phys. Rev.*, D78:036003, 2008.

[72] Valentin Hirschi, Rikkert Frederix, Stefano Frixione, Maria Vittoria Garzelli, Fabio Maltoni, et al. Automation of one-loop QCD corrections. *JHEP*, 1105:044, 2011.

[73] Leif Lonnblad. ARIADNE version 4: A Program for simulation of QCD cascades implementing the color dipole model. *Comput. Phys. Commun.*, 71:15–31, 1992.

[74] Steffen Schumann and Frank Krauss. A Parton shower algorithm based on Catani-Seymour dipole factorisation. *JHEP*, 0803:038, 2008.

[75] Walter T. Giele, David A. Kosower, and Peter Z. Skands. A simple shower and matching algorithm. *Phys. Rev.*, D78:014026, 2008.

[76] S. Catani, F. Krauss, R. Kuhn, and B.R. Webber. QCD matrix elements + parton showers. *JHEP*, 0111:063, 2001.

[77] Leif Lonnblad. Correcting the color dipole cascade model with fixed order matrix elements. *JHEP*, 0205:046, 2002.

[78] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, and Michele Treccani. Matching matrix elements and shower evolution for top-quark production in hadronic collisions. *JHEP*, 0701:013, 2007.

[79] Stefano Frixione, Fabian Stoeckli, Paolo Torrielli, Bryan R. Webber, and Chris D. White. The MCatNLO 4.0 Event Generator. 2010. arXiv:1010.0819.

[80] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[81] Keith Hamilton and Paolo Nason. Improving NLO-parton shower matched simulations with higher order matrix elements. *JHEP*, 1006:039, 2010.

[82] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 1212:061, 2012.

[83] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983.

[84] B.R. Webber. A QCD Model for Jet Fragmentation Including Soft Gluon Interference. *Nucl. Phys.*, B238:492, 1984.

[85] Jan-Christopher Winter, Frank Krauss, and Gerhard Soff. A Modified cluster hadronization model. *Eur. Phys. J.*, C36:381–395, 2004.

[86] M.G. Bowler. e+ e- Production of Heavy Quarks in the String Model. *Z. Phys.*, C11:169, 1981.

[87] J. Beringer et al. (Particle Data Group). Review of Particle Physics. *Phys. Rev. D*, 86:010001, 2012.

[88] ATLAS collaboration. Comparison of Monte Carlo generator predictions for bottom and charm hadrons in the decays of top quarks and the fragmentation of high p_{T} jets. ATL-PHYS-PUB-2014-008, 2014.

[89] Andy Buckley, Hendrik Hoeth, Heiko Lacker, Holger Schulz, and Jan Eike von Seggern. Systematic event generator tuning for the LHC. *Eur. Phys. J.*, C65:331–357, 2010.

[90] Rick Field. Early LHC Underlying Event Data - Findings and Surprises. 2010.

[91] Marisa Sandhoff and Peter Z. Skands. Colour annealing - a toy model of colour reconnections. 2005. FERMILAB-CONF-05-518-T.

[92] Spyros Argyropoulos and Torbjörn Sjöstrand. Effects of color reconnection on $t\bar{t}$ final states at the LHC. *JHEP*, 11:043, 2014.

[93] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[94] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[95] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[96] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[97] Jr. Alves, A. Augusto et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[98] Fabienne Marcastel. CERN's Accelerator Complex. La chaine des accelerateurs du CERN. Oct 2013. General Photo.

[99] CMS Collaboration. Public CMS Luminosity Information. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults` (2015-2-21).

[100] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. CERN, Geneva, 2006.

[101] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[102] Serguei Chatrchyan et al. Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV. *JINST*, 8:P09009, 2013.

[103] Serguei Chatrchyan et al. The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC. *JINST*, 8:P11002, 2013.

[104] Data Parking and Data Scouting at the CMS Experiment. (CMS-DP-2012-022), Sep 2012.

[105] CMS Collaboration. Measurement of the b-tagging efficiency using $t\bar{t}$ events. *CMS Physics Analysis Summary*, CMS-PAS-BTV-11-003, 2012.

[106] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. Adaptive vertex fitting. *J. Phys.*, G34:N343, 2007.

[107] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009.

[108] CMS Collaboration. Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 TeV. CMS Physics Analysis Summary CMS-PAS-PFT-10-002, CERN, 2010.

[109] David Barney. `https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=5697` (2015-2-21).

[110] Serguei Chatrchyan et al. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.

[111] Vardan Khachatryan et al. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10(06):P06005, 2015.

[112] Andreas Hocker, J. Stelzer, F. Tegenfeldt, H. Voss, K. Voss, et al. TMVA - Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[113] Matteo Cacciari and Gavin P. Salam. Dispelling the N^3 myth for the k_{T} jet-finder. *Phys. Lett. B*, 641:57, 2006.

[114] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_{T} jet clustering algorithm. *JHEP*, 04:063, 2008.

[115] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993.

[116] Yuri L. Dokshitzer, G.D. Leder, S. Moretti, and B.R. Webber. Better jet clustering algorithms. *JHEP*, 9708:001, 1997.

[117] M. Wobisch and T. Wengler. Hadronization corrections to jet cross-sections in deep inelastic scattering. 1998. arXiv:hep-ph/9907280.

[118] Serguei Chatrchyan et al. Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST*, 6:P11002, 2011.

[119] Denis Rathjens. *Jet energy calibration and a search for supersymmetry with vector boson fusion channel like sign di-τ_h final states.* PhD thesis.

[120] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J.*, C72:1896, 2012.

[121] CMS Collaboration. 8 TeV Jet Energy Corrections and Uncertainties based on 19.8 fb^{-1} of data in CMS. CMS-DP-2013-033, 2013.

[122] CMS Collaboration. Calculation of Residual Energy Correction for b Jets Using Z+b Events in 8 TeV pp Collisions. 2014. CMS-PAS-JME-13-001.

[123] Kristin Goebel. *Probing supersymmetry based on precise jet measurements at the CMS experiment*. PhD thesis, U. Hamburg, Dept. Phys., 2015. DESY-THESIS-2015-003.

[124] CMS Collaboration. Jet Energy Corrections for Multiple Cone Sizes. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/MultipleConeSizes14` (2015-2-21).

[125] Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, $t\bar{t}$ and boosted topology events. CMS Physics Analysis Summary CMS-PAS-BTV-13-001, CERN, 2013.

[126] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.

[127] Vardan Khachatryan et al. Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV. *JINST*, 10(02):P02006, 2015.

[128] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP*, 03:015, 2013.

[129] Serguei Chatrchyan et al. Measurement of the underlying event activity at the LHC with $\sqrt{s} = 7$ TeV and comparison with $\sqrt{s} = 0.9$ TeV. *JHEP*, 09:109, 2011.

[130] S. Jadach, Z. Was, R. Decker, and Johann H. Kuhn. The tau decay library TAUOLA: Version 2.4. *Comput. Phys. Commun.*, 76:361–380, 1993.

[131] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *JHEP*, 09:111, 2009.

[132] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J. C*, 71:1547, 2011.

[133] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[134] S. Agostinelli et al. GEANT 4 – a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.

[135] John M. Campbell and R. K. Ellis. MCFM for the Tevatron and the LHC. *Nucl. Phys. Proc. Suppl.*, 205-206:10, 2010.

[136] Ryan Gavin, Ye Li, Frank Petriello, and Seth Quackenbush. W Physics at the LHC with FEWZ 2.1. *Comput. Phys. Commun.*, 184:208–214, 2013.

[137] Nikolaos Kidonakis. Top Quark Production. In *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013)*, pages 139–168, 2014.

[138] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, et al. HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR. *Comput. Phys. Commun.*, 182:1034–1046, 2011.

[139] P. Kant et al. HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions. *Comput. Phys. Commun.*, 191:74–89, 2015.

[140] Jörn Lange. *Differential Top-Quark-Pair Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV with CMS and Charge Multiplication in Highly-Irradiated Silicon Sensors*. PhD thesis. DESY-THESIS-2013-042.

[141] Martin Görner. *Differential Cross Sections for Top-Quark-Pair Production in the e/μ+Jets Final State at $\sqrt{s} = 8$ TeV in CMS*. PhD thesis. DESY-THESIS-2014-023.

[142] Standard Model Cross Sections for CMS at 8 TeV. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat8TeV` (2015-2-21).

[143] Serguei Chatrchyan et al. Measurement of the $t\bar{t}$ production cross section in pp Collisions at 7 TeV in lepton + jets events using b-quark jet identification. *Phys. Rev. D*, 84:092004, 2011.

[144] Serguei Chatrchyan et al. Measurement of the mass difference between top and antitop quarks. *JHEP*, 06:109, 2012.

[145] Scott Stuart Snyder. *Measurement of the top quark mass at D0*. PhD thesis, 1995. FERMILAB-THESIS-1995-27.

[146] B. Abbott et al. Direct measurement of the top quark mass at D0. *Phys. Rev. D*, 58:052001, 1998.

[147] J. Abdallah et al. Measurement of the mass and width of the W boson in e^+e^- collisions at $\sqrt{s}$ = 161 – 209 GeV. *Eur. Phys. J. C*, 55:1, 2008.

[148] T. Aaltonen et al. Measurement of the top-quark mass in all-hadronic decays in $p\bar{p}$ collisions at CDF II. *Phys. Rev. Lett.*, 98:142001, 2007.

[149] V. M. Abazov et al. Measurement of the top quark mass in the lepton + jets channel using the ideogram method. *Phys. Rev. D*, 75:092001, 2007.

[150] Jet energy scale uncertainty correlations between ATLAS and CMS. Technical Report CMS-PAS-JME-14-003, CERN, Geneva, 2014.

[151] Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV with the CMS Detector. 2012. CMS-PAS-JME-12-002.

[152] A. Heister et al. Study of the fragmentation of b quarks into B mesons at the Z peak. *Phys. Lett. B*, 512:30–48, 2001.

[153] J. Abdallah et al. A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z Pole. *Eur. Phys. J. C*, 71:1557, 2011.

[154] J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, Pavel M. Nadolsky, et al. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 0207:012, 2002.

[155] Michiel Botje, Jon Butterworth, Amanda Cooper-Sarkar, Albert de Roeck, Joel Feltesse, et al. The PDF4LHC Working Group Interim Recommendations. 2011. arXiv:1101.0538.

[156] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, et al. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.

[157] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, et al. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244–289, 2013.

[158] Georges Aad et al. Measurement of the $t\bar{t}$ production cross-section as a function of jet multiplicity and jet transverse momentum in 7 TeV proton-proton collisions with the ATLAS detector. *JHEP*, 01:020, 2015.

[159] ATLAS collaboration. Comparison of Monte Carlo generator predictions for gap fraction and jet multiplicity observables in top-antitop events. ATL-PHYS-PUB-2014-005, 2014.

[160] Serguei Chatrchyan et al. Measurement of jet multiplicity distributions in t t-bar production in pp collisions at $\sqrt{s}$ = 7 TeV. *Eur. Phys. J.*, C74:3014, 2014.

[161] Measurement of the Jet Multiplicity in dileptonic Top Quark Pair Events at 8 TeV. Technical Report CMS-PAS-TOP-12-041, CERN, Geneva, 2013.

[162] A. Heister et al. Studies of QCD at e+ e- centre-of-mass energies between 91-GeV and 209-GeV. *Eur. Phys. J.*, C35:457–486, 2004.

[163] J. Abdallah et al. A Study of the energy evolution of event shape distributions and their means with the DELPHI detector at LEP. *Eur. Phys. J.*, C29:285–312, 2003.

[164] P. Achard et al. Studies of hadronic event structure in e^+e^- annihilation from 30-GeV to 209-GeV with the L3 detector. *Phys. Rept.*, 399:71–174, 2004.

[165] G. Abbiendi et al. Measurement of event shape distributions and moments in e+ e- $\rightarrow$ hadrons at 91-GeV - 209-GeV and a determination of alpha(s). *Eur. Phys. J.*, C40:287–316, 2005.

[166] Georges Aad et al. Measurement of jet shapes in top-quark pair events at $\sqrt{s}$ = 7 TeV using the ATLAS detector. *Eur. Phys. J.*, C73:2676, 2013.

[167] Top LHC working group. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TheorySystematics` (2014-10-14).

[168] Nikolaos Kidonakis. Differential and total cross sections for top pair and single top production. In *Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012)*, pages 831–834, 2012.

[169] Georges Aad et al. Measurements of normalized differential cross sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *Phys. Rev.*, D90(7):072004, 2014.

[170] Serguei Chatrchyan et al. Measurement of differential top-quark pair production cross sections in *pp* colisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 73:2339, 2013.

[171] Measurement of differential top-quark pair production cross sections in the lepton+jets channel in pp collisions at 8 TeV. Technical Report CMS-PAS-TOP-12-027, CERN, Geneva, 2013.

[172] Vardan Khachatryan et al. Measurement of the Differential Cross Section for Top Quark Pair Production in pp Collisions at $\sqrt{s} = 8$ TeV. 2015. arXiv:1505.04480.

[173] Paolo Nason. Presentation at Top LHC WG meeting, `https://indico.cern.ch/event/301787/session/10/contribution/23/material/slides/0.pdf`.

[174] Peter Zeiler Skands. Tuning Monte Carlo generators: The Perugia tunes. *Phys. Rev. D*, 82:074018, 2010.

[175] Peter Z. Skands and Daniel Wicke. Non-perturbative QCD effects and the top mass at the Tevatron. *Eur. Phys. J. C*, 52:133, 2007.

[176] G. Aad et al. Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC. *New J. Phys.*, 13:053033, 2011.

[177] Vardan Khachatryan et al. Charged particle multiplicities in *pp* interactions at $\sqrt{s} = 0.9$, 2.36, and 7 TeV. *JHEP*, 1101:079, 2011.

[178] A. Karneyeu, L. Mijovic, S. Prestel, and P.Z. Skands. MCPLOTS: a particle physics resource based on volunteer computing. *Eur. Phys. J.*, C74:2714, 2014.

[179] Combination of the CMS top-quark mass measurements from Run 1 of the LHC. Technical Report CMS-PAS-TOP-14-015, CERN, Geneva, 2014.

[180] Measurement of the Top Quark Mass from $\sqrt{s} = 7$ TeV ATLAS Data using a 3-dimensional Template Fit. Technical Report ATLAS-CONF-2013-046, CERN, Geneva, May 2013.

[181] T. Aaltonen et al. Precision Top-Quark Mass Measurements at CDF. *Phys. Rev. Lett.*, 109:152003, 2012.

[182] Paul Joseph Lujan. *Precision measurement of the top quark mass in the lepton + jets channel using a matrix element method with Quasi-Monte Carlo integration*. PhD thesis, 2009. FERMILAB-THESIS-2009-26.

[183] Victor Mukhamedovich Abazov et al. Precise measurement of the top-quark mass from lepton+jets events at D0. *Phys. Rev.*, D84:032004, 2011.

[184] L. Lyons, D. Gibaut and P. Clifford. How to combine correlated estimates of a single physical quantity. *Nucl. Instrum. Meth.*, A270:110, 1988.

[185] Victor Mukhamedovich Abazov et al. Precision measurement of the top-quark mass in lepton+jets final states. *Phys. Rev. Lett.*, 113:032002, 2014.

[186] Victor Mukhamedovich Abazov et al. Jet energy scale determination in the D0 experiment. *Nucl. Instrum. Meth.*, A763:442–475, 2014.

[187] Tevatron Electroweak Working Group. Combination of CDF and D0 results on the mass of the top quark using up to 9.7 fb^{-1} at the Tevatron. 2014. FERMILAB-CONF-14-227-E-TD.

[188] ATLAS and CMS Collaborations. Combination of ATLAS and CMS results on the mass of the top quark using up to 4.9 fb^{-1} of data. CMS Physics Analysis Summary CMS-PAS-TOP-13-005, CERN, 2013.

[189] Andy Buckley, Jonathan Butterworth, Leif Lonnblad, David Grellscheid, Hendrik Hoeth, et al. Rivet user manual. *Comput. Phys. Commun.*, 184:2803–2819, 2013.

[190] Matt Dobbs and Jorgen Beck Hansen. The HepMC C++ Monte Carlo event record for High Energy Physics. *Comput. Phys. Commun.*, 134:41–46, 2001.

[191] Georges Aad et al. Measurement of $t\bar{t}$ production with a veto on additional central jet activity in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *Eur. Phys. J.*, C72:2043, 2012.

[192] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J.*, C74(8):3024, 2014.

[193] G Barker, E Ben-Haim, M Feindt, U Kerzel, P Roudeau, L Ramler, and A Savoy-Navarro. DELPHI-2002-069-CONF-603, Jul 2002.

[194] G. Abbiendi et al. Inclusive analysis of the b quark fragmentation function in Z decays at LEP. *Eur. Phys. J.*, C29:463–478, 2003.

[195] Koya Abe et al. Measurement of the b quark fragmentation function in Z0 decays. *Phys. Rev.*, D65:092006, 2002.

[196] Michelangelo Mangano. Top mass and its interpretation: open TH issues – Anatomy of top production and decay, and impact on top-quark mass measurement, 2012. Talk given at TOP LHC WG meeting.

[197] Georges Aad et al. Measurement of the mass difference between top and anti-top quarks in pp collisions at $\sqrt(s) = 7$ TeV using the ATLAS detector. *Phys. Lett.*, B728:363–379, 2014.

[198] Measurement of the top - antitop mass difference in pp collisions at $\sqrt{s} = 8$ TeV. 2012. CMS-PAS-TOP-12-031.

[199] Study of the underlying event, b-quark fragmentation and hadronization properties in $t\bar{t}$ events. Technical Report CMS-PAS-TOP-13-007, CERN, Geneva, 2013.

[200] Simone Alioli, Patricia Fernandez, Juan Fuster, Adrian Irles, Sven-Olaf Moch, et al. A new observable to measure the top-quark mass at hadron colliders. *Eur. Phys. J.*, C73:2438, 2013.

[201] Georges Aad et al. Determination of the top-quark pole mass using $t\bar{t} + 1$-jet events collected with the ATLAS experiment in 7 TeV pp collisions. 2015. arXiv:1507.01769.

Acknowledgments

I am greatly indebted to Prof. Dr. Peter Schleper who gave me the opportunity to work in his research group and to write this thesis. In addition to his helpful advice, he served as first referee of the thesis. I would also like to thank Prof. Dr. Wolfgang Wagner, Prof. Dr. Ian Brock, Prof. Dr. Ludwig Mathey, Prof. Dr. Sven-Olaf Moch, and Dr. Andreas Meyer for reviewing my thesis and thesis defense.

The groundwork necessary for this measurement was laid by numerous colleagues from the CMS collaboration, and especially by the UHH top analysis team, consisting of Dr. Eike Schlieckau, Dr. Martin Görner, Dr. Holger Enderle, and Dr. Jörn Lange. Thank you very much!

I also thank Dr. Henning Kirschenmann and Dr. Denis Rathjens for being great office mates and profound sources of knowledge about jet energy corrections.

Finally, special thanks go to my advisor Dr. Hartmut Stadie for a countless number of useful discussions as well as for important know-how about (surviving) the day-to-day business in particle physics.