Constraints on the top quark's charge with the CMS experiment

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften

vorgelegt von

Diplom-Physikerin Yvonne Küssel

aus Heidelberg

II

Contents

Zusammenfassung

Das Standardmodell der Elementarteilchenphysik wird am LHC Ringbeschleuniger in Genf auf die Probe gestellt. Unter anderem werden Top-Quark-Paare in großer Anzahl erzeugt, womit Präzisionsmessungen von Top-Quark-Eigenschaften möglich werden.

Die erwartete elektrische Ladung des Top-Quarks als elektroschwacher Isospin-Partner des Bottom-Quarks beträgt $+2/3\,$e. Es wird eine Messung präsentiert, die die Standardmodell-Erwartung von einem Szenario mit einer Ladung des Top-Quarks von $-4/3\,$e abgrenzt. Hierfür werden Top-Quark-Paarzerfälle verwendet, in denen das eine Top-Quark in ein Bottom-Quark, ein hochenergetisches Myon und ein Myon-Neutrino zerfällt und das andere in ein Bottom-Quark und zwei weitere leichtere Quarks. Die Messung wird mit Proton-Proton-Kollisionen aus dem Jahr 2011 durchgeführt, die mit dem CMS Experiment aufgezeichnet wurden. Die Datenmenge entspricht $5.0\,\mathrm{fb}^{-1}$.

Die Korrelation zwischen dem hochenergetischen Myon aus dem harten Prozess und einem Myon aus einem der beiden Bottom-Quarks wird genutzt, um die Ereignisse in eine $+2/3\,$e- oder $-4/3\,$e-Kategorie einzuordnen. Die normalisierte Asymmetrie A zwischen diesen beiden Kategorien dient als Test-Statistik, um das exotische Szenario auszuschließen, welches zu einer Asymmetrie von $A = -1$ führen würde. Die gemessene Asymmetrie von $A = 1.10 \pm 0.12(\text{stat.}) \pm 0.34(\text{sys.})$ schließt das $-4/3\,$e-Szenario mit hoher Signifikanz aus und entspricht der Standardmodell-Erwartung von $A = +1$.

Abstract

The standard model of particle physics is being probed at the LHC collider ring at Geneva. Amongst others, top-quark pairs are produced in large quantities. This makes precision measurements of top-quark properties possible.

The top quark as the electroweak-isospin partner of the bottom quark is expected to have an electric charge of $+2/3\,$e. A measurement is presented that discriminates between the top quark charge hypotheses of $+2/3\,$e and $-4/3\,$e using the muon+jets final state of top-pair events $t\bar{t} \rightarrow (b\mu\nu)(bqq')$. The measurement is performed with proton-proton collisions at $\sqrt{s} = 7\,$TeV recorded by the CMS detector in the year 2011, corresponding to an integrated luminosity of $5.0\,\text{fb}^{-1}$.

Charge correlations between the high-p_{T} muons from the hard process and a muon from one of the bottom quarks are exploited to sort the events in a $+2/3\,$e or $-4/3\,$e category. The normalized asymmetry between both categories is calculated. The top quark charge scenario of $-4/3\,$e would correspond to an asymmetry of $A = -1$. This is excluded with high significance while the measured asymmetry of $A = 1.10 \pm 0.12(\text{stat.}) \pm 0.34(\text{sys.})$ confirms the standard-model expectation of $A = +1$.

Chapter 1

Introduction

The standard model of particle physics describes the constituent parts of matter and the interactions between them. Matter and interactions are expressed as quantized fields. The quantum particles of matter carry a spin of $1/2$. The interaction field particles carry a spin 1. The standard model contains two kinds of interactions. Firstly, the strong interaction that is e.g. responsible for the binding forces within nucleons. Its mediator particles are gluons that couple to all particles carrying color charge. Secondly, there is the electro-weak interaction that comprises the electro-magnetic force transmitted via photons and the weak interaction transmitted by the W and Z^0 bosons involved in particle decays. A summary of the gauge bosons is given in Table 1.1.

Table 1.1: The gauge bosons within the standard model. The masses are given in natural units and taken from [1]. Q denotes the electric charge in units of the elementary charge. The column color denotes the charge of the strong interaction that can take eight combinations of color-anticolor pairs of green, red and blue as value.

Boson	Interaction	Mass [GeV]	Q [e]	Color
gluon	strong	0	0	✓
W	weak	80.385 ± 0.015	± 1	–
Z^0	weak	91.1876 ± 0.0021	0	–
γ	electric	0	0	–

The fermions are categorized into the group of strong interacting particles, the quarks, and all others, the leptons. The lightest fermions are stable due to conservation laws. Building the first generation, these are the up and down quarks and in the leptonic sector the electron and its neutrino. There are two more generations with increasing masses. An overview of the three generations of fermions and some of their quantum numbers is given in Table 1.2. All fermions have corresponding anti-particles with the same mass but with some quantum numbers inverted. In particular the electric charge is of opposite sign. In a plain gauge symmetry of $SU(3) \times SU(2) \times U(1)$ all particles are predicted to be massless in contradiction to the experimental measurements represented in Figure 1.1.

Table 1.2: The three generations of fermions within the standard model. Q denotes the electric charge in units of the elementary charge, T_3 the third component of the weak isospin. The column color denotes the charge of the strong interaction that can take the values green, red and blue.

Fermiontype	I.	II.	III.	Q [e]	T_3	Color
Quarks	$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	2/3	1/2	✓
				−1/3	−1/2	✓
	u_R	c_R	t_R	2/3	0	✓
	d_R	s_R	b_R	−1/3	0	✓
Leptons	$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L$	0	1/2	−
				−1	−1/2	−
	e_R^-	μ_R^-	τ_R^-	−1	0	−

The Higgs mechanism solves this deficiency. The Higgs field breaks the $SU(2) \times U(1)$ electro-weak symmetry. A consequence of the Higgs mechanism is the existence of new boson, the so-called Higgs boson. It generates masses to the W and Z^0 bosons and to the charged fermions by interacting with them with gauge and Yukawa couplings respectively. A new particle so far consistent with the Higgs boson has recently been discovered [2]. Further investigations are ongoing to determine its properties and therewith its role within the standard model. Neutrino masses that are experimentally proven by evidences for neutrino oscillations [1] cannot be generated by the Yukawa coupling but imply e.g. the existence of sterile, right-handed neutrinos or that neutrinos are their own anti-particles.
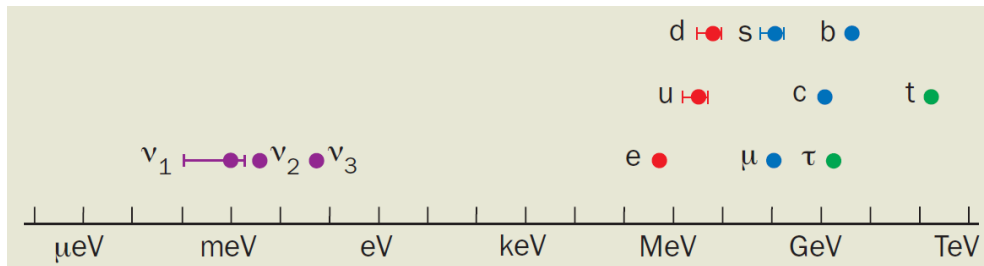
Figure 1.1: Fermion masses. Neutrino-oscillation experiments have not measured the mass of neutrinos absolutely yet but the mass difference between the different neutrino types. The absolute values of the neutrino masses shown here are just a guess. [3]

The top quark, as the last missing quark and the heaviest elementary particle, was discovered in 1995 at the Tevatron accelerator. As the weak-isospin partner of the bottom quark, it was predicted for a long time. So far it seems to conform to all expectations imposed by the standard model. However, the standard model prediction for its electric charge of $+2/3$ e still needs to be confirmed. An electrical charge of $-4/3$ e would be consistent with the charge-conservation law with respect to the decay particles which are the W boson and the b quark. This scenario is already excluded at a 95 % confidence level by the Tevatron [4]. Recent results from the ATLAS experiment at the LHC indicate an

exclusion of the exotic quark scenario at a 5σ confidence level [5]. The result presented here complements these results.

This analysis uses proton-proton collisions data at $\sqrt{s} = 7\,\text{TeV}$ provided by the Large Hadron Collider in the 2011 runtime recorded by the CMS detector. $t\bar{t}$ events corresponding to an integrated luminosity of $5.0\,\text{fb}^{-1}$ are considered in the muon+jets final state $t\bar{t} \to (b\mu\nu)(bqq')$. In addition to a cut based standard top pair selection [6], a b-tagging algorithm [7] is used to identify the two most b-like jets of each top-pair candidate event. The top mass is used to associate an identified b jet to the b quark from the hadronically decaying side. This allows to associate the second identified b jet in the event to the leptonically decaying side. A top quark charge hypothesis is built by combining the charges of the W boson and the b quark. The charge of the leptonically decaying W boson corresponds to the charge of the high-p_T muon from the hard process. The b-quark charge is only indirectly accessible. Hadronization prevents a direct charge measurement. However, charge correlated jet observables allow to estimate the b-quark charge. A muon associated with the semi-leptonic B hadron decay within the b jet ('soft muon') can be used to establish a robust discriminant between b- and $\bar{b}$-like jets. The discrimination power of the soft muon is diluted, e.g., due to cascade decays with semi-leptonic decaying D-mesons. Additionally, in this context, the p_T-weighted sum of all charges in a jet ('jet charge') is investigated as an alternative discriminant.

After a short introduction about the top quark and the experimental setup of the CMS experiment in Chapter 2 and Chapter 3 details of the used data and simulation samples are given in Chapter 4. Chapter 5 gives a summary of the reconstructed physics objects used in this analysis and the applied event selection. In addition the algorithm employed to associate the identified b jets to the two decay sides of the top pair is explained. In Chapter 6 the charge measurement of the b quark is presented before the top-charge measurement is performed in terms of an event categorization, described in Chapter 7. Results are presented in Chapter 8 together with a statistical interpretation in terms of excluding an exotic top quark charge scenario and a break down of the systematic uncertainties. The main results of this analysis are published by the CMS Collaboration as a Physics Analysis Summary [8].

System of Units

As it is common practice in particle physics this analysis uses natural units ($\hbar = c = 1$). Units of commonly used observables in this convention are:

$$[\text{Energy}] = [\text{Momentum}] = [\text{Mass}] = [\text{Time}]^{-1} = [\text{Length}]^{-1} = \text{eV} .$$

Chapter 2

Phenomenology of the Top Quark

2.1 Top-Quark Production

In $\sqrt{s} = 7\,\text{TeV}$ proton-proton collisions, top quarks are predominantly produced in pairs through strong interaction processes, whereas single top quarks are produced through weak interaction with about half the production cross section [9]. The background for single-top events is much more dominant than the background for top-pair events. This analysis focuses on top pair events.

The production process of top-quark pairs can be described by a factorization in long- and short-distance effects defined by the energy scale μ_F. Therefore, the total production cross section can be calculated as a convolution of the probability densities $f_{i,j}(x, \mu_F)$ for finding two partons i and j inside the colliding hadrons a and b and the cross section $\hat{\sigma}(\mu_F)$ for their interaction:

$$\sigma_{a,b \to t\bar{t}} = \sum_{i,j} \int_0^1 \int_0^1 \mathrm{d}x_i \mathrm{d}x_j f_i f_j \hat{\sigma}_{i,j \to t\bar{t}} \ . \tag{2.1}$$

x denotes the fraction of the hadron's longitudinal momentum, that is carried by the parton. The probability densities have to be measured experimentally while the partonic cross sections can be calculated from pertubative QCD using feynman diagrams. The diagrams for the leading order processes of top-pair production in proton-proton collisions, gluon-gluon fusion and quark-antiquark annihilation, are shown in Figure 2.1.

The production of top pairs resonates at a center of mass energy that corresponds to twice the top-quark mass. At LHC energies only a small fraction x of the proton's longitudinal momentum is needed for resonant production. The parton distribution function predicts a large gluon fraction at this regime. Therefore, the dominant production mechanism at LHC is gluon-gluon fusion.

At next-to-leading-order (NLO), there are additional contributions from qg and $\bar{q}g$-scattering processes [10]. The total top pair production cross section in proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$ has been calculated up to approximate NNLO to be 163^{+11}_{-10} pb [11]. Figure 2.2 shows that this is in good agreement with the current experimental results.
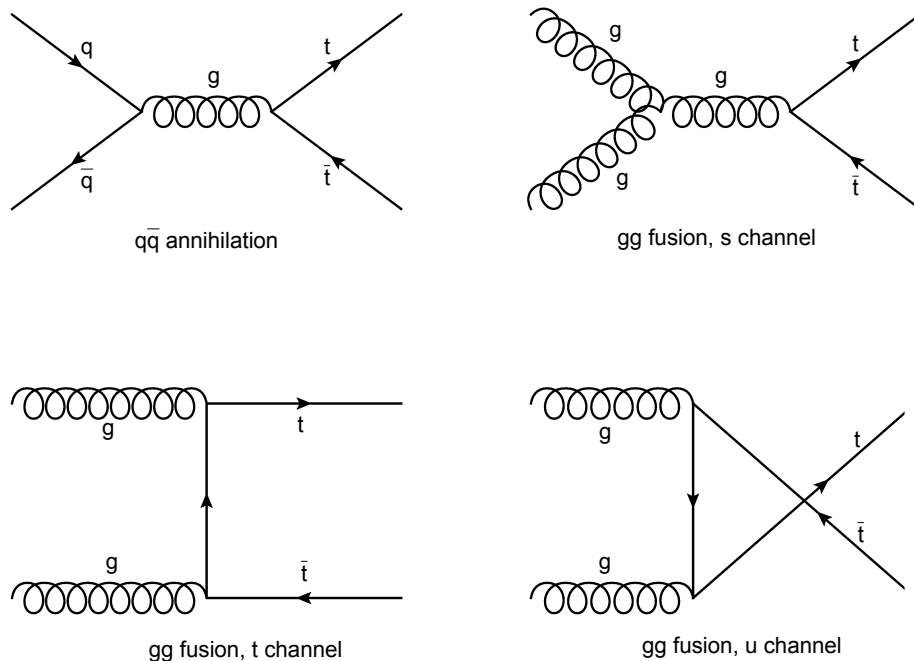
Figure 2.1: Leading order feynman diagrams for the pair production of top quarks.

2.2 Top-Quark Decay

The top quark decays through weak interaction. Due to forbidden flavor changing neutral currents in the standard model, the top quark decays via a W boson. The quark-mixing matrix (CKM) determines the branching ratios for the different decay channels. The CKM matrix is a unitary 3×3 matrix. The free parameters of the CKM matrix have been determined by various experiments. With the precise determination of V_{ts} and V_{td} it follows that top quarks decay to nearly a hundred per cent into b quarks. All direct measurements of V_{tb} confirm this so far [13] [14] [15].

The signature of top-pair events in the detector depends on the subsequent decay modes of the W boson. The W boson can decay into nine different isospin doublets in the leptonic and hadronic sector, which have approximately the same branching ratios. The fully-hadronic channel, where the W bosons from both top quarks decay into quarks, is enhanced by a factor of three due to the additional degree of freedom from color charge. This channel suffers from many combinatorial possibilities in the event reconstruction, a large amount of multijet background and strong sensitivity to uncertainties in the jet energy calibration. If both W bosons decay leptonically, referred to as dileptonic channel, the two high-p_T leptons provide a clear signature but the two neutrinos lead to a four-fold ambiguity in the kinematic event reconstruction. The semileptonic channel with a high-p_T lepton and four jets in the final state has a distinctive signature with a compatible branching ratio. The branching fractions of all channels are summarized in Table 2.1.
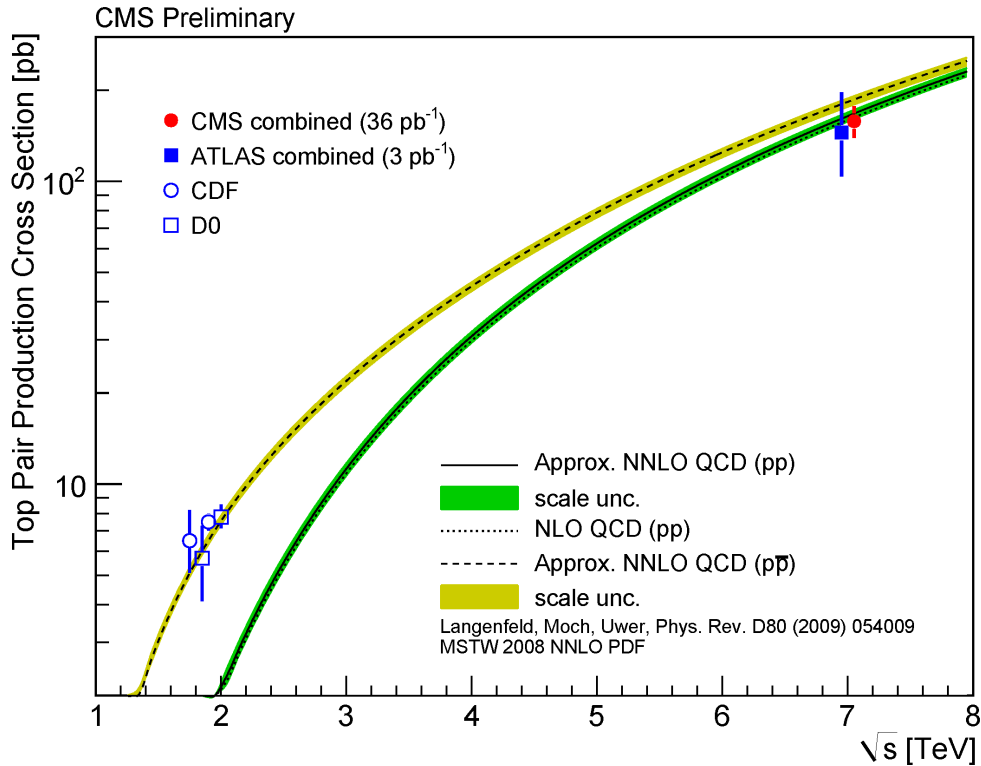
Figure 2.2: Standard model predictions and experimental results for the top pair production cross section in proton-(anti)proton collisions [12]. The data points are slightly displaced for better visibility.

2.3 Properties of the Top Quark

2.3.1 Mass

The high mass of the top quark leads to many prominent properties that define its special role within the standard model. For instance, the high mass determines the short lifetime of the top quark. Thereby, it becomes the only quark that decays before it hadronizes, thus, the only bare quark that can be investigated. Consequently, the common convention to define its mass is the pole mass as for a bare unstable lepton [10].

In perturbation theory a massive fermion propagator with a four momentum p has a pole at $\sqrt{p^2} = m - \frac{i}{2}\Gamma$. The real part of it corresponds to the invariant mass of the decay products, the so-called pole mass. The imaginary part corresponds to the decay width Γ. Nevertheless, the expectation that the top quark decays quickly enough to escape effects from non-pertubative QCD is not exact due to quark confinement [16]. In contrast to a bare lepton, the top quark carries a color charge. The color is passed on to the b quark. At some point of the hadronization process an interaction with the proton remnant is induced to form color neutral hadrons. Therefore, the invariant mass formed by these detected color-neutral decay products contains partially contributions from particles not arising from the direct top-decay chain. These long-range effects cause an irreducible ambiguity in the order of 200 MeV [17].

In contrast to the top quark, all other quark masses are defined by the short-distance

Table 2.1: Branching ratios of the different decay channels of the two W bosons from top-pair decays.

$t\bar{t} \rightarrow (W^+ b)\,(W^- \bar{b})$			W^+				
			$e^+\,\nu_e$	$\mu^+\,\nu_\mu$	$\tau^+\,\nu_\tau$	$u\,\bar{d}$	$c\,\bar{s}$
			1/9	1/9	1/9	3/9	3/9
	$e^-\,\bar{\nu}_e$	1/9	1/81	1/81	1/81	3/81	3/81
	$\mu^-\,\bar{\nu}_\mu$	1/9	1/81	1/81	1/81	3/81	3/81
W^-	$\tau^-\,\bar{\nu}_\tau$	1/9	1/81	1/81	1/81	3/81	3/81
	$\bar{u}\,d$	3/9	3/81	3/81	3/81	9/81	9/81
	$\bar{c}\,s$	3/9	3/81	3/81	3/81	9/81	9/81

mass convention. Avoiding such a theoretical ambiguity, in this convention the mass can be determined with arbitrary accuracy. Short distances are equivalent to high energies at which perturbation theory is applicable and can be used to calculate the quark mass. Arising divergences can than be removed using renormalization schemes under which the most common one is the so-called modified minimal subtraction ($\overline{\text{MS}}$) [18].

The difference in the top mass between the two conventions, $\overline{\text{MS}}$ and pole mass, is in the order of 6% [19]. Current measurements at hadron colliders preferably determine the top quark mass by the invariant mass of the decay products with a calibration to simulation that is based on leading-order matrix element calculations and parton showering. For that reason, the measured top mass is interpreted as a pole mass from the Particle Data Group [20]. In a possible future e^+e^--collider a precise determination of the top $\overline{\text{MS}}$ mass will become possible by scanning the production threshold of top pairs [10].

Meanwhile, the most precise determination of the top quark mass is performed by the Tevatron experiments with $m_t = (173.2 \pm 0.9)\,\text{GeV}$ [21].

2.3.2   Decay width and lifetime

The decay width is a measure for the particle decay probability and can be calculated analogical to cross sections.

At leading order perturbation theory, neglecting the b-quark mass, the decay width of the top quark is given by [19]

$$\Gamma(t \rightarrow bW)/V_{tb}^2 = \frac{G_F m_t^3}{8\pi\sqrt{2}} \left(1 - 3\frac{m_W^4}{m_t^4} + 2\frac{m_W^6}{m_t^6} \right) . \tag{2.2}$$

In leading order the decay width corresponds to the width of the invariant mass distribution of the top decay products. Given the masses of the top quark and the W boson, the decay width is in the order of $\Gamma \approx 1.4\,\text{GeV}$ which is beyond the resolution at current hadron colliders.

The decay width determines the lifetime via the relation $\Gamma = \frac{1}{\tau}$. Therefore, due to its high mass the top quark decays very quickly within $\sim 10^{-25}\,\text{s}$. Its lifetime is even shorter than

the time necessary to build a hadron ($\sim 10^{-23}$ s) [22] or to flip its spin [19] making the top quark a unique probe to test standard model predictions for a bare quark.

2.3.3 Spin

Due to angular momentum conservation and the known spins of the b quark and the W boson, the spin of the top quark is determined to be $1/2$. The measured differential cross sections so far are consistent with this prediction. Given that the top quark does not hadronize before the decay, spin polarization and correlation are passed to the decay products, resulting in characteristic angular distributions [23] [24]. The spin correlation can be measured through the spin correlation coefficient in terms of a coefficient A, defined as the fractional difference in the number of events where the top and antitop quark spins are aligned and those where the top quark spins have opposite alignment. ATLAS measures the strength of the spin correlation to be $A_{\text{helicity}} = 0.34^{+0.15}_{-0.11}$ [25]. CMS measures $A_{\text{helicity}} = 0.24^{+0.02}_{-0.08}$ [26]. Both are in agreement with the standard model prediction of $A^{\text{SM}}_{\text{helicity}} = 0.31$ [27].

2.3.4 Electric charge

In particle physics the electric charge defines the strength of the electromagnetic coupling. It is a conserved quantum number resulting from the global gauge invariance of the electromagnetic field. This was confirmed experimentally by searches for non-conserving decays [28]. As the top quark is interpreted as the isospin partner of the b quark it is assumed to carry a charge of $2/3$ e. While the b quark's charge was experimentally confirmed by the R-ratio measurement in Z^0 to $b\bar{b}$ processes by various experiments listed in [29], the center of mass energy of e^+e^- colliders has not been sufficiently large to produce $t\bar{t}$ pairs and therewith to confirm the charge of the top quark in an analogue way. A direct electric charge measurement at hadron colliders can be performed by investigating $t\bar{t}$ production with an additional photon radiated off the top quark which introduces the top charge at the top-photon vertex into the feynman diagram, see [30] [31].

An indirect evidence for the top quark to carry the electric charge of $+2/3$ e can be provided by exploiting the charge conservation law which leaves the possibility for the top quark with a charge of $-4/3$ e illustrated in Figure 2.3. This scenario has been excluded previously [4] [5]. In this thesis an exclusion with the highest significance is performed using data collected by the CMS experiment.
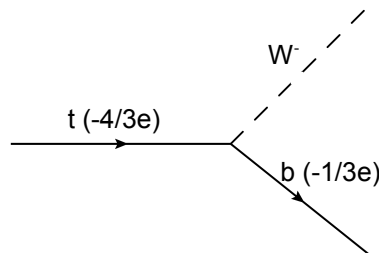

Figure 2.3: Decay of a top-like particle carrying an electric charge of $-4/3$ e.

Chapter 3

Experimental Setup

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [32] is a collider ring for protons and heavy ions operated by the European Organization for Nuclear Research (CERN). The two beam pipes of the LHC are located in a tunnel 100 m underground in the Swiss-French border region close to Geneva, as depicted in Figure 3.1. The counter-rotating beams inside the ring of 26.7 km circumference are brought to collision at four interaction points each surrounded by a complex detector system: There are two multipurpose detectors, ATLAS [33] [34] and CMS [35], one detector specialized for heavy-ion collisions, ALICE [36], and one detector dedicated to b-quark measurements, LHCb [37].
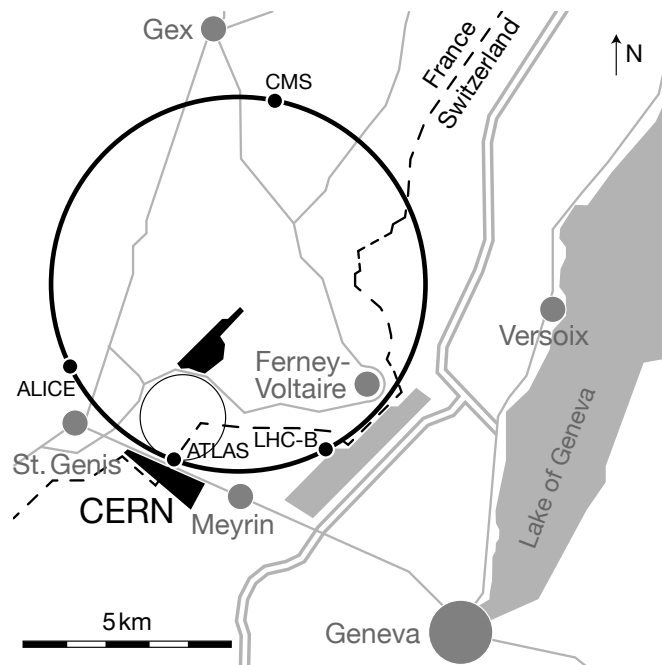

Figure 3.1: Location of the LHC and CERN sites near Geneva [38].

Pre-accelerated by a chain of linear accelerators and synchrotrons, the protons are injected into the LHC with an energy of 450 GeV. The data used in this thesis were collected in the 2011 runtime period with protons further accelerated by the LHC to an energy

of 3.5 TeV. The LHC is designed to reach proton energies of 7 TeV. With the help of 1232 superconducting dipole magnets providing a magnetic field up to 8.3 T these high energetic protons are bent onto their circular orbit.

In addition to the center of mass energy, the second main parameter characterizing the collider performance is the instantaneous luminosity $\mathcal{L}$. It determines the event rate N of a certain process with a cross section σ:

$$N = \sigma \cdot \mathcal{L} \tag{3.1}$$

The luminosity lifetime of a typical run in 2011 was about 6 hours. The decrease in the number of protons during a run is only partially due to collisions but also due to interactions with the environment, e.g., remaining beam gas in the super high vacuum of about $1.3 \cdot 10^{-13}$ bar inside the beam pipes.

In addition to the number of protons per bunch N_p, the instantaneous luminosity depends on the number of bunches n_b, the revolution frequency f and the beam profile A at the interaction point:

$$\mathcal{L} = \frac{N_p^2 \cdot n_b \cdot f}{A} \tag{3.2}$$

A depends on beam parameters such as the longitudinal bunch size, the transverse beam size and the beam crossing angle. The protons are traveling with almost the speed of light, leading to f in the order of 11 kHz.

In 2011 at maximum $1.45 \cdot 10^{11}$ protons per bunch were used and up to 1380 colliding bunches with 50 ns spacing. Figure 3.2 and 3.3 show the evolution of instantaneous and integrated luminosity of the LHC in 2011. At the end of 2011, an integrated luminosity of $5.5\,\text{fb}^{-1}$ was accumulated from which $5.0\,\text{fb}^{-1}$ are used in this analysis.

3.2   The Compact Muon Solenoid

In this thesis proton-proton collisions of the LHC are investigated using the Compact Muon Solenoid (CMS) apparatus [35]. CMS is a cylindrically shaped detector, radially symmetric around the beam axis and symmetric in both directions along the beam axis from the interaction point. It has a length of 21.6 m and a radius of 7.3 m. Figures 3.4 and 3.5 illustrate the arrangement of the main CMS components in the symmetry planes of the detector given in CMS coordinates.

CMS uses a right-handed coordinate system with the origin at the nominal interaction point, the center of the detector. The x-axis points to the center of the LHC ring, the y-axis points upwards perpendicular to the LHC plane, and the z-axis along the beam line in counterclockwise direction. The azimuthal angle, ϕ, is measured in the x-y plane relative to the x-axis in the range of $-\pi < \phi < \pi$. The polar angle, θ, is measured from the positive z-axis perpendicular to ϕ in the range of $0 < \theta < \pi$. The polar angle is often expressed in terms of the pseudo-rapidity η defined as $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$. Correspondingly, $\Delta R = \sqrt{\phi^2 + \eta^2}$ is used as measure for an angular distance.

3.2.1   Magnet

One central feature of the CMS experiment is a superconducting solenoid [41]. Together with the good tracker resolution it allows for precision measurements of charged parti-
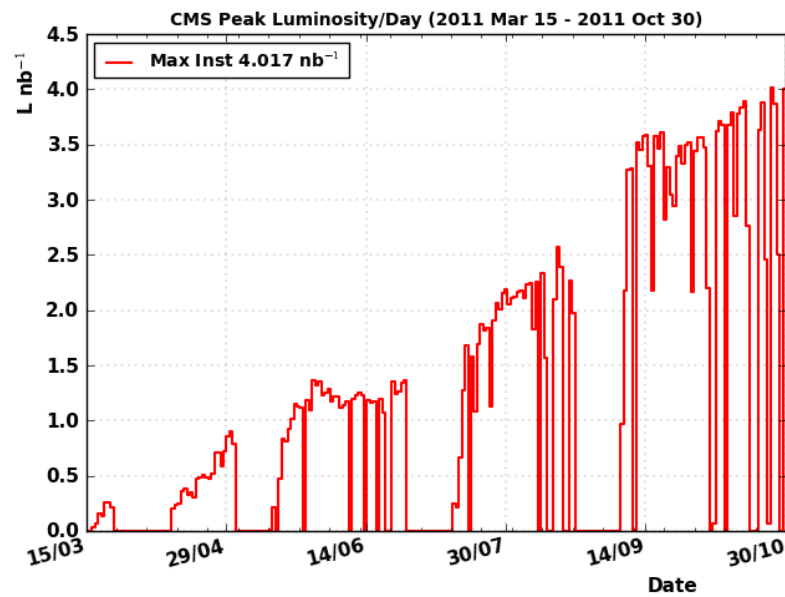
Figure 3.2: Maximum instantaneous luminosity per day delivered to CMS as a function of operation date in the 2011 run. [39]
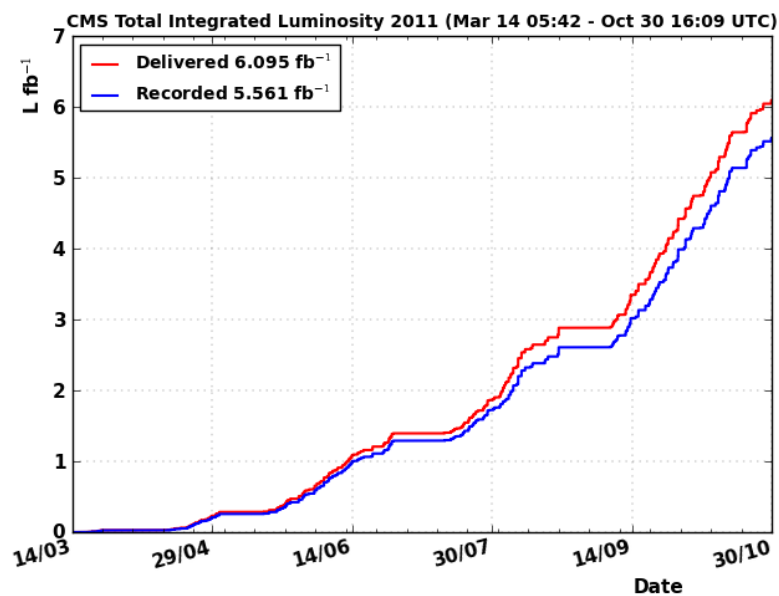

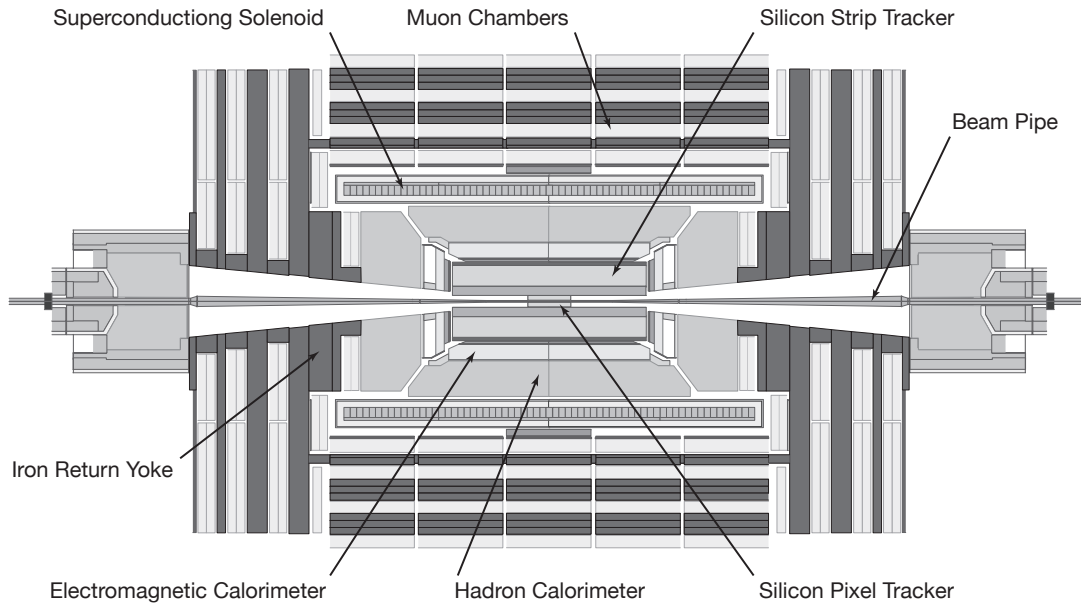**Figure 3.3:** Total integrated luminosity as a function of operation date in the 2011 run. [39]

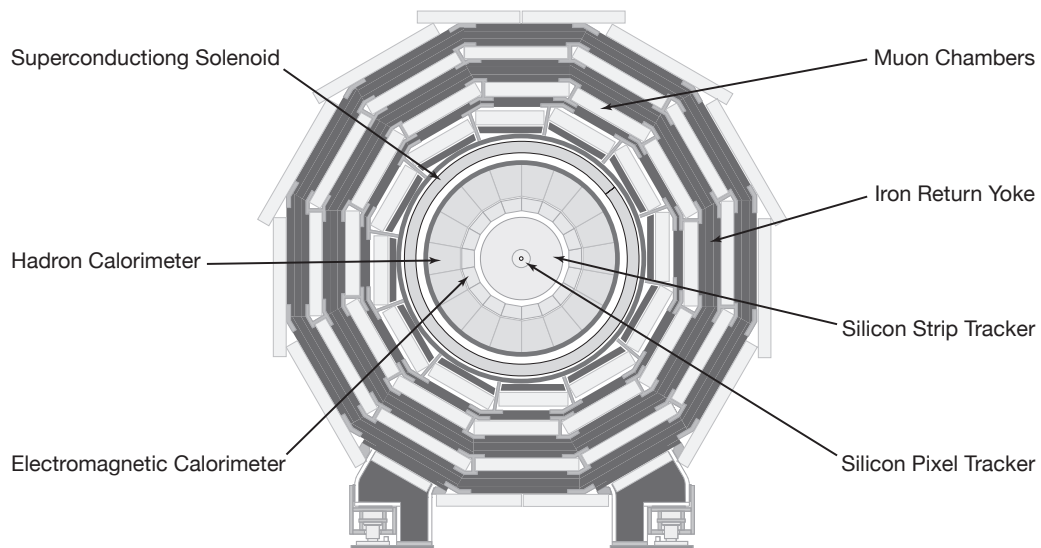**Figure 3.4:** Cross section of the CMS detector in the r-z plane [40].

Figure 3.5: Cross section of the CMS detector in the r-ϕ plane [40].

cle's momenta up to an energy of 1 TeV by bending their trajectories. To achieve this performance the solenoid provides a homogeneous field of 3.8 T within its inner volume of 3 m radius and 12.5 m length. The magnet comprises the silicon trackers and the main calorimeters. In this way, more sensitive material is located inside the magnet preventing a loss in precision due to multiple scattering. Additionally, the magnet material absorbs hadrons preventing them to enter the muon system. The support structure of the muon system serves as massive iron yoke to return the magnetic flux. The resulting magnetic field in the muon system has a strength of about 2 T and bends the muon tracks in the opposite direction as in the inner detector parts.

3.2.2 Inner tracking detectors

The tracking system [42] is the innermost part of CMS and therefore closest to the collision point. The small material budget in front of the sensors minimizes the amount of multiple scattering processes and therefore, enables a precise measurement of momenta and charge signs. A precise track measurement is essential for a high resolution in the reconstruction of interaction vertices to distinguish between primary and secondary vertices as well as between multiple interactions per bunch crossing, so-called pileup.

The inner tracking detectors cover a range of $|\eta| < 2.5$ and consist of a pixel detector starting at a radius of about 4 cm away from the beam axis up to a radius of about 10 cm and a strip detector surrounding it reaching a radius of 1.15 m, as shown in Figure 3.6. The active material of pixel and strip detectors is based on silicon.
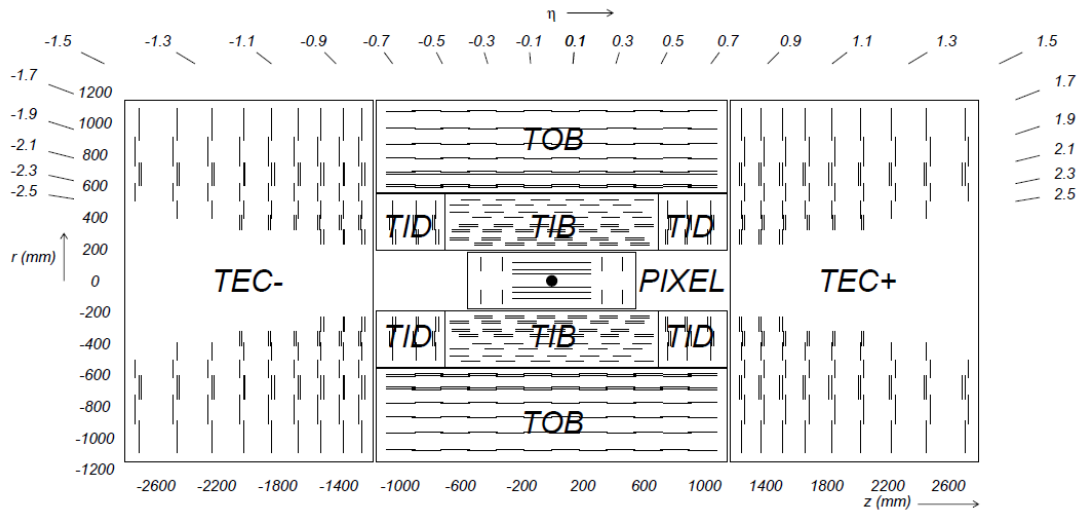

Figure 3.6: Inner tracking detectors of CMS in the r-z plane [43]. The pixel detector, the tracker inner barrel (TIB), outer barrel (TOB), inner disks (TID) and endcaps (TEC) are shown. The segments represent tracker modules. There are eight layers of double-sided strip modules along the whole z range.

The pixel detector system consists of 66 million pixels building an active area of about 1 m^2. They are arranged on three barrel layers and two discs at each end of the barrel. Therefore, the tracks within $|\eta| < 2.2$ have typically three pixel hits and tracks in the range of $2.2 < |\eta| < 2.5$ have typically two hits in the pixel detector. The pixel modules in the endcaps have a declination angle of 20° leading to a higher resolution. With a

dimension of $100 \times 150 \mu m^2$ of each pixel, an analogue readout and exploiting the Lorentz effect a single-point resolution of $15 - 20 \mu m$ is reached.

The silicon strip detector consists of 9.3 million strips building an active area of about $200 \, m^2$. The detector modules with either 768 or 512 strips are arranged in barrel layers and discs to provide typically around 9 hits for tracks with $|\eta| < 2.4$. The silicon strip detector is divided into four subsystems (TIB, TOB, TID, TEC) as depicted in Figure 3.6. These subsystems differ in the arrangement and types of modules. The modules up to a radius of $60 \, cm$ are $320 \mu m$ thick. With increasing radius the strips become longer leading to more noise. In order to guarantee a good signal-to-noise ratio, thicker sensors ($500 \mu m$) are used at radii larger than $60 \, cm$.

In some regions of the strip detector the modules are double-sided, see Figure 3.6. They consist of single-sided modules mounted back-to-back with a stereo angle of $100 \, mrad$. This way they provide information about the coordinate along the strips. With a strip spacing of $80 - 183 \mu m$ a single-point resolution of up to $23 \mu m$ in the r-ϕ plane and $230 \mu m$ in the z-direction is reached.

The expectation from detector simulation studies is that the momentum resolution of the tracker in the central region ranges from 1 to 5% at $1 \, GeV$ and $1 \, TeV$, respectively. The impact parameter resolution for high-momentum tracks is expected to be close to $10 \mu m$ [10]. The studies performed with data confirm this expected performance of the tracker [43]. Figure 3.7 shows as an example the agreement between data and simulation in the primary vertex resolution.
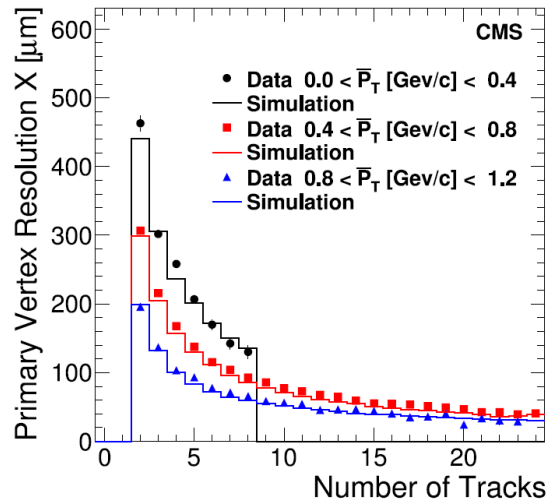

Figure 3.7: Primary vertex resolution in x as function of the number of tracks. [43]

3.2.3 Calorimeters

The tracking detectors are surrounded by two calorimetry layers to measure the particle energies, the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). The incoming particles deposit their energy in the calorimeters by a cascade of secondary particles. The compact layout of CMS with the calorimeters inside the solenoid requires materials capable to decelerate particles on a relatively short distance.

Starting at a radius of 1.29 m the ECAL [44] measures the energy of electrons and photons in a range of $|\eta| < 3$. The arrangement of the different ECAL subsystems is shown in Figure 3.8. An ECAL preshower system covers the range of $1.653 < \eta < 2.6$. This sampling calorimeter is dedicated to a discrimination between π^0 mesons, that decay into nearby photon pairs, and single photons by the use of silicon strip detector layers.
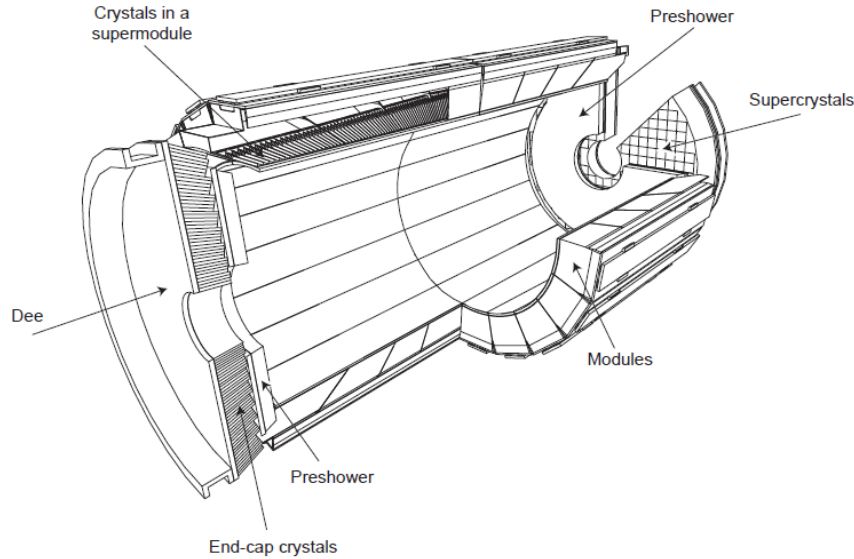

Figure 3.8: Schematic drawing of the electromagnetic calorimeter of CMS [45].

The remaining parts of the ECAL are based on scintillator crystals that are made of lead tungstate ($PbWO_4$). Lead tungstate has a radiation length of 0.89 cm and a Molière radius of 2.2 cm determining the dimension of the single crystals with a front face of $22 \times 22\,\mathrm{mm}^2$ at the one end and a length of 23 cm equivalent to 25.8 radiation lengths. The 61 200 crystals in the barrel and 7 325 crystals in the endcaps are pointing to the center of the detector while a small tilt in the axes reduces the number of particles going along gaps between the crystals.

Due to the small light emitting efficiency of the crystals (4-5 photoelectrons per MeV) the signal is amplified directly at the photon detectors by silicon avalanche photodiodes in the barrel and vacuum phototriodes in the endcaps.

The energy resolution of the ECAL can be parameterized as function of the incoming particle's energy E as [45]:

$$\left(\frac{\sigma(E)}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12\%}{E}\right)^2 + (0.3\%)^2 \quad (E \text{ in GeV}) . \qquad (3.3)$$

The first term accounts for the stochastic shower development, the second is due to the electronic noise. The constant term that comprises uncertainties in the calibration, dominates the resolution at high energies. The numbers in Equation 3.3 were obtained without a magnetic field using an electron test beam. Studies with the first $250\,\mathrm{nb}^{-1}$ of data at $\sqrt{s} = 7\,\mathrm{TeV}$ show that the relative difference between measured and simulated energy scale of the ECAL is in the order of 1% in the barrel and 3% in the endcaps [46].

The ECAL is enclosed by the hadronic calorimeter (HCAL) [47]. The chosen design of a sampling calorimeter is very compact and fits inside the solenoid. Plastic scintillators

serve as active material. The scintillator light is converted by integrated wave-length shifting fibers before being transmitted to the hybrid photo diodes through clear fibers. As absorber material brass has been chosen as it is nonmagnetic with a short interaction length. Figure 3.9 shows that the inner parts of the HCAL correspond to more than five interaction lengths. There is another layer of calorimetry outside the solenoid, denoted as outer hadron calorimeter (HO). It is used to detect late showering hadrons. The solenoid conforms to about three additional interaction lengths in this case.

The HCAL can be divided into a barrel part, endcaps and a forward-calorimeter.

With the forward-calorimeter a range up to $|\eta| < 5$ is covered nearly hermetically which enables a good reconstruction of missing transverse energy. The region covered by the forward-calorimeter is exposed to strong radiation. Therefore, a special design with iron and quartz has been chosen to detect the particles by their Cherenkov light.

The HCAL granularity is almost everywhere $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ providing a good di-jet mass resolution. For gauging the HCAL performance, it is common practice to study the performance of measuring the missing transverse energy presented in [48] and to study the jet-p_T resolution shown in Figure 5.2 or in [49].

3.2.4 Muon system

High energetic muons are not showering in the calorimeters like electrons due to their higher mass. As described by the Bethe equation they pass through the inner detectors with only minor energy loss by ionization, traverse the solenoid and enter the muon system [50]. In the muon system their curvature is opposite due to the return flux of the magnetic field through the iron yoke. In addition to the compactness of the detector, the iron yoke serves as shield against neutrons from the cavern and isolates the muon stations from the accompanying electro-magnetic showers.

Sufficient material is necessary in front of the muon stations to prevent hadronic punch-throughs and to guarantee a clear muon signature. At CMS there are at least 10 interaction lengths in front of the muon system and between 18 and 27 at the end, see Figure 3.9.

Three different gas detectors in different η-regions are embedded into the return yoke, see Figure 3.10. In the barrel region of $|\eta| < 0.8$ with a low muon rate and low residual magnetic field, drift tubes (DT) are situated. In the endcaps there is a high rate of muons and neutrons. Therefore, up to $|\eta|$ of 2.4, cathode strip chambers (CSC) are used which have a better spatial and time resolution (one bunch crossing). Between both the DTs and the CSC layers, there are resistive-plate chambers (RPCs) with an excellent timing resolution dedicated to triggering. 370 RPCs are installed in the barrel region and 432 in the endcaps. The resolution of the muon system as a function of the total momentum including information of the inner tracker is [50]

$$\frac{\Delta p}{p} = 0.045\sqrt{p} \ \ (p \text{ in TeV}) \ . \tag{3.4}$$

This equation describes the resolution for barrel and endcap region in the whole momentum range except for momenta below $70\,\text{GeV}$ in the endcap where the relative momentum resolution becomes a constant of 1.2%.
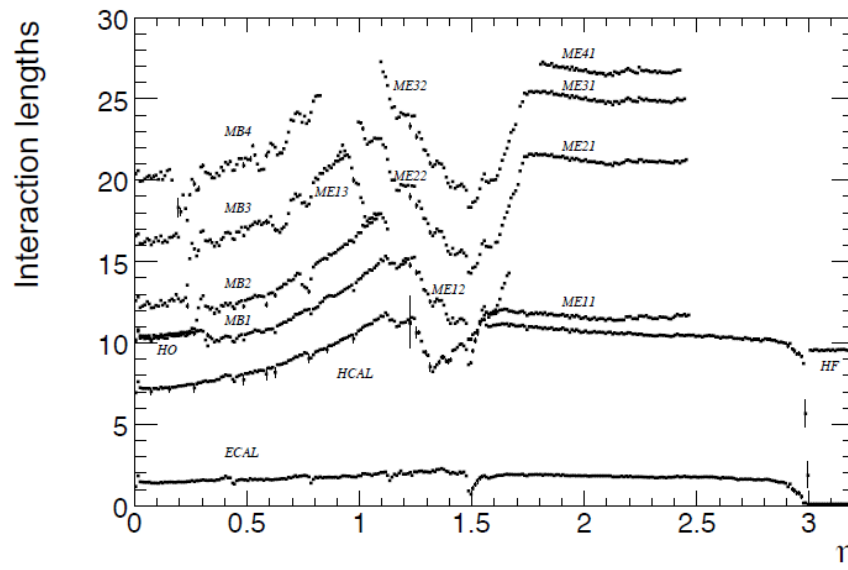
Figure 3.9: Material thickness in units of interaction lengths as a function of pseudorapidity after the ECAL, HCAL, and after each muon station. [35].

3.2.5   Trigger and data acquisition

The trigger and data acquisition systems of CMS [51] are designed to cope with 10^{10} interactions per second. Only $O(10^2)$ interactions per second can be written to archival media. This challenging reduction in data rate is achieved by a trigger and data acquisition system that consists of four parts: the detector electronics, the hardware-based level-1 (L1) trigger, the readout network and the software-based high-level trigger (HLT).

The L1 trigger uses information from the calorimeters and muon detectors with reduced granularity to select the most interesting events in less than $1\,\mu$s depending on primitive candidate definitions for photons, electrons, muons, taus and jets. Global sums of transverse energy and missing transverse energy are employed, too. While the L1 algorithms are running, the full detector information is buffered for about $3\,\mu$s.

The High Level Trigger further decreases the event rate from around $100\,$kHz to around $300\,$Hz before data storage. The full detector information is available for the HLT algorithms which use the L1 objects as seeds. These algorithms are simplified versions of the full reconstruction algorithms and produce similar results in the ideal case. The HLT processes the data in different paths. After each trigger step decisions are taken that can stop paths. Practically, this breaks the HLT down into a multi-level trigger system.
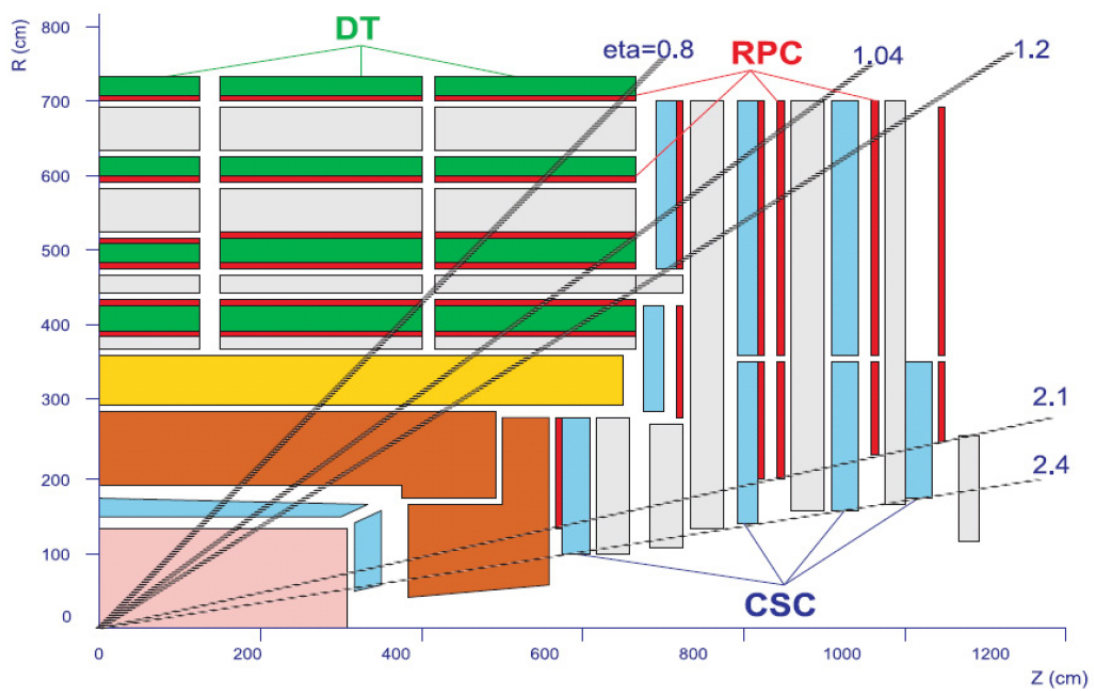
Figure 3.10: Cross section of the muon system in one quarter slice of the CMS detector with the drift tube chambers (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC) in the muon barrel (MB) and endcap (ME) systems [35].

Chapter 4

Collision Data and Event Simulation

The certified runs of proton-proton collision data taken in 2011 are considered for this analysis. The data collected during this period correspond to an integrated luminosity of $L = 5.0\,\text{fb}^{-1}$.

Simulation is necessary to compare the data with theoretical predictions and therefore to confirm or adjust theoretical predictions on a high precision level. The simulation needs to incorporate the statistical nature of a collision process and the detector response. The basic simulation steps and the used software is described in Section 4.1. The simulation of multi interactions per bunch crossing is explained in Section 4.2. Section 4.3 gives an overview of all simulated processes considered in this analysis. In the following the simulation is often referred to as MC.

4.1 Simulation Software

The simulation is performed in several steps, starting with the event generation that relies on perturbation theory. By the use of Feynman rules and Monte-Carlo methods it simulates the interaction process involving high-energetic, non-collinear particles. The outcome is represented as a list of four-momenta of the incoming and outgoing particles. In this analysis mostly the Monte Carlo matrix-element generator MADGRAPH [52] is used together with the leading order parton distribution function set CTEQ6L1 [53]. TAUOLA [54] is used for modeling the decay of tau leptons. The PYTHIA and POWHEG [55] [56] generators are used for some specific processes explained in Section 4.3.

In a next step, PYTHIA [57] models the initial and final state radiation (ISR/FSR) and the hadronization. These involve low energetic and collinear processes that can not be calculated with perturbation theory and are described by phenomenological models. In order to study systematic effects due to the hadronization model, another signal sample is investigated that uses HERWIG for the showering and hadronization. The MLM matching procedure [58] is performed to avoid double counting in the showers from matrix element generation and PYTHIA showering. The Z2 tune is used [59] to model the underlying event.

After the generation the events are passed to the GEANT4 [60] detector simulation. The output is in the same format as the measured data.

4.2   Pileup Reweighting

The simulated events are overlaid with a random number of pileup events. The events need to be re-weighted to describe the measured pileup environment in data. In Figure 4.1 the number of reconstructed vertices is shown without re-weighting the simulation. After re-weighting there is a better agreement between data and MC, see Figure 4.2 and Figure 4.3.

Two different pileup reweighting methods are investigated: *1BX* and *3D*. While the *1BX* procedure only considers pileup within one bunch crossing the *3D* method takes also pileup into account that is due to the bunch crossings before and after the main bunch crossing therefore it is a more precise description.

Figure 4.2 shows that the total event yield is better modeled by the *3D* method. Figure 4.3 shows that the shape is better described by the *1BX* method. The *3D* method is used in the top-charge analysis since it is more elaborated. The data-driven validation on multijet data described in Section 6.4 uses simulation with an older software version. This allows only the usage of the *1BX* method.

The difference in the top-charge result by using the *1BX* method is taken as systematic uncertainty. Additionally, the *3D* reweighting is varied within its uncertainties, see Section 8.2. This way the possible effect on the result due to the bad data-MC agreement in the vertex multiplicity is covered by systematic uncertainties. Since the total yield of events is not relevant in this analysis all following data-MC comparisons are performed by normalizing the total number of simulated events to the total number of data events. This is indicated in the plots by the label "Scaled to Data".
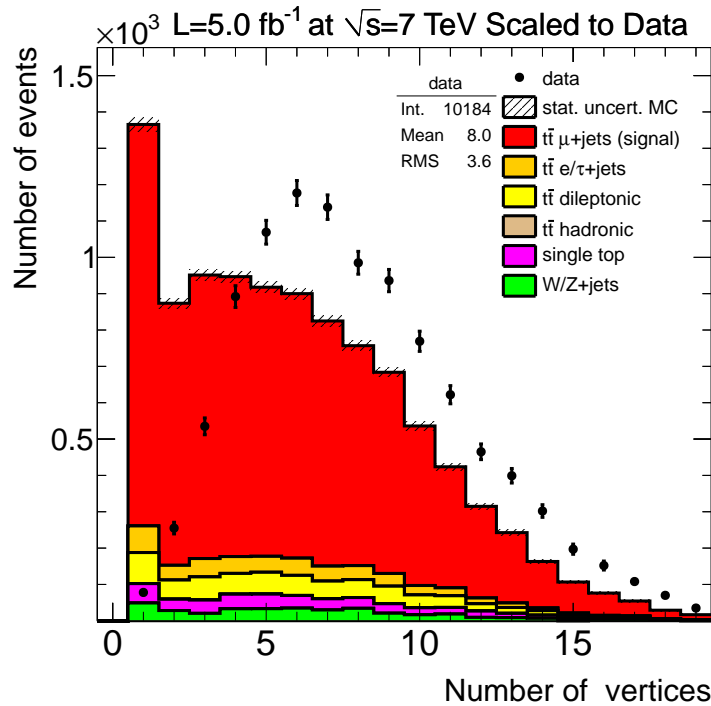

Figure 4.1: Data-MC comparison of the number of reconstructed vertices without re-weighting after the top-pair selection.
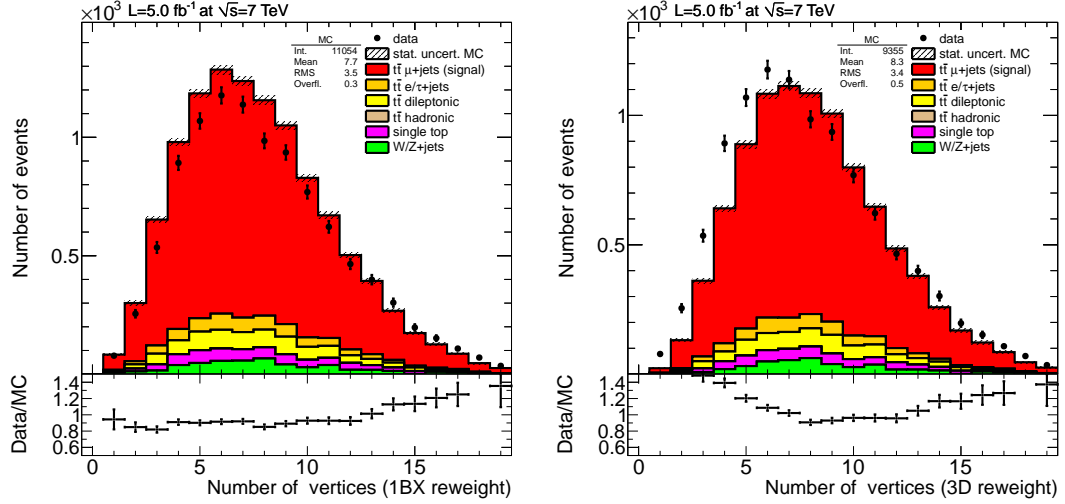
Figure 4.2: Data-MC comparison of the number of reconstructed vertices with *1BX* (left) and *3D* (right) re-weighting after the top-pair selection normalized to the integrated luminosity.
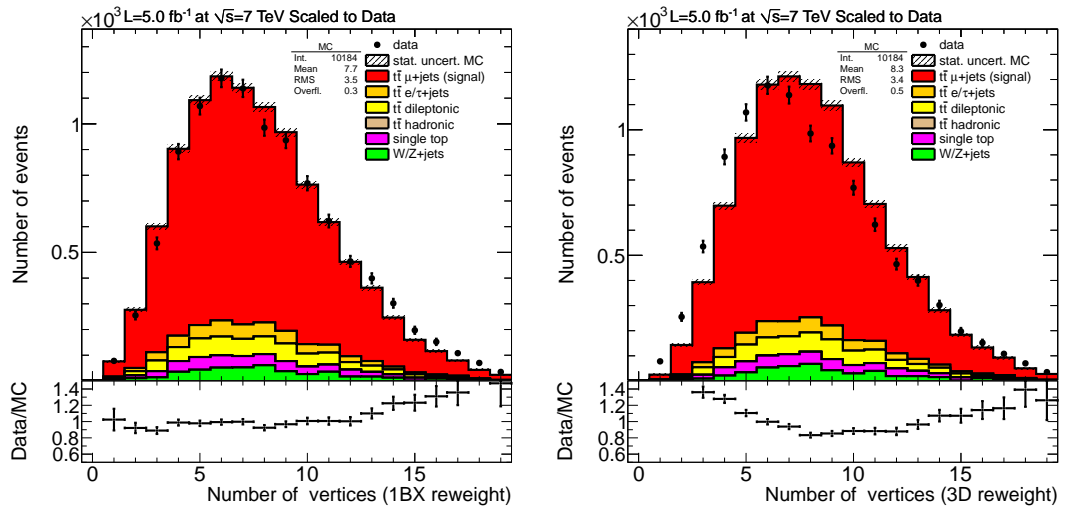

Figure 4.3: Data-MC comparison of the number of reconstructed vertices with *1BX* (left) and *3D* (right) re-weighting after the top-pair selection normalized to data.

4.3　Simulated Processes

The simulated processes considered in this analysis consist of a top pair signal and signal-like background events.

The signal signature is shown in Figure 4.4. It consist of a high-p_T muon, a muon neutrino, two b jets and two light jets in the final state in the following referred to as muon+jets channel.
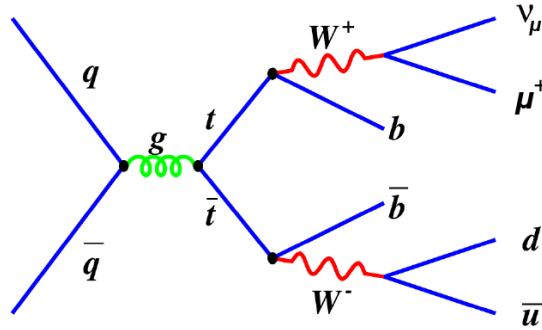

Figure 4.4: Decay topology of the considered signal process. This analysis uses a muon in the final state.

A leptonically decaying W boson associated with jets (W+jets) has a signal-like signature with four jets and an isolated muon in the final state. A leptonically decaying Z^0 or γ associated with jets (DY+jets) can resemble a signal event as well if one one of the leptons escapes due to the limited detector acceptance. A multi-jet event (QCD) can be misinterpreted as signal for instance due to leptonically decaying B and D mesons. The multi-jet events are modeled by PYTHIA with a muon required on generator level with p_T larger than $20\,\text{GeV}$. This enriches events relevant for the analysis before the time-consuming simulation step. Single top events are generated with POWHEG [55] [56]. Diboson processes are neglected in this analysis due to their relatively small cross section. Table 4.1 and 4.2 summarizes the assumed cross sections for the generated samples. The cross section for top pairs and single top quarks are determined with MCFM [61] within the CMS Collaboration to be consistent with the event simulation. The cross sections for the W/DY+jets processes are determined with FEWZ [62]. The assumed top-quark mass is $172.5\,\text{GeV}$. Signal samples with a top-quark mass of $169\,\text{GeV}$ and $175\,\text{GeV}$ are generated for systematic studies.

Table 4.1: Predicted cross sections of the considered physics processes. The given uncertainties include factorization and renormalization scale uncertainties and uncertainties on PDFs and α_S.

Sample	$t\bar{t}$	W+jets	DY+jets	QCD
σ [pb]	157.5^{+23}_{-24}	31314 ± 1600	3048 ± 130	84679.3
Order	NLO	NNLO	NNLO	LO

Table 4.2: Predicted cross sections of the considered single top processes. The given uncertainties include factorization and renormalization scale uncertainties and uncertainties on PDFs and α_S.

Sample	$t_{s\text{-channel}}$	$\bar{t}_{s\text{-channel}}$	total t channel	$t_{\text{tW}}^{(-)}$
σ [pb]	$2.7^{+1.1}_{-1.0}$	$1.4^{+0.09}_{-0.08}$	$64.6^{+3.4}_{-3.2}$	7.9 ± 0.6
Order	NLO	NLO	NLO	NLO

Chapter 5

Event Reconstruction and Selection

In proton-proton collisions thousands of particles arise that enter the detector. The raw data that they produce consist of hits and energy depositions. The specific arrangement of the various subdetectors allows to distinguish between the particles by the use of complex algorithms that exploit the characteristic particle signatures.

The physics objects in this analysis are reconstructed via the particle-flow algorithm [63]. Starting with the most distinct signatures, like those arising from muons, each particle in the detector is identified. In a next step, the algorithm sorts the individually reconstructed particles in object collections by requiring certain quality criteria. Again, this is performed starting from simple to more complex particle identification.

The definition of these object collections has to be adopted to the analysis needs. In the following sections the definitions for the physics object used in this analysis are given. Using the so-defined objects a selection is setup to identify top-pair events in the muon+jets decay channel, see Section 5.6. Finally in Section 5.7, the event topology is reconstructed as a whole by assigning the jets to the final state objects of the top-quark decay by characteristic b-jet properties and by exploiting the top-quark mass.

5.1   Vertex

Primary vertices at CMS [64] are reconstructed from tracks that are selected based on their impact parameter with respect to the beamline, the number of hits in the inner tracking system and the normalized track χ^2. The tracks are combined according to their z-coordinate at the point of closest approach to the beamline. If their nearest neighbor is within a distance of $z = 1\,\mathrm{cm}$ the tracks are grouped together. Each vertex candidate with more than two tracks is fitted with an adaptive vertex fit, in which the tracks are contributing according to a weight w_i between 1 and 0 depending on their compatibility with the common vertex. In addition to the best estimate for the vertex position the fit provides indicators for the success of the fit such as the number of degrees of freedom defined as $n_{dof} = 2\sum_{\text{tracks}} w_i - 3$. The vertex with the highest sum of track p_T is considered as primary vertex. A primary vertex in this analysis fulfills $n_{dof} > 4$ and is located within $|z| < 24\,\mathrm{cm}$ and $\rho < 2\,\mathrm{cm}$ around the nominal interaction point. Moreover, all charged objects pointing to pileup vertices are rejected in the subsequent reconstruction steps.

5.2 Muons

There are three different types of reconstruction algorithms for muons in CMS [65]. Besides the so-called standalone muons, which are reconstructed using only information from the muon system, 99 % of all muons that occur in collisions are based on information from the inner tracking system and the muon stations. Two approaches exist to combine the indeptly reconstructed tracks of the two systems:

Global muon (outside-in): Starting from the standalone muon track in the muon stations a matching inner track is associated. A global track is determined by a fit using the hits from both tracks.

Tracker muon (inside-out): Tracks are extrapolated to the muon system by taking into account the expected energy loss and multiple scattering effects. If at least one muon segment matches, the track is assigned as a tracker muon.

In this analysis a muon must be identified as both tracker and global muon. The inner tracks must have at least 10 hits out of which one has to be in the pixel detector. The global track must have at least one hit in the muon system. The global fit should have a normalized χ^2 of less than 11. The distance to the primary vertex is required to be $\Delta z < 0.5$ cm and the two-dimensional impact parameter with respect to the average beam spot must be smaller than 0.02 cm. The muons should be isolated to distinguish between muons that appear in jets due to leptonic b and c decays. A combined isolation is defined as:

$$\text{relIso} = \frac{\text{hadronIso} + \text{photonIso}}{p_\text{T}} \tag{5.1}$$

hadronIso (photonIso) is the p_T sum of all hadrons (photons) within a cone of $\Delta R < 0.4$ around the muon. relIso should be smaller than 0.125. In addition, the muons should have at least a distance of $\Delta R > 0.3$ to the next jet as defined in Section 5.4.3.

The high-p_T muon is expected to trigger the event. Additional requirements guarantee that the selection is tighter than the trigger and thereby avoid systematic uncertainties in the modeling of the trigger turn-on regions. The unprescaled trigger bit used in this analysis has a threshold of $p_\text{T} > 24$ GeV, therefore the muons should have at least $p_\text{T} > 26$ GeV. At least two muon stations should have a matched segment, since the DT require two segments to trigger. Finally, the muons should be in the central region $|\eta| < 2.1$ up to which the triggers are installed.

There is a veto using a looser muon definition to suppress backgrounds e.g. leptonic decaying Z^0 bosons with two leptons in the final state. The muons should be global with $p_\text{T} > 10$ GeV, $|\eta| < 2.5$ and relIso < 0.2.

Another class of muons plays an important role in this analysis, the muons that occur in B-meson decays within b jets. They are allowed to fulfill looser quality criteria which are described in Section 6.2.

5.3 Electrons

The electron reconstruction at CMS uses information from the inner tracking system and the electromagnetic calorimeter. A description of the reconstruction algorithms can be

found in [66]. Electrons are only considered in the selection to suppress backgrounds e.g. from dileptonic top pair events. There is a veto on electrons that have a $E_T > 15\,\mathrm{GeV}$, $|\eta| < 2.5$ and relIso < 0.2.

5.4 Jets

The quarks and gluons from the hard-scattering process cannot be detected directly due to confinement. Through gluon radiation and splitting they build a collection of color neutral hadrons before entering the detector. The resulting particles originating from a high-energetic quark or gluon are called jets due to their collinearity. They serve as an estimator for energy and momentum of the initial parton from the hard scattering process. The jets (PF jets) at CMS are reconstructed in the last step of the particle-flow algorithm. Having sorted out all isolated particles in single particle collections the remaining non-isolated particles are clustered into jets using the anti-k_t clustering algorithm.

5.4.1 The anti-k_T algorithm

The anti-k_T algorithm [67] defines two distance measures: d_{ij} between each pair (i,j) of entities (particles or pseudojets) and d_{iB} between each particle and the beamline. If the smallest of all distances is d_{ij}, i and j are combined. If it is d_{iB}, i is called a jet and is removed from the list. The two measures are defined as

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta R^2(ij)}{\rho^2} \,, \tag{5.2}$$

$$d_{iB} = k_{ti}^{2p} \,. \tag{5.3}$$

k_{ti} corresponds to the transverse momentum of particle i. ρ is a radius parameter and is set to $\rho = 0.5$ in this analysis. The parameter p determines the relative power between energy (k) and spatial scales ($\Delta R = \sqrt{(\phi_i - \phi_j)^2 + (y_i - y_j)^2}$, ϕ is azimuth, y is rapidity). Its value of -1 is eponymous for the algorithm. Among other advantages this algorithm provides regular cone sizes, which are useful for example in the calibration of the jets, and it is infrared and collinear safe. This means stability against small changes in the set of input particles by adding or removing low energetic particles or if two particles are measured as one. This is particularly important due to the limited detector resolution and granularity as well as for low energetic, statistical processes such as pileup interactions or soft gluon radiation.

5.4.2 Jet-energy corrections

Imperfections in the measured jet energy are due to the non-uniform and non-linear response of the CMS calorimeters. Furthermore, additional energy can arise e.g. from pileup interactions. A calibration is necessary to account for these effects. The aim is to reproduce as accurately as possible the energy of the true particle jet. A true particle jet is based on all stable particles from the fragmenting parton and from the underlying event activity, clustered with the same algorithm applied to the detector jets. CMS uses a factorized approach to calibrate the jet energies [49].

In a first step, the total energy of the jet is corrected to account for extra energy not arising from the high-p_T scattering such as from electronic noise and pileup. Then the bulk of the η and p_T dependence of the jet-energy response is removed by using the simulated QCD events. The correction factors for PF jets are within the interval of $[1.0, 1.1]$ and much smaller than in a previous jet definition (CALO) that uses calorimeter information only.
A comparison of the jet-energy response between data and MC shows differences e.g. up to 10% depending on η. Therefore, in data another correction step is needed to account for these differences. These residual corrections are extracted from data. By exploiting the transverse-momentum balance in dijet events the η dependence of the relative jet response can be removed. The difference in total jet energy response is removed by exploiting the transverse-momentum balance in Z^0/γ+jet events. The energy of Z^0 or γ are used as reference objects to calibrate the jet because their energy can be measured with much higher precision. The total corrections factors are shown in Figure 5.1.
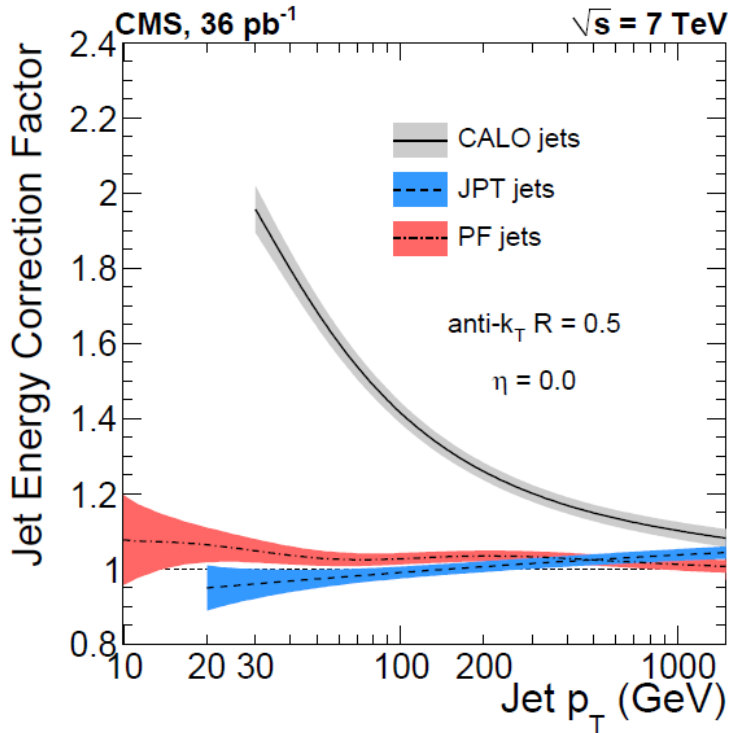

Figure 5.1: Total jet energy correction factor as a function of the jet p_T exemplified for $\eta = 0$. JPT jets are calorimeter jets with a corrected energy response and resolution by incorporating tracking information. [49]

5.4.3   Jet selection

The following identification criteria help to distinguish real hadronic jets from instrumental noise while retaining the vast majority of real jets. This is validated in pure noise, non-collision data from e.g. cosmic triggers or from triggers on empty LHC bunches [68]. The charged hadron energy fraction should be larger than zero, while the charged and neutral electromagnetic energy fraction should be smaller than 99 %. The neutral hadron energy fraction is required to be smaller than 99 % among others to discriminate jets from

neutral pions. The PF jets should have at least two constituents and one of them must be charged. Finally, the jets should have $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$.

5.4.4 Jet-energy resolution

The jet energy resolution is found to be larger in data than in simulation [49]. Therefore, the jet energies in simulated events are corrected according to the measured resolution. In Figure 5.2 the resolution of the transverse momentum is shown before and after increasing the resolution for reconstructed jets matched to jets on generator level (genJet). In this analysis the jet-resolution correction is performed after the top-pair selection for practical reasons. Large systematic uncertainties are assumed on the background fraction to cover the influence of the jet energy resolution on the top-pair selection.
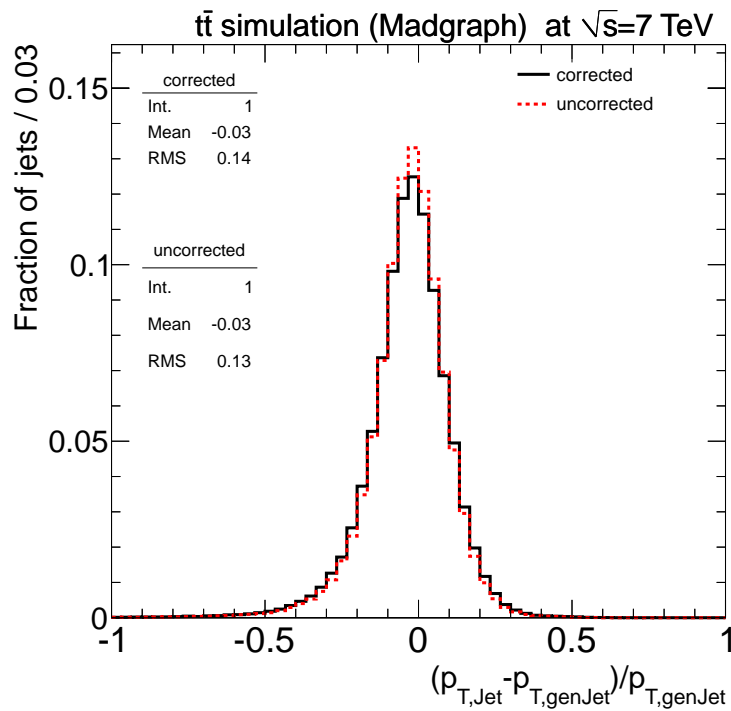
Figure 5.2: Jet resolution before and after correction normalized to unit area.

5.5 b-Jet Identification

The identification of the two b jets in top-pair decays is a prerequisite to the charge determination of the b quark and finally to the charge determination of the top quark. b quarks form distinctive jets in the detector that differ significantly from jets arising from light quarks or gluons. The b-jet characteristics arise from a relatively long lifetime and the relatively large mass of the b quark. The weak decay of the b quark is suppressed by the CKM matrix. Therefore, the b quark hadronizes and travels some distance before it decays. The high resolution of the pixel detector allows to measure the decay length. As a result displaced tracks originating from a secondary vertex are strong hints on the

existence of a b jet. Several algorithms exploit these characteristics to identify b jets [69]. In this analysis the robust track-counting algorithm is used.

5.5.1 Track-counting algorithm

The track-counting algorithm (TC) uses well-reconstructed tracks of a jet to calculate its b-likeliness. Details on the track-selection criteria can be found in [7]. For each selected track the three dimensional impact parameter and its uncertainty are calculated (see Figure 5.3). The algorithm ranks the tracks by decreasing impact parameter significance. The significance of the n-th track serves as discriminator. Usually $n = 2$ is chosen to reach a high efficiency (track counting high efficiency, TCHE) while $n = 3$ results in a high purity of the b tag. The sign of the discriminator value is given by the sign of the scalar product between the jet direction and the direction of the impact parameter. The distribution of this TCHE discriminator is shown in Figure 5.4.

Standard operating points of the TCHE discriminator at CMS are defined in [7]. The medium working point of TCHE > 3.3 is used for this analysis. At this working point light flavored jets from QCD Monte-Carlo simulation with a p_{T} of about $80\,\mathrm{GeV}$ are accepted in about $1\,\%$ of the cases. The resulting b-tagging performance in this analysis is validated in Section 5.5.3.
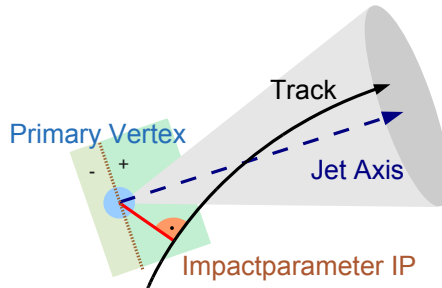

Figure 5.3: A sketch to illustrate the impact parameter definition of a track contributing to the jet. The second highest track impact parameter significance is used as discriminator in the TCHE algorithm. [70]

5.5.2 Generator matching

For testing purposes it is useful to link the reconstructed object to the original generated particle referred to as generator matching. There is no direct link between the two, a result of the generation and simulation process. Therefore, an algorithmic assignment procedure needs to be developed.

In this analysis a generator matching of the b jets is useful to test the performance of b tagging, jet association and b charge assignment. This analysis assumes that the two jets with the highest b tag originate from the two b quarks involved in the hard process. Therefore, only these two jets and the two generator b quarks from the top-pair decay are considered for the matching. The jet with highest b tag, referred to as first b jet, is matched to the b quark which is closest in ΔR. The jet with the second highest b tag, referred to as second b jet, is matched to the remaining b quark.
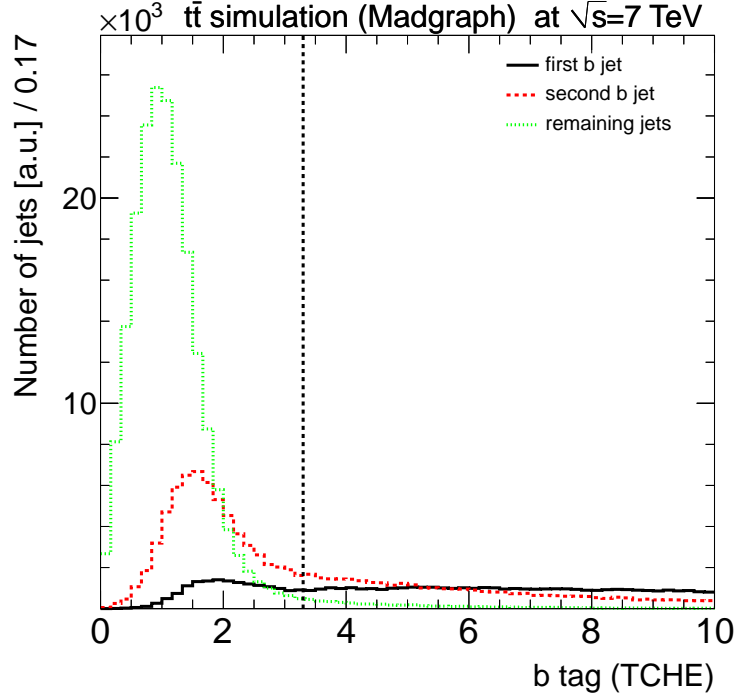
Figure 5.4: Distribution of the b discriminator (TCHE) for the first and second b jet in the event and all remaining jets. The vertical dashed line indicates the working point of this analysis.

Figure 5.5 shows the performance of this procedure in terms of ΔR between the b jet and the b quark for both possible b-jet-b-quark combinations. Obviously, the combination that results from the algorithm peaks at zero whereas the other combination, a random assignment, accumulates at approximately π. Figure 5.6 shows that the relative transverse momentum resolutions of the matched b-jet-b-quark pairs is balanced. The energy resolution between the b quark and the b jet is approximately $20\,\%$. The b-jet matching is defined to be successful if $\Delta R(b\,\text{jet}, b\,\text{quark}) < 0.25$.

5.5.3   Performance of the b-tagging algorithm

Two quantities are defined to validate the b-tagging performance in this analysis. At first, P_{btag} gives the probability that a b-tagged jet matches a b quark. Figure 5.7 shows that P_{btag} reaches a plateau at a b tag > 3.3, the working point of this analysis. ϵ_{btag} defines the selection efficiency of a real b jet to be selected by the b-tagging algorithm. Figure 5.8 shows that the efficiency at the working point is about 90%.

All events are required to have at least two b jets. They are considered to stem from the two b quarks of the top pair decay process. The b jet with the higher b tagging value is referred to as first $b$jet in the following. The first b jet matches in $P_{\text{btag}}^{\text{bjet1}} = 88.3 \pm 0.2(\text{stat.})\%$ of the cases a b quark, the second b jet in $P_{\text{btag}}^{\text{bjet2}} = 77.0 \pm 0.2(\text{stat.})\%$.
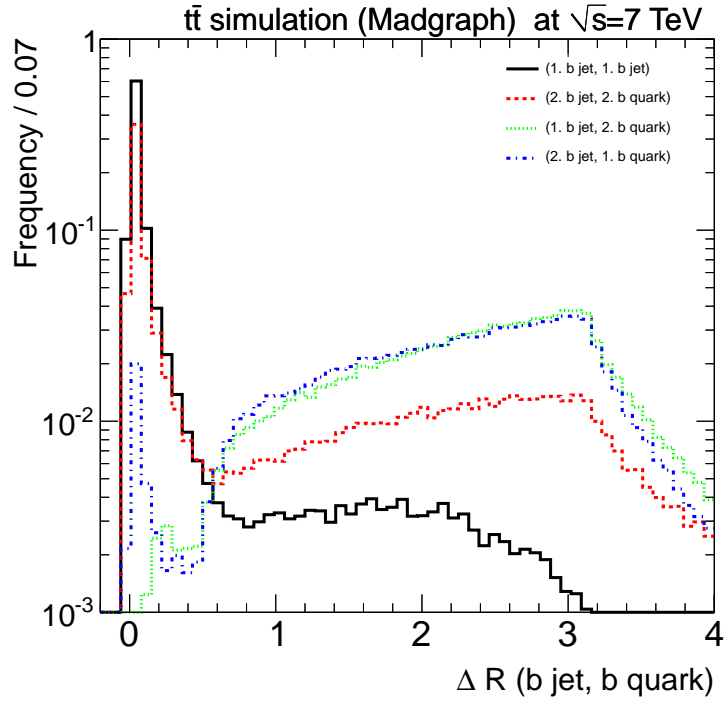
Figure 5.5: ΔR between the first b jet (1. b jet) and its matched b quark (1. b quark), ΔR between the second b jet (2. b jet) and its matched b quark (2. b quark) and ΔR of the unmatched combinations (1. b jet, 2. b quark) and (2. b jet, 1. b quark). The distributions are normalized to unit area.
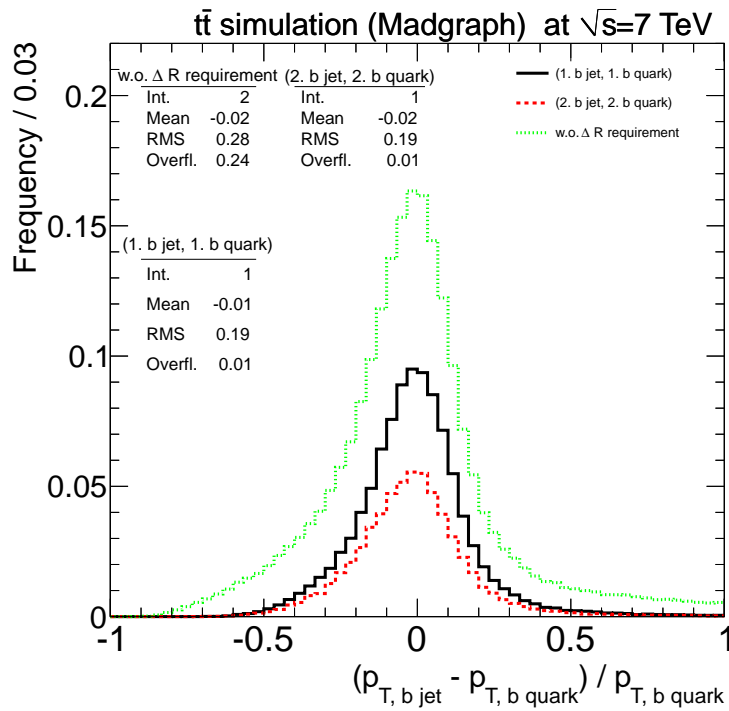

Figure 5.6: The relative difference in p_{T} between the two b jets and their matched generator b quarks. With and without the matching requirement of $\Delta R < 0.25$.
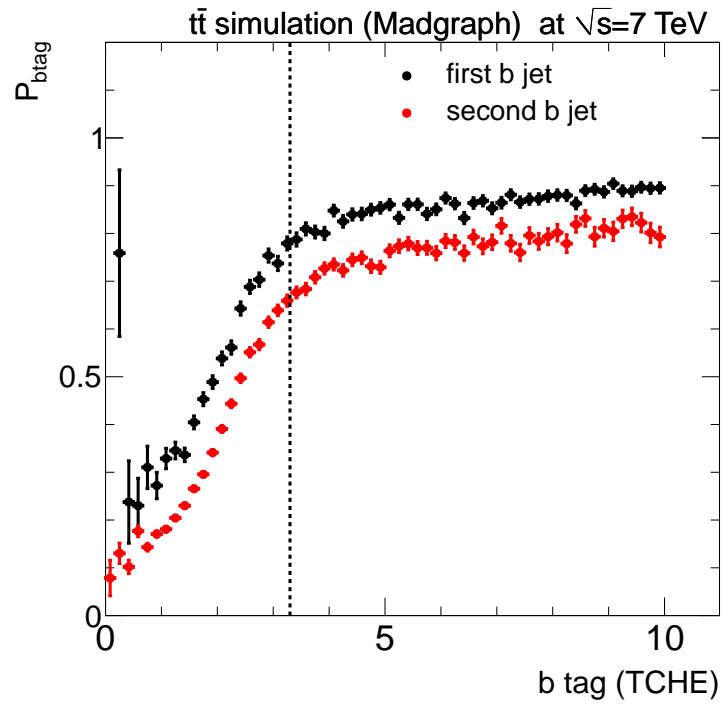
Figure 5.7: P_{btag} as function of the b tag value. The vertical dashed line indicates the working point of this analysis.
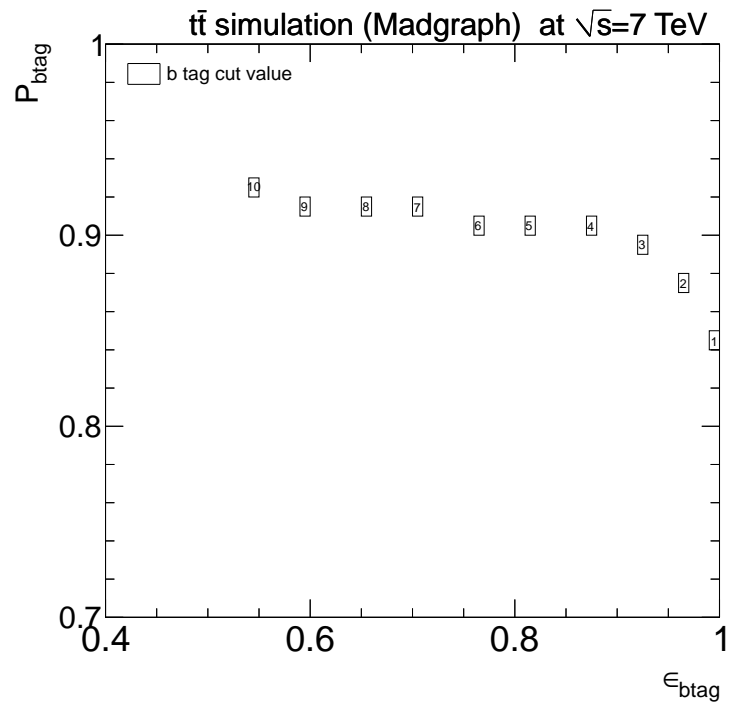
Figure 5.8: P_{btag} as a function of ϵ_{btag}. The marker in the form of a number corresponds to the respective b-discriminator value (TCHE).

5.5.4 b-Tagging corrections

The efficiencies and mistag rates of the b-tagging at CMS have been measured previously [7]. The results are p_T and η dependent correction factors SF_i to correct the single-jet efficiencies. The method and implementation to apply these factors in order to correct a two-jet efficiency is adopted from a top pair cross section analysis [71].

The two-jet efficiency $\epsilon_{\geq 2\text{b-tags}}$ is expressed in terms of the single-jet efficiencies as shown in Equation (5.4) and (5.5).

$$\epsilon_{\geq 2\text{b-tags}} = 1 - \epsilon_{<2\text{b-tags}} \tag{5.4}$$

$$
\begin{aligned}
\epsilon_{<2\text{b-tags}} = &\prod_i^{N_b}(1-\epsilon_{b,i}) \prod_j^{N_c}(1-\epsilon_{c,j}) \prod_k^{N_l}(1-\epsilon_{l,k}) \\
&+ \sum_h^{N_b} \epsilon_{b,h} \prod_{i\neq h}^{N_b}(1-\epsilon_{b,i}) \prod_j^{N_c}(1-\epsilon_{c,j}) \prod_k^{N_l}(1-\epsilon_{l,k}) \\
&+ \sum_h^{N_c} \epsilon_{c,h} \prod_i^{N_b}(1-\epsilon_{b,i}) \prod_{j\neq h}^{N_c}(1-\epsilon_{c,j}) \prod_k^{N_l}(1-\epsilon_{l,k}) \\
&+ \sum_h^{N_l} \epsilon_{l,h} \prod_i^{N_b}(1-\epsilon_{b,i}) \prod_j^{N_c}(1-\epsilon_{c,j}) \prod_{k\neq h}^{N_l}(1-\epsilon_{l,k})
\end{aligned}
\tag{5.5}
$$

The first line in Equation (5.5) corresponds to the number of cases where no jet is tagged as a b jet. The remaining three rows correspond to the cases where exactly one of the jets, either a b, c or light jet is tagged as a b jet. N_b, N_c and N_l denote the number of b, c, and light jets in the event. It is taken into account that the single jet efficiencies are p_T and η dependent. In Equation (5.4) the fraction of events, where at most one jet is tagged, is subtracted from unity. This corresponds to the efficiency where at least two jets are tagged.

The final correction factor is applied in terms of an event weight shown in Equation (5.6), the corrected two-jet efficiency divided by the uncorrected one.

$$SF_{\geq 2\text{b-tags}} = \frac{\epsilon_{\geq 2\text{b-tags}}(SF_i \cdot \epsilon_{i,sim})}{\epsilon_{\geq 2\text{b-tags}}(\epsilon_{i,sim})} \tag{5.6}$$

The uncorrected single-jet efficiencies $\epsilon_{i,sim}$ are determined on each simulated sample respectively as a function of the jet p_T and η after all top pair selection steps except for the b-tagging requirement.

5.6 Top Pair Event Selection

The previously defined objects are used to select top-pair events in the muon+jets channel. The single selection steps exploit the distinctive signature of this channel to suppress the background processes described in Section 4.3. An event display of such a selected top pair candidate is depicted in Figure 5.9.
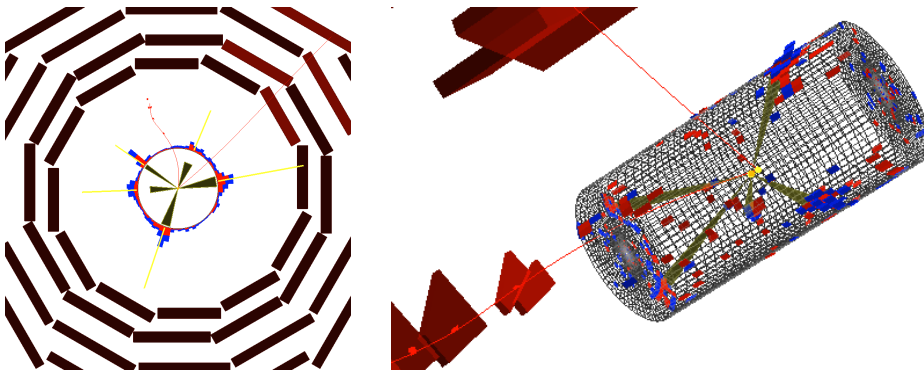
Figure 5.9: Event display of a selected top candidate event in the CMS detector. An isolated muon can be identified as straight global track. Additionally, more than four high-p_T jets are indicated. The soft muon used for the b-charge determination is associated with the most upward jet in the r-ϕ plane shown on the left.

The event selection of this analysis was developed by the CMS top group. It is based on the selection described in [6]. The CMS top group implementation is used. The cut flow consists of the following steps:

1 - Trigger: pass an unprescaled single muon trigger with an isolation criteria and a minimum muon p_T of 24 GeV.

2 - Vertex: require a well identified primary vertex.

3 - Muon: exactly one isolated muon with $p_T > 26$ GeV.

4 - Muon Veto: veto on low energetic loose muon with $p_T > 10$ GeV.

5 - Electron Veto: veto on electrons with $E_T > 15$ GeV.

6 - Jets: at least four well-identified jets with $p_T > 30$ GeV.

7 - Min2BJets: at least two b jets.

A comparison between data and simulation of observables relevant for the measurement is shown in Figure 5.10-5.13. The selected events show a good agreement. The distribution of the jet-charge observable, described in Section 6.1, is not well modeled by simulation but it is not used for the final result of this analysis.

5.7 Jet Association

The aim of this analysis is to combine the charge of the b quark from the leptonic decay side with the charge of the high-p_T muon to reconstruct the charge of the top quark. Thus, it is necessary to associate the two b-jet candidates with the leptonic or hadronic decaying top quark. The jet-association algorithm can be divided into the following steps:

1. Assign the two b jets to the two b quarks in the top-pair event.
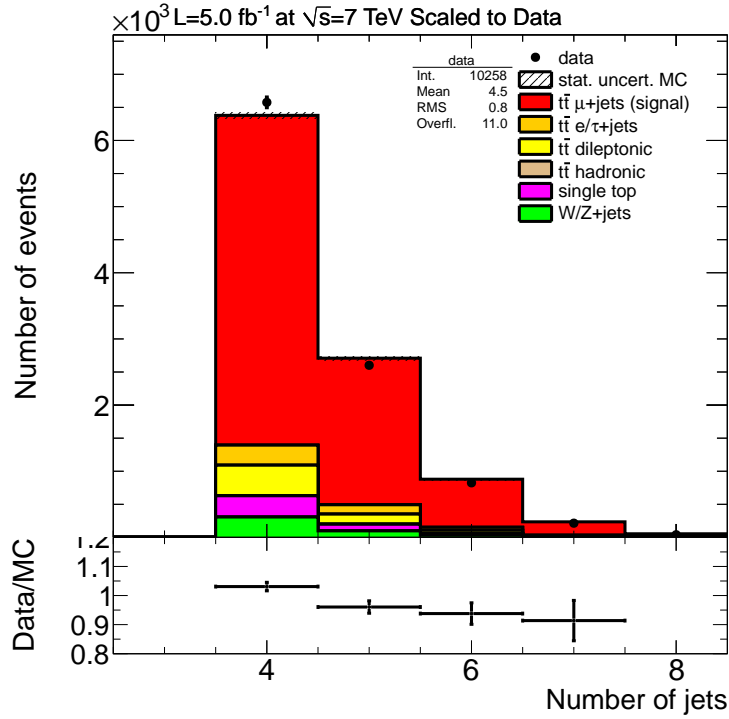
Figure 5.10: Data-MC comparison of the jet multiplicities of the high-quality jets.
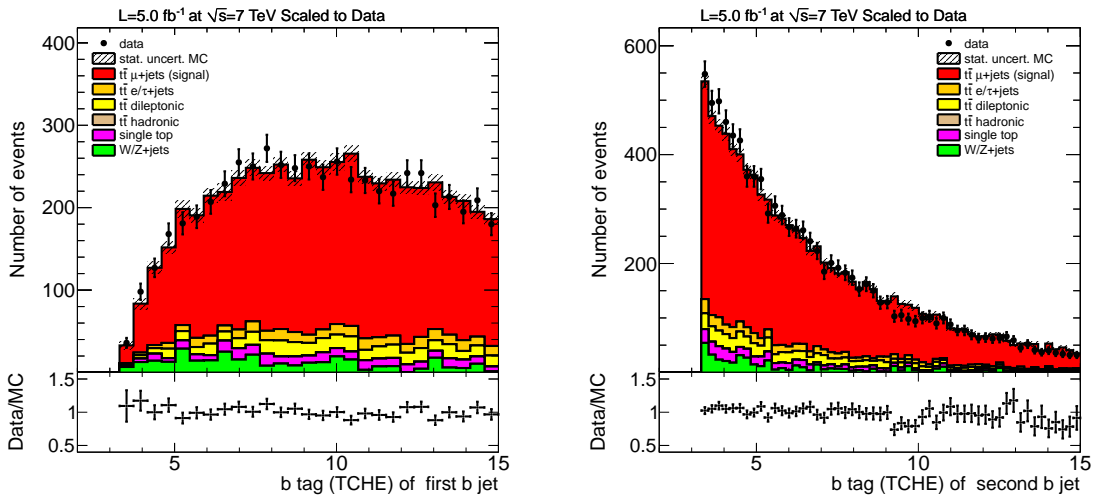

Figure 5.11: Data-MC comparison of the b-tag discriminator values of the first and second b jet.

2. Build with each of the b jets and the remaining high energetic jets all possible three-jet combinations.

3. Find the combination with an invariant mass closest to the top-quark mass of 172.5 GeV for simulated events and 173.5 GeV in data [72].

4. The b jet in this combination is considered to come from the hadronically decaying top quark.

5. The other two non-b jets in this combination are considered to stem from the light quarks of the W boson decay.

6. The remaining b jet is considered to come from the leptonic decay side.

This method has been tested on simulated events by defining the pairing probability P_{pair} to associate both b jets correctly to the b quarks. This requires that both b jets have been successfully matched to a b quark. P_{pair} is shown in Figure 5.14 as a function of the reconstructed top mass. The reconstructed top mass of correct pairings is often closer to the real top mass. Therefore, only events with a reconstructed top mass less than 250 GeV are considered for the top-charge measurement. The invariant mass $m(\mu, b_{\text{lep}})$ of the b jet from the leptonic decay side and the high-p_{T} muon is restricted by the top mass [73]

$$m(\mu, b_{lep}) \leq \sqrt{m_{\text{top}}^2 - m_W^2} \approx 156\,\text{GeV} \, . \tag{5.7}$$

Figure 5.15 shows P_{pair} as function of $m(\mu, b_{\text{lep}})$. There is a veto on events where $m(\mu, b_{\text{lep}})$ exceeds 150 GeV to suppress wrong assignments. Applying these quality criteria a pairing probability of $P_{\text{pair}} = 0.820 \pm 0.003$ (stat.) is reached in the selected top pair events.
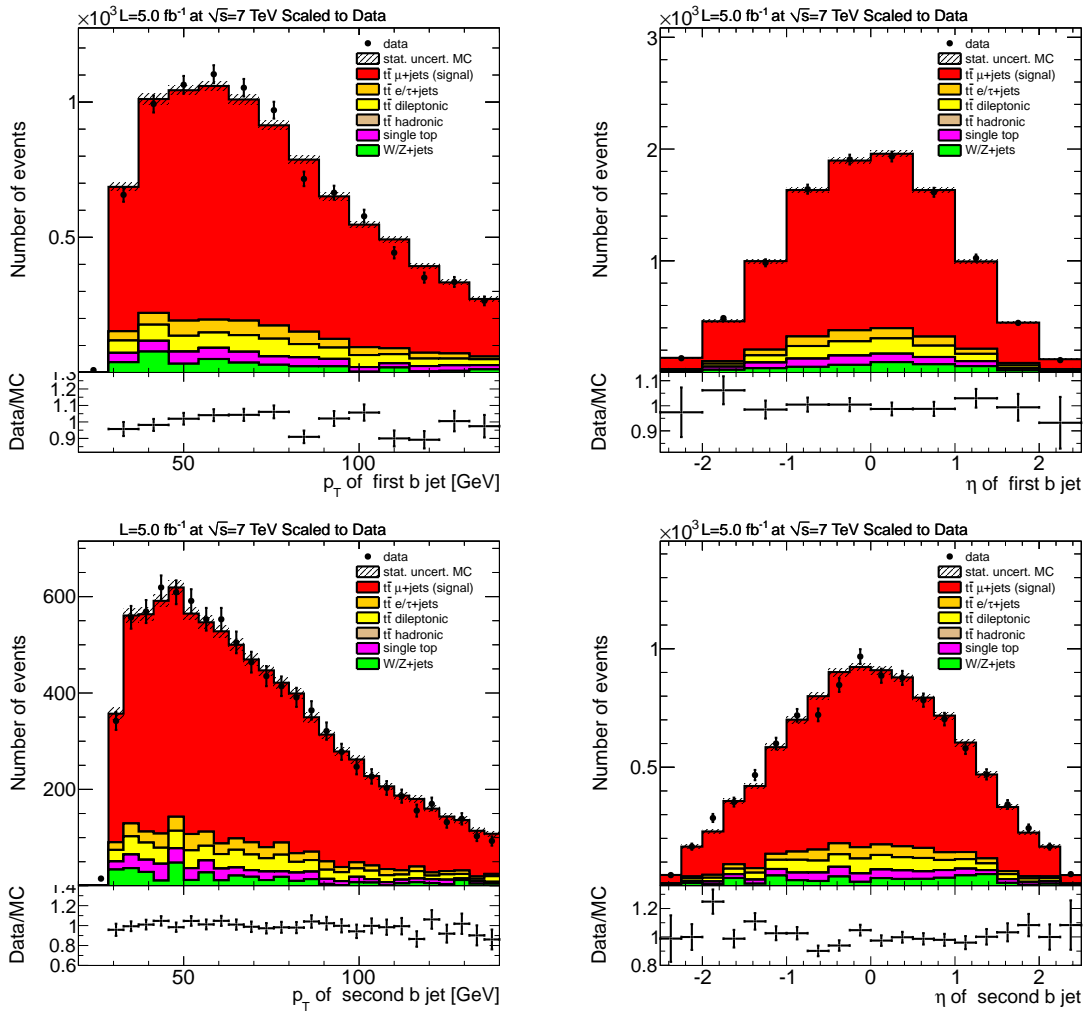
Figure 5.12: Data-MC comparison of the p_T (left) and η (right) distributions of the first and second b jet.
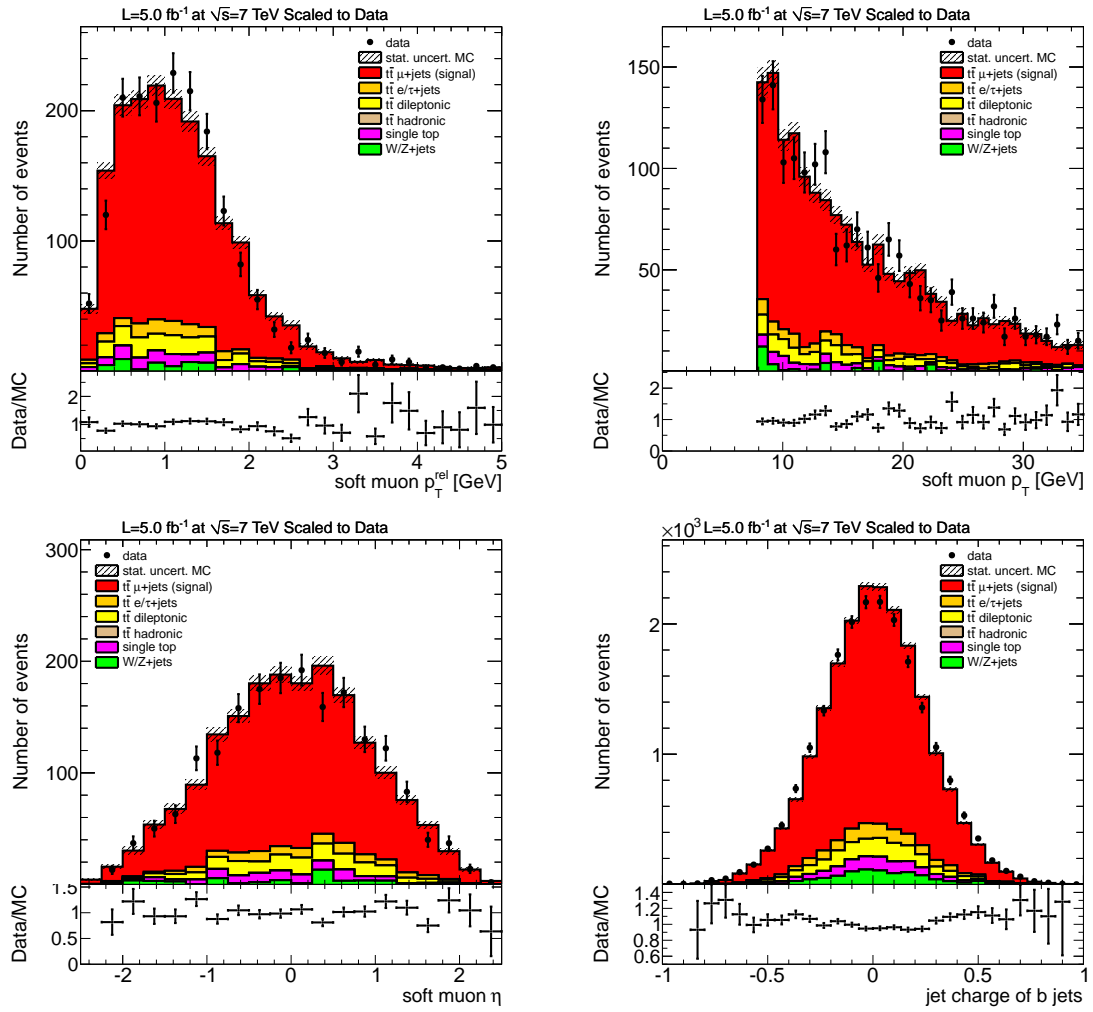
Figure 5.13: Data-MC comparison of variables relevant for the b-charge determination. p_T^{rel}, p_T and η distributions of the soft muons out of the two b jets and the jet-charge distribution of the two b jets.
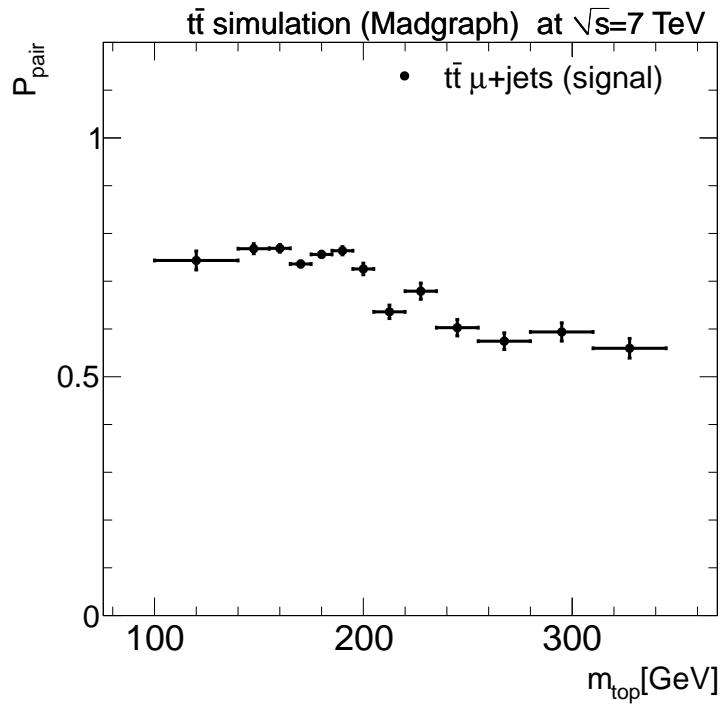
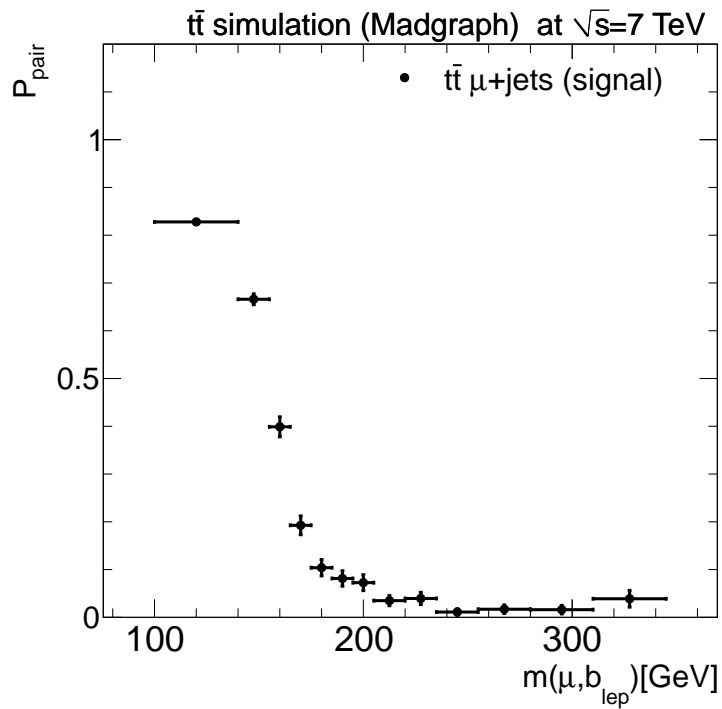**Figure 5.14:** Pairing probability versus the reconstructed top mass on the hadronic decay side.

Figure 5.15: Pairing probability versus the reconstructed invariant mass between the b jet associated to the leptonic decay side and the high-p_T muon.

Chapter 6

b-Charge Determination

Due to confinement it is not possible to measure the charge of the bare b quark directly. Particles arising in this hadronization and decay process form the final b jet that is measured by the detector.

These b jets have characteristics that depend on whether the initial parton is a b quark or a $\bar{b}$ quark. Thus, there is a correlation between b-jet observables and the sign of the b quark's charge. These observables have already been used at LEP [74] and before.

The most sensitive observables for this analysis are investigated in more detail: the so-called jet charge JC and the charge q_μ of a muon from B-meson decays. In the following the general behavior of the two observables is studied and their performance is tested in terms of the probability P_{bC} for correct b charge determination. The discriminators are optimized in terms of statistical significance as described in Section 6.3. Finally, the discriminator with the overall best performance is chosen for this analysis.

6.1 Jet-Charge Discriminator

During the hadronization process, particles are produced that are charge-correlated with the b quark. This is sketched in Figure 6.1.

There are experimentally confirmed fragmentation models that sort the produced particles in a list. The hadron containing the initial quark takes the highest rank in the list. The correlation to the original quark charge is stronger for particles being higher in the list. Those particles are more likely to be produced with large momentum and rapidity with respect to the initial quark axis [74].

Therefore, a jet-charge variable (JC) has been defined that weights the charges q of the tracks with their momentum p:

$$ JC = \frac{\sum\limits_{\text{tracks}} (p \cdot \vec{j})^x \cdot q}{\sum\limits_{\text{tracks}} (p \cdot \vec{j})^x} \ . \tag{6.1} $$

The sum refers to all tracks associated to the b jet. The term $(p \cdot \vec{j})$ corresponds to the weight. In the example given here it is the track momentum p projected on the jet axis $\vec{j}$. The parameter x defines the strength of the momentum weights and is used to optimize the discrimination performance of JC.
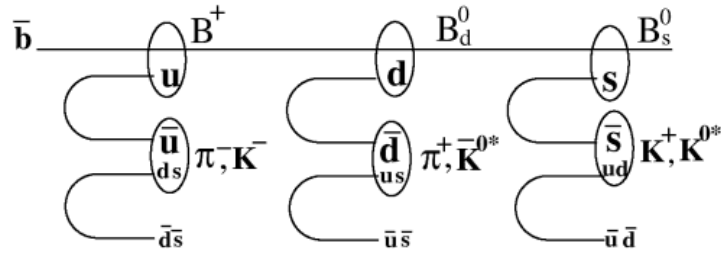
Figure 6.1: "Schematic picture of $\bar{b}$ fragmentation illustrating the charge correlation of the parent $\bar{b}$ quark to the various leading fragmentation particles that can appear." [74]

This analysis interprets *JC* in terms of a discriminant, that allows to distinguish between b and $\bar{b}$ jets. If *JC* is positive the b-quark charge is considered to be $+1/3$ e, if *JC* is negative the b-quark charge is taken to be $-1/3$ e. In Figure 6.2 the jet charge is shown for jets matched to b and $\bar{b}$ quarks with the jet charge configuration leading to the best discriminating performance as explained in the following section.
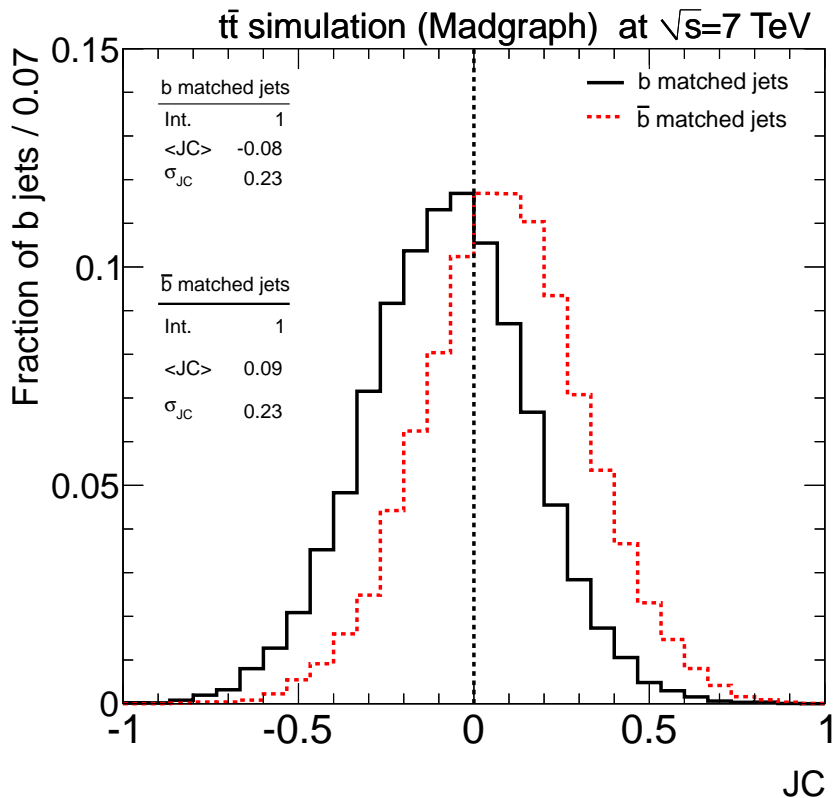
Figure 6.2: *JC* of b- and $\bar{b}$-matched jets normalized to unit area.

6.1.1 Performance

Different configurations of the jet-charge variable have been tested and are optimized for the highest significance ($< JC > /\sigma_{\mathrm{JC}}$) of the jet charge :

Jet-tracks associations: The exact definition of tracks to be considered for the calculation of *JC* is essential. As a first option all tracks are associates to the jet which lie within $\Delta R < 0.5$. This option is referred to as cone-based *JC* in the following. The second option associates all tracks to the jet, that belong to the charged particles forming the particle flow jet, referred to as particle-based *JC*. Figure 6.3 shows the distribution of the track weights for all tracks that are considered in the jet-charge variable for the cone and particle based *JC* definition. The track weight corresponds to a probability that the track originates from a specific vertex. In this case, the weights for the primary vertex are shown. The number of tracks with a zero weight decreases for the particle based *JC*. This is caused by the pileup correction step during the particle flow algorithm. Therefore, the particle-based option is expected to be less sensitive to pileup effects.

Type of weight: The momentum transverse ($p_{\mathrm{T}}^{\mathrm{rel}}$) and longitudinal ($p_{\mathrm{L}}^{\mathrm{rel}}$) to the jet axis and the total momentum can serve as weight for the track charges. The transverse momentum is motivated by the high mass of the B meson. The longitudinal momentum is motivated by the fact that the b quark is produced in the hard process. The hadron containing the b quark has a higher momentum than subsequent particles in the decay chain. The total momentum hardly differs from the longitudinal momentum. Therefore, the plots are only shown for the longitudinal momentum.

Strength of weight: The exponent x in Equation (6.1) defines the strength of the weight.

Figure 6.4 shows that the following configuration is optimal:

- Jet-tracks association: particle based *JC* definition

- Type of weight: $p_{\mathrm{L}}^{\mathrm{rel}}$

- Strength of weight: $x = 0.7$

Additional track requirements that were applied in previous analyses [75] are tested for an improvement in the b-charge reconstruction. Figure 6.5 and Figure 6.6 show P_{bC} as a function of a requirement on these observables. Figure 6.5 shows that P_{bC} can not be enhanced by requiring the tracks to have a minimum p_{T}. A restriction on tracks from certain quality categories can not improve P_{bC} neither as shown in Figure 6.6. Requiring that the track must have a maximum impact parameter d_0 has been tested in Figure 6.5. Since non of the mentioned requirements significantly improve P_{bC} but decrease the charge reconstruction efficiency, none of these requirements are placed on the tracks. The jet charge distribution of the final configuration is shown in Figure 6.2.
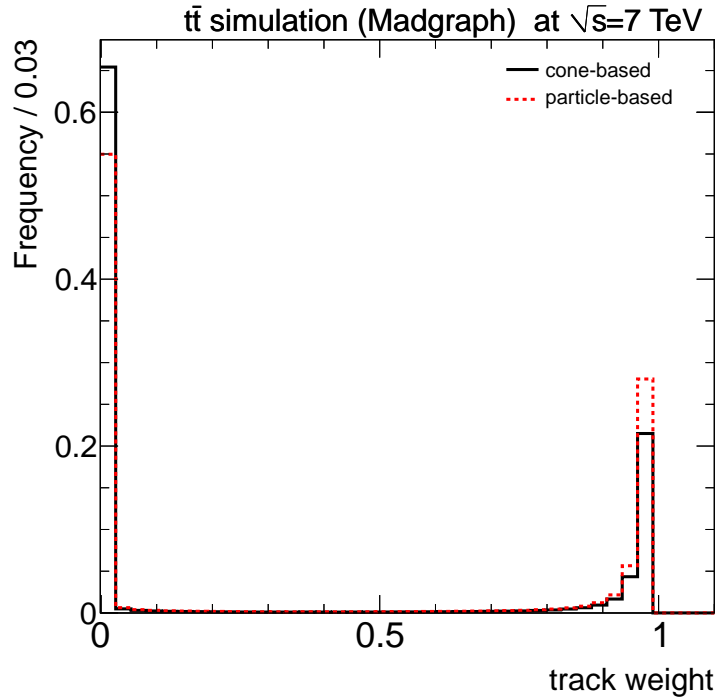
Figure 6.3: Primary vertex track weight of all tracks that are considered for the jet charge for cone and particle based jet-track association.
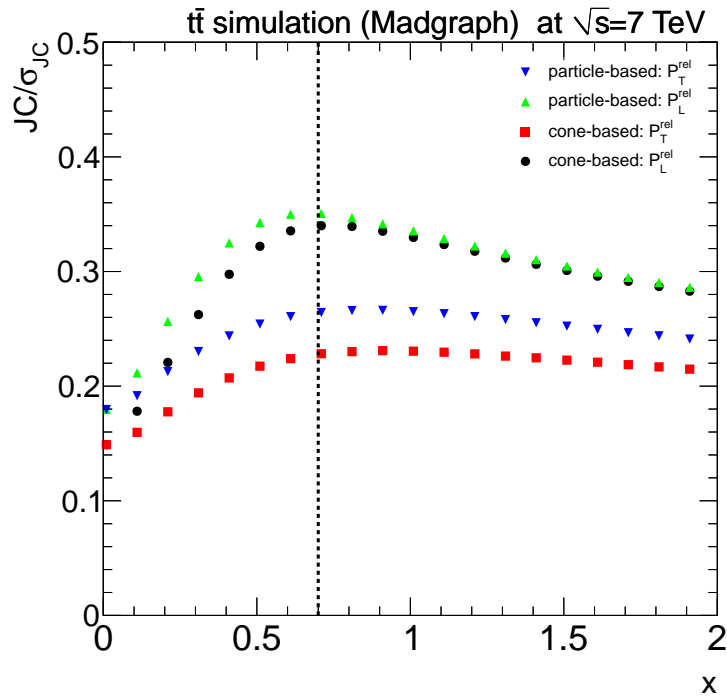

Figure 6.4: Jet charge significance for the two selected b jets in the events with different jet-track association methods and different momentum weights. The dashed line corresponds to the chosen value for x.
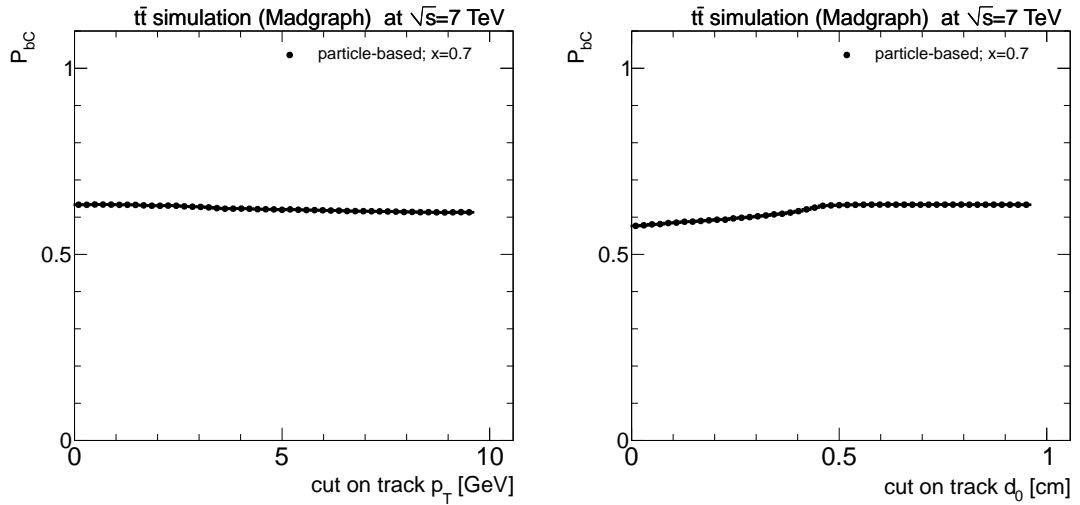
Figure 6.5: P_{bC} as a function of the minimum transverse momentum (left) and maximum impact parameter (right) requirement of the tracks considered for the jet-charge observable.
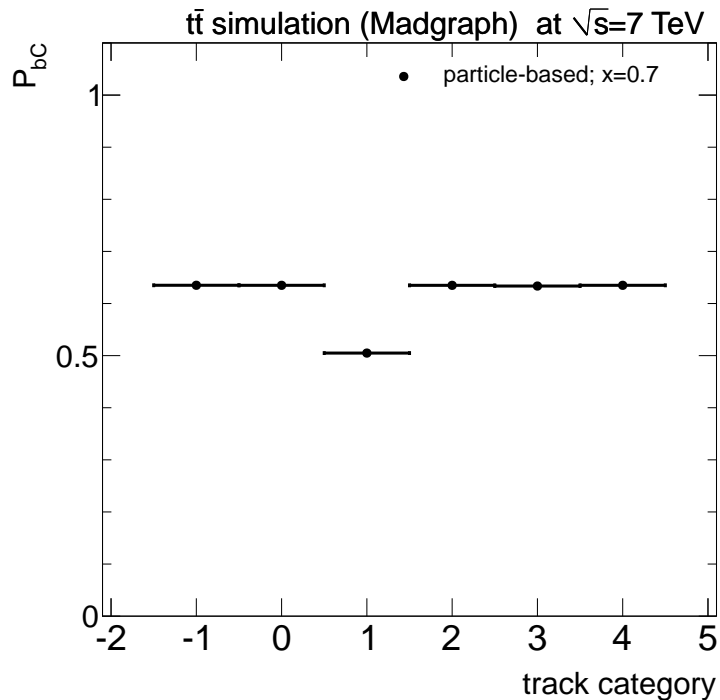
Figure 6.6: P_{bC} as a function of the track quality categories. The quality categories are the following: undefined quality=-1, loose=0, tight=1, high purity=2, confirmed=3, good iterative=4. The categories imply different requirements on how many layers contribute to the track, the quality of the track fit and the compatibility with the primary vertex. The restriction on tracks of these categories does not increase P_{bC}.
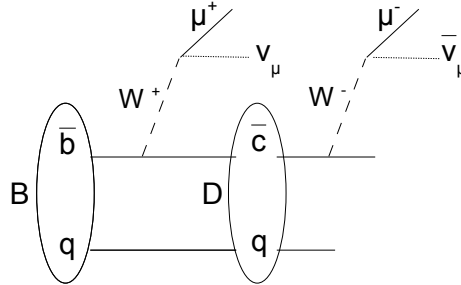
Figure 6.7: Sketch of a B-meson decay chain with leptons that can occur in a direct decay or cascade decay via a D meson.

6.2 Soft Muon Charge Discriminator

Figure 6.7 shows the double-leptonic decay of a B meson in a cascade via a D meson. Muons arising from the direct decay carry a charge q_μ with the same sign as the initial b-quark charge which qualifies q_μ as b-charge discriminator. The branching ratio of the direct decay is about 11 % [5]. If the b-quark decays sequentially via a D-meson the sign will be opposite. The branching ratio of such a cascade decay is about 10 % [5]. If the B meson is neutral and oscillates before it decays the sign of the muon charge will be different as well.

Figure 6.8 shows an example decay chain of the signal sample that uses PYTHIA [57] with the Lund-String model for the simulation of the hadronization process.

Obviously, it is not possible to connect the b quark from the hard process with the B meson after the hadronization directly due to the string formation. There are also b quarks that arise in gluon radiations. All sources of muons are classified in Table 6.1 by the quark content of the hadron that they are originating from. The gluon radiations off the b quark modeled in the signal sample are mostly soft and collinear. They contribute to the b jet from the hard process. Therefore, the reference b quark in the following is the b quark that stems directly from the top quark before having radiated any gluons.

After reconstruction, those muons can be identified as non-isolated muons inside a jet cone of $\Delta R < 0.4$. They must be global muons and have a $p_T > 8\,\text{GeV}$ to guarantee a good reconstruction and to be comparable to the muons from the data driven validation in Section 6.4. The muons within b jets are denoted as soft muons in the following though the soft muons have relatively high p_T (see Figure 5.13). This notation is motivated by the fact that the p_T of the soft muons is generally lower than the p_T of the isolated muons that arise in the hard process, the so-called high-p_T muons.

6.2.1 Performance

The muons stemming from direct B-meson decays often have a large momentum. On the one hand this is due to the hard fragmentation characteristics of b quarks and the low

```
p+ (0, 0, 3500; 3500)
  g (0.635035, 1.71445, 724.632; 724.634)
    g (43.8447, 2.54018, 473.557; 475.589)
      t (-107.767, -113.374, -19.3436; 233.662)
        b (-36.4298, -69.2289, 55.018; 95.754)
          g (-1.21124, -1.24077, 0.344723; 1.76789)
            string (-25.5002, -78.4112, -6.71851; 161.442)
              pi- (0.206244, 0.314617, -0.585789; 0.710034)
              rho+ (0.667644, 0.0228945, -8.33434; 8.3932)
              rho- (1.89499, -0.110095, -11.5147; 11.7064)
              Kbar0 (0.562772, -0.165278, -3.91794; 3.99274)
              K*0 (0.70154, -0.0509931, -9.80805; 9.87165)
              rho0 (0.53656, -0.833033, -4.83313; 4.98257)
              pi0 (0.0942151, -0.144297, -0.234083; 0.320488)
              pi+ (0.45937, -0.761411, -4.30547; 4.39856)
              eta' (2.15288, -2.18408, -9.73833; 10.2546)
              rho- (0.440562, -0.945659, -2.64518; 2.95005)
              rho+ (0.469156, -1.19229, -2.90203; 3.24858)
              rho- (1.34579, -1.09739, -0.832106; 2.00006)
              rho0 (-0.609179, -1.04281, -0.377163; 1.46952)
              pi+ (0.0725278, 0.0179702, -0.0870763; 0.18068)
              rho- (0.115973, -1.51625, -0.247698; 1.74407)
              nbar0 (0.0628145, -2.1429, 0.833638; 2.4847)
              n0 (-0.861006, -0.603487, 0.463284; 1.48423)
              Kbar0 (-1.29575, -4.86847, 3.44693; 6.12454)
              K*+ (-5.20192, -8.88949, 8.01362; 13.0799)
              rho- (-2.18244, -4.61937, 2.84182; 5.91516)
              B*bar0 (-25.133, -47.5993, 38.0453; 66.1299)
                Bbar0 (-24.7043, -46.8033, 37.4206; 65.0309)
                  nu_mu (-2.69458, -5.76442, 4.35346; 7.70986)
                  mu+ (-6.81927, -16.1591, 13.2876; 22.0043)
                  D*_2- (-15.1904, -24.8798, 19.7795; 35.3167)
                gamma (-0.428729, -0.796083, 0.624676; 1.09899)
          g (0.139759, -1.11166, 0.642165; 1.29139)
          b (-34.9753, -66.623, 53.2908; 92.3249)
            string (-25.5002, -78.4112, -6.71851; 161.442)
              pi- (0.206244, 0.314617, -0.585789; 0.710034)
              rho+ (0.667644, 0.0228945, -8.33434; 8.3932)
              rho- (1.89499, -0.110095, -11.5147; 11.7064)
              Kbar0 (0.562772, -0.165278, -3.91794; 3.99274)
              K*0 (0.70154, -0.0509931, -9.80805; 9.87165)
              rho0 (0.53656, -0.833033, -4.83313; 4.98257)
              pi0 (0.0942151, -0.144297, -0.234083; 0.320488)
              pi+ (0.45937, -0.761411, -4.30547; 4.39856)
              eta' (2.15288, -2.18408, -9.73833; 10.2546)
              rho- (0.440562, -0.945659, -2.64518; 2.95005)
              rho+ (0.469156, -1.19229, -2.90203; 3.24858)
              rho- (1.34579, -1.09739, -0.832106; 2.00006)
              rho0 (-0.609179, -1.04281, -0.377163; 1.46952)
              pi+ (0.0725278, 0.0179702, -0.0870763; 0.18068)
              rho- (0.115973, -1.51625, -0.247698; 1.74407)
              nbar0 (0.0628145, -2.1429, 0.833638; 2.4847)
              n0 (-0.861006, -0.603487, 0.463284; 1.48423)
              Kbar0 (-1.29575, -4.86847, 3.44693; 6.12454)
              K*+ (-5.20192, -8.88949, 8.01362; 13.0799)
              rho- (-2.18244, -4.61937, 2.84182; 5.91516)
              B*bar0 (-25.133, -47.5993, 38.0453; 66.1299)
                Bbar0 (-24.7043, -46.8033, 37.4206; 65.0309)
                  nu_mu (-2.69458, -5.76442, 4.35346; 7.70986)
                  mu+ (-6.81927, -16.1591, 13.2876; 22.0043)
                  D*_2- (-15.1904, -24.8798, 19.7795; 35.3167)
                gamma (-0.428729, -0.796083, 0.624676; 1.09899)
        W+ (-71.3373, -44.1455, -74.3617; 137.908)
```

Figure 6.8: Excerpt of an example decay chain of a top pair signal event with semimuonically decaying b quarks simulated with PYTHIA [57] hadronization and showering. Obviously, sometimes more than one b quark arises from one decaying top e.g. in a gluon splitting process. The four momenta for each decay particle is shown in brackets.

Table 6.1: Sources of muons in b decays. The points indicate that there can be gluon radiations.

Denotation	Decay
Direct	$b \rightarrow ... \rightarrow$ *string* $\rightarrow b \rightarrow \mu^-$
Cascade	$b \rightarrow ... \rightarrow$ *string* $\rightarrow c \rightarrow \mu^+$
W Cascade	$b \rightarrow ... \rightarrow$ *string* $\rightarrow \overline{c} \rightarrow \mu^-$
Anti	$b \rightarrow ... \rightarrow$ *string* $\rightarrow \overline{b} \rightarrow \mu^+$
Tau	$b \rightarrow ... \rightarrow$ *string* $\rightarrow \tau^- \rightarrow \mu^-$

decay-particle multiplicity in semileptonic b-quark decays. On the other hand it is due to the relatively high mass of the B meson which leads to a large transverse component of the muon momentum with respect to the b quark axis, in the following denoted as p_T^{rel}. The reference b quark is the b quark originating directly from the top quark before any gluon radiation. If the muon stems from a D meson its p_T^{rel} is smaller due to the smaller mass of the D meson. Figure 6.9 shows the muon p_T^{rel} spectra on generator level for the different semimuonic b decay chains introduced in Table 6.1. The relative frequencies of the decay chains represent the corresponding branching fraction realized in simulation. Figure 6.10 shows the longitudinal momentum component with respect to the b quark axis, denoted as p_L^{rel}.
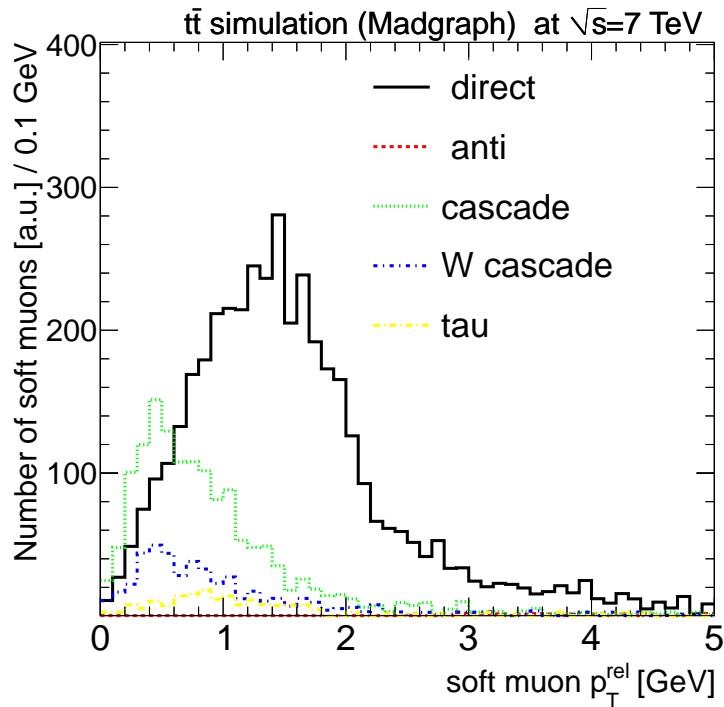
Figure 6.9: p_T^{rel} of the generated soft muons with $p_T > 8$ GeV stemming from the different semimuonic b decay chains listed in Table 6.1.
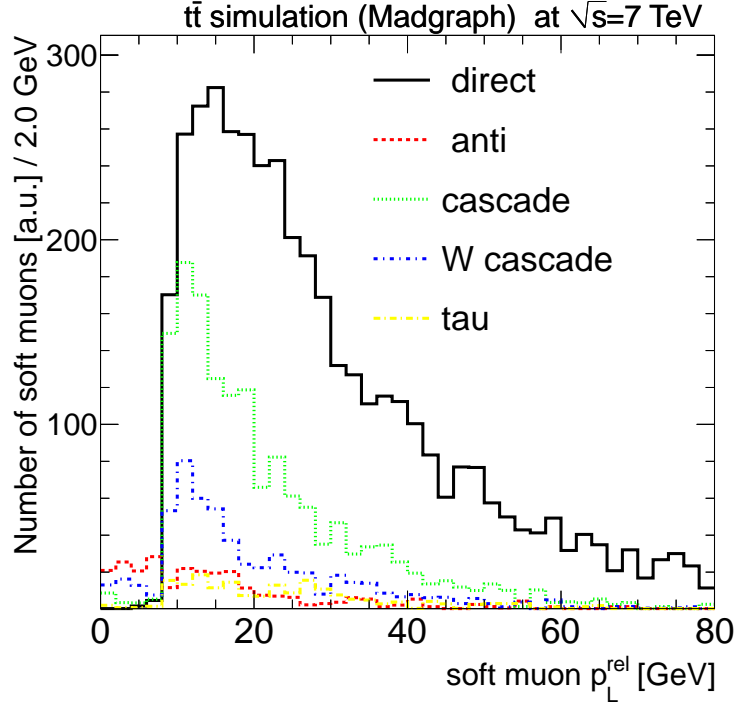
Figure 6.10: p_L^{rel} of the generated soft muons with $p_T > 8$ GeV stemming from the different semimuonic b decay chains listed in Table 6.1.

Figure 6.11 and Figure 6.12 show that P_{bC} increases with rising p_T^{rel} and p_L^{rel}. Since the mass of the B meson restricts the p_T^{rel} of the soft muon there is a tendency that P_{bC} decreases again for p_T^{rel} values larger than 2 GeV. The results on generator and reconstruction level agree.

6.3 Choice and Optimization of the b-Charge Discriminator

Both the *JC* and the soft muon discriminators show a correlation to the initial b quark's charge. In the following their discrimination power is optimized and it is investigated if a combination of the two discriminators improves the result.

6.3.1 Considered b-charge discriminators

$q_\mu(p_T^{rel})$: As shown in Figure 6.11 P_{bC} increases with increasing p_T^{rel}. An optimal requirement of a minimum p_T^{rel} value for the soft muons is investigated.

$q_\mu(p_L^{rel})$: As shown in Figure 6.12 P_{bC} increases with increasing p_L^{rel}. An optimal requirement of a minimum p_L^{rel} value for the soft muons is investigated.

JC: The plain jet charge and a possible cut value on its absolute value is investigated. Although the charge correlation is weaker for this discriminator a lot more events remain for the measurement without the requirement of a muon inside the b jet.
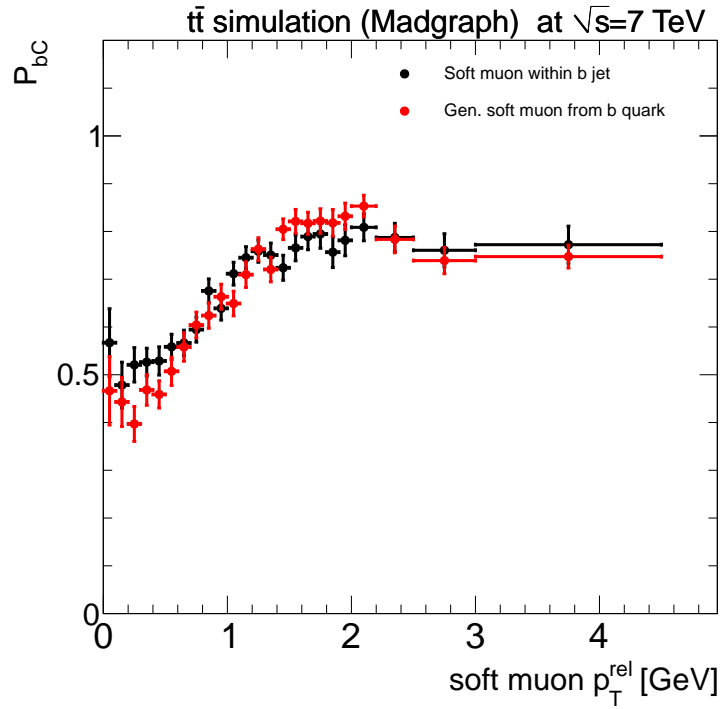
Figure 6.11: P_{bC} as a function of muon p_T^{rel} for generator muons from the decay of the reference b quarks and on reconstruction level for muons within b jets.
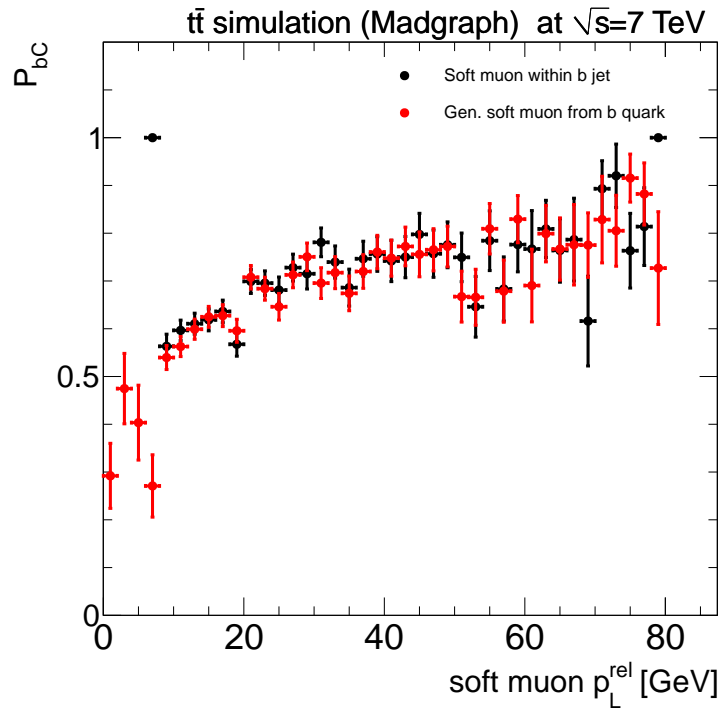

Figure 6.12: P_{bC} as a function of muon p_L^{rel} for generator muons from the decay of the reference b quarks and on reconstruction level for muons within b jets.

***JC* x *q*$_\mu$:** In contrast to the soft muon discriminator which appears in the decay of the B meson the jet charge comprises additional information of the flavor at production of the B meson since tracks arising from the fragmentation process contribute to it. Therefore, the jet-charge observable contains independent information. Figure 6.13 shows the jet charge distribution for b and $\bar{b}$ jets with either a positive or negative soft muon inside. There is still separation power to discriminate the b and $\bar{b}$ jets by applying a cut on JC. Therefore, an additional requirement on $JC \times q_\mu$ is investigated.

Two versions of this combination are studied, including the soft muon in the jet-associated tracks ($JC_{\text{incl},\mu}$) and excluding it ($JC_{\text{excl},\mu}$). Figure 6.13 shows the $JC_{\text{incl},\mu}$ distribution for b and $\bar{b}$ jets including the soft muon in the jet associated tracks. Figure 6.14 shows the $JC_{\text{excl},\mu}$ the corresponding distribution with the soft muon removed from the jet associated tracks. It can be seen that the distributions are shifted with respect to Figure 6.13 due to the removed soft muon but the separation power remains.
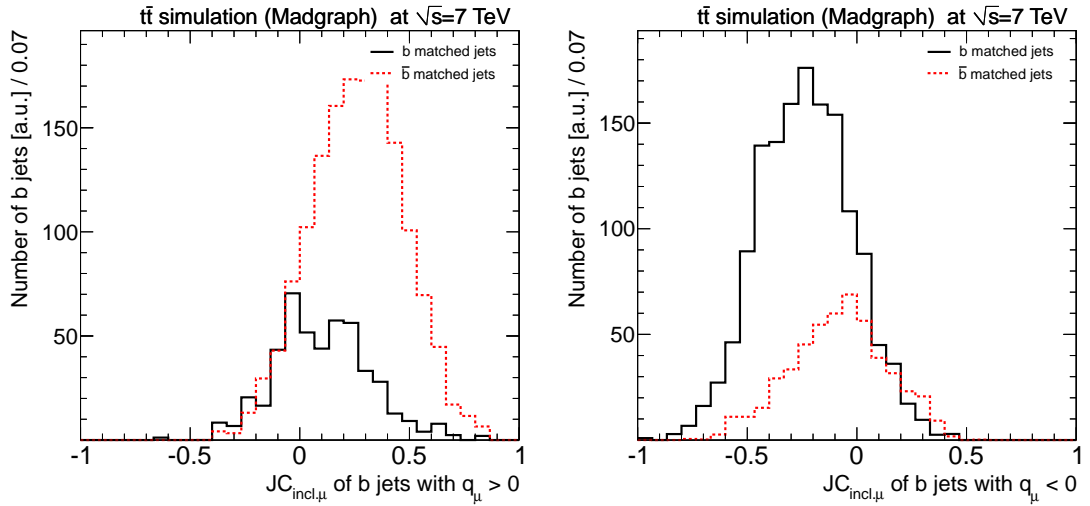

Figure 6.13: *JC* of b- and $\bar{b}$-matched jets including the soft muon from the associated tracks for jets containing a positive (left) and a negative (right) soft muon.

6.3.2 Optimization procedure

First the discrimination power is optimized in terms of statistical significance. Then the impact of systematic uncertainties is investigated. Finally, the discriminator with the best overall performance is chosen for this analysis.

The discriminators are optimized for maximal statistical significance of the final test statistics. The test statistics A of the analysis is the asymmetry between the number of events N_{SM} with a standard model like reconstructed top charge and the number of events N_{XM} with an exotic reconstructed top charge

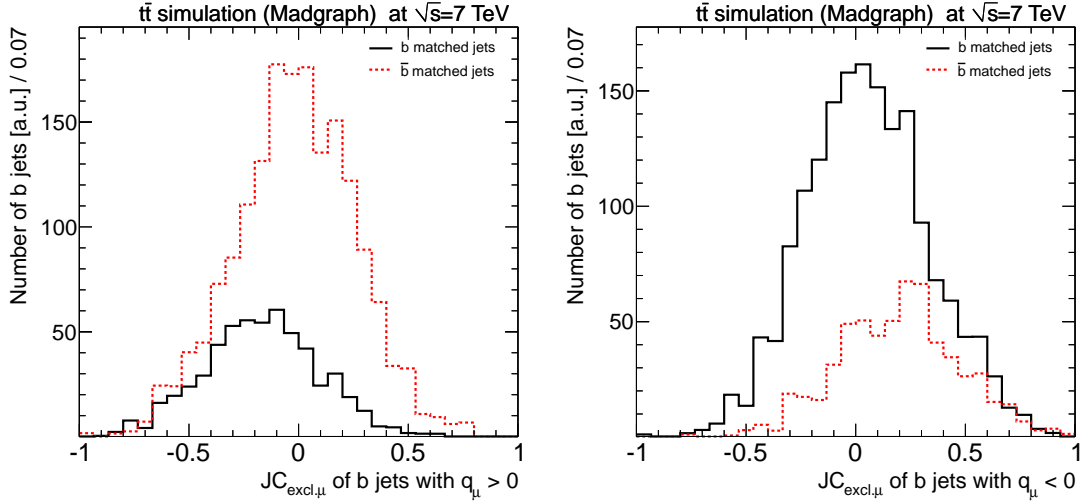$$A = \frac{N_{\text{SM}} - N_{\text{XM}}}{N_{\text{SM}} + N_{\text{XM}}} = 2P - 1 \ . \tag{6.2}$$

Figure 6.14: *JC* of b- and $\bar{b}$-matched jets excluding the soft muon from the associated tracks for jets containing a positive (left) and a negative (right) soft muon.

P is the fraction of SM like events.

Assuming that both scenarios could be realized in nature, the true asymmetry between them is diluted by imperfect reconstruction. The true and measured asymmetries are related

$$A_{\text{true}} = \frac{A_{\text{meas}}}{D} \tag{6.3}$$

via the dilution

$$D = \frac{N_{\text{correct}} - N_{\text{wrong}}}{N_{\text{correct}} + N_{\text{wrong}}} \ . \tag{6.4}$$

Assuming the standard model (SM) to be true, a correctly reconstructed event is reconstructed with a top charge of $+2/3\,$e. In the alternative exotic scenario (XM) a correctly reconstructed event would lead to a top charge of $-4/3\,$e. Therefore, $A(\text{SM}) = -A(\text{XM})$ applies since N_{correct} and N_{wrong} are inverted for the two scenarios.

The difference d of the asymmetries for the two hypothesis in units of standard deviations corresponds to the statistical significance of the result. Therefore, the optimal parameter values maximize the figure of merit d.

$$d = \frac{|A(\text{SM}) - A(\text{XM})|}{\sigma_A} = \frac{2A}{\sigma_A} \ . \tag{6.5}$$

The statistical uncertainty on the asymmetry σ_A is calculated using binomial statistics, a common approximation [76]

$$\sigma_A = \sqrt{\frac{1 - A^2}{N_{\text{sel.}}}} \ . \tag{6.6}$$

$N_{\text{sel.}}$ is the number of selected events for the measurement.

In addition to d, an alternative figure of merit used in an analysis of the CDF experiment is tested [75]. The aim of the CDF approach is to minimize a quantity that scales with the statistical uncertainty on the true asymmetry A_{true} between $-4/3\,\text{e}$ and $+2/3\,\text{e}$ charged top quarks in nature.

The error on the true asymmetry is (see Equation 6.6)

$$\sigma_{A_{\text{true}}} = \sqrt{\frac{1 - A_{\text{meas}}^2 \cdot D^2}{D^2 \cdot \epsilon \cdot N}} \ . \tag{6.7}$$

N is the number of events before the b-charge specific selection and ϵ is the efficiency of the b-charge specific selection. Within the CDF approach the quantity ϵD^2 is maximized.

6.3.3 Results

The figure of merit d is shown for the different discriminators described before as function of the variable to be optimized:

- Figure 6.15 shows d as a function of the absolute cut value on the *JC*. The jet charge with an additional cut of $|JC| > 0.2$ optimizes the performance.

- Figure 6.16 shows d as function of the cut value on $JC \times q_\mu$. The combination of *JC* and the soft muon charge is better for the jet charge definition which includes the soft muon track in the collection of associated tracks. The combined discriminator performance is optimal for cut values in the range of $[0., 0.3]$. In the following both versions of this discriminator must fulfill $JC \times q_\mu > 0$.

- Figure 6.17 shows d as function of a minimum $p_{\text{T}}^{\text{rel}}$ requirement. d is maximal for the soft muon discriminator at $p_{\text{T}}^{\text{rel}} > 0.85\,\text{GeV}$. The highest significance for the combined discriminator is reached for the jet charge definition that includes the soft muon in the associated tracks for $p_{\text{T}}^{\text{rel}}$ cut values smaller than $p_{\text{T}}^{\text{rel}} = 1\,\text{GeV}$.

- With Figure 6.18 it has been tested which requirement on $p_{\text{L}}^{\text{rel}}$ is optimal. The combined discriminator with $JC_{\text{incl.}\mu}$ shows the best performance with no requirement on the $p_{\text{L}}^{\text{rel}}$ of the soft muon.

Comparing all parameter variations the plain jet charge with an additional cut on the absolute value of $JC > 0.2$ shows the best performance in terms of statistical significance. Figure 6.19 indicate the outcome of applying the CDF approach in this analysis. Both approaches lead to consistent results. The maxima of ϵA^2 and d are located at the same cut parameter values.

Taking all systematic uncertainties into account the measurement is not limited by statistics. The significance can only be improved by increasing A and therewith P, see Equation 6.2, to reduce the impact of systematic uncertainties. The optimization is performed on signal events only, therefore P is labeled P_{signal}. The resulting value of P_{signal} needs to be taken into account in the decision for the optimal parameters for the b-charge discrimination. Figure 6.20 to 6.23 show P_{signal} as function of the variable to be optimized.

A quality requirement on the soft muon to have a $p_{\text{T}}^{\text{rel}} > 0.85\,\text{GeV}$ provides an optimal trade-off between statistical significance and low sensitivity to systematic uncertainties.
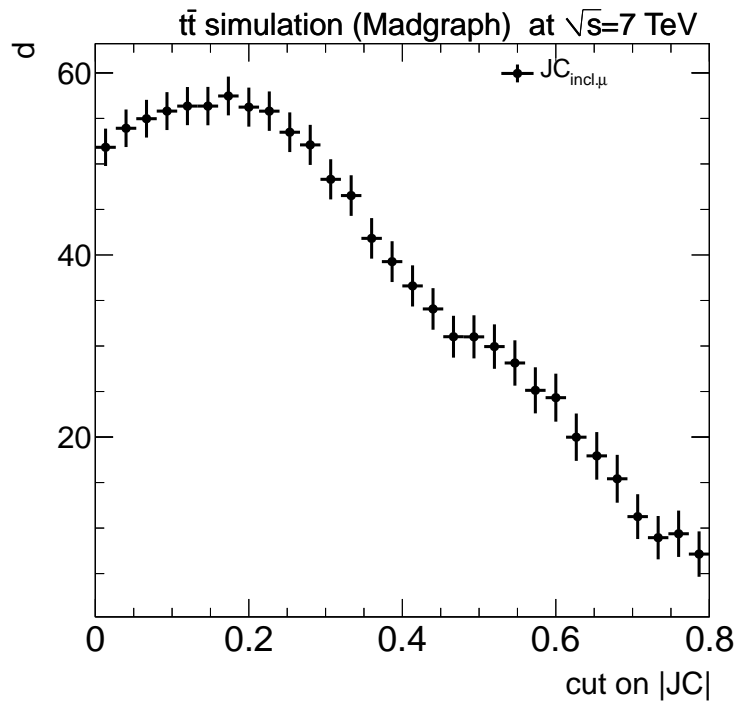
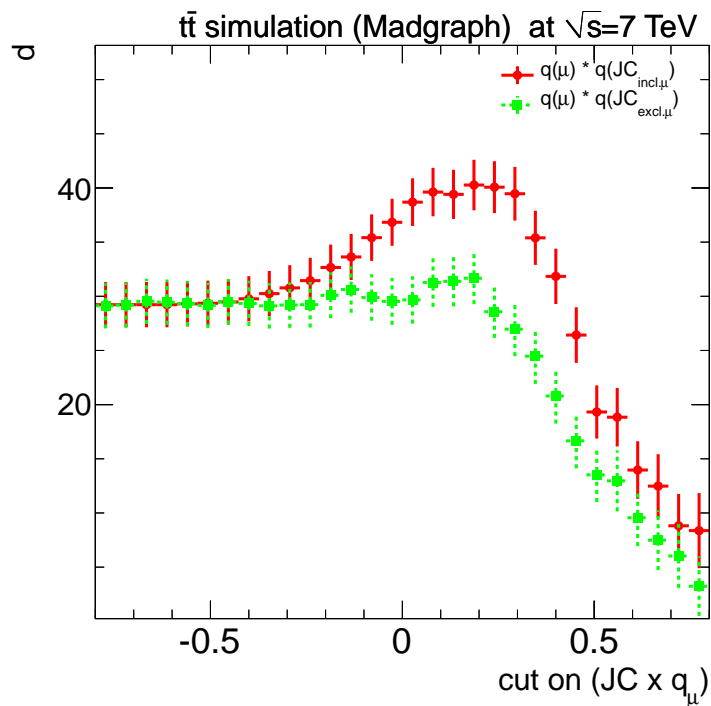**Figure 6.15:** d as a function of the minimal absolute jet charge cut value.

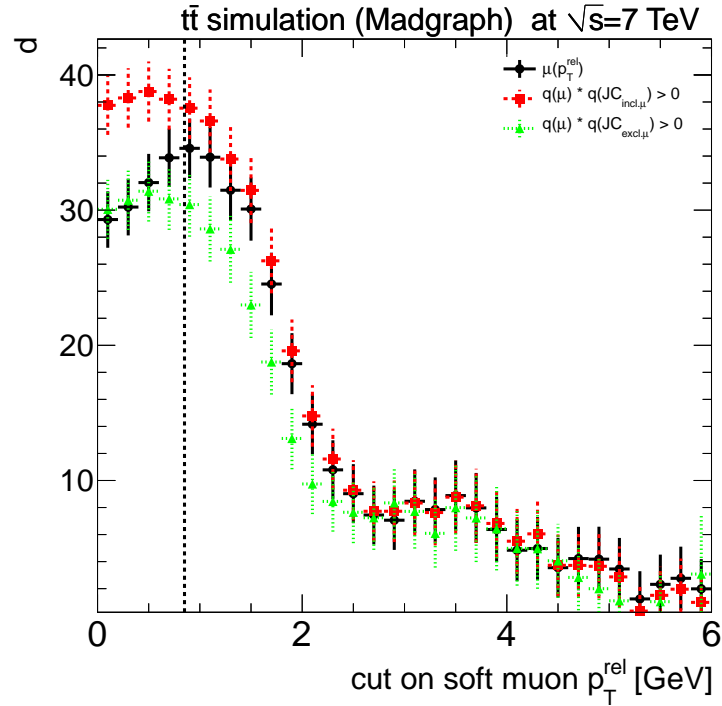**Figure 6.16:** d as a function of the minimal cut value on $JC \times q_\mu$.

Figure 6.17: d as a function of the cut on p_T^{rel} of the soft muon. Three different discriminators described before are investigated. The highest significance for the soft muon discriminator is reached by a minimum cut value of $p_T^{rel} = 0.85\,\text{GeV}$ indicated by a dashed line.
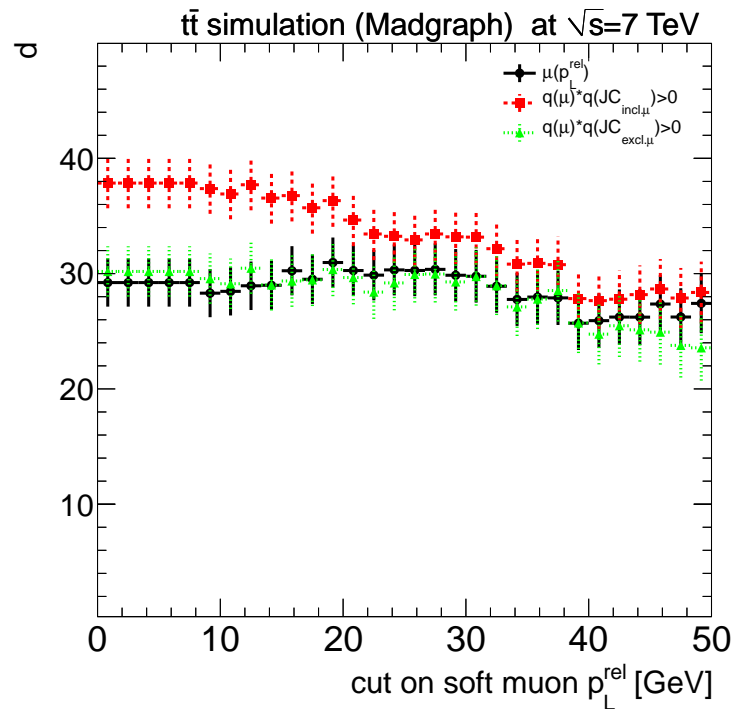

Figure 6.18: d as a function of the soft muon p_L^{rel} cut. Three different discriminators described in the text are investigated.
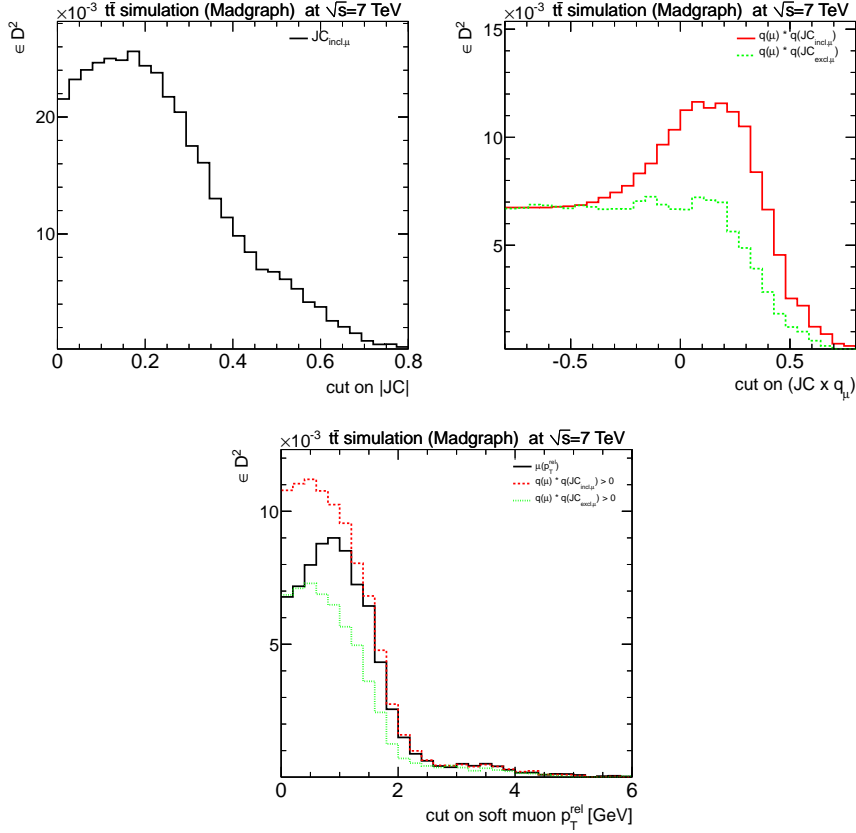
Figure 6.19: ϵA^2 as a function of the minimal absolute jet charge cut value (upper left) as a function of a cut on $JC \times q_\mu$ (upper right) and as a function of the minimum p_T^{rel} of the soft muon (below). Three different discriminators described in the text are investigated.

The combined discriminator is more sensitive to systematic uncertainties and has not a significantly higher P_{signal}. Thus, the soft muon is chosen as b discriminator for the measurement. Events where at least one of the two b jets fulfills these requirements are selected for the measurement. The resulting P_{bC} is $P_{bC} = 0.746 \pm 0.007$ (stat.) .

6.4   Data Driven Performance Validation

Being produced at energy scales at which the strong coupling is small and pertubative theory is applicable the b quark radiates gluons before the fragmentation starts due to confinement. Through the self interaction of gluons the color field density grows until new quark pairs fluctuate from the vacuum. The b quark loses energy while it separates from its color partner to scales where the strong coupling is large [74].
Thus, the fragmentation of b quarks is based on low-energetic QCD models as the Lund-string model [77] applied in this analysis. Those models suffer from large uncertainties. Therefore, it is necessary to validate the b charge determination in a data-driven way.
A double-tag method is used that relies on the existence of two oppositely charged b quarks. It has already been applied at LEP e.g. for the measurement of the forward-backward asymmetry in Z^0 to $b\bar{b}$ events [74]. At the LHC it is more difficult to select
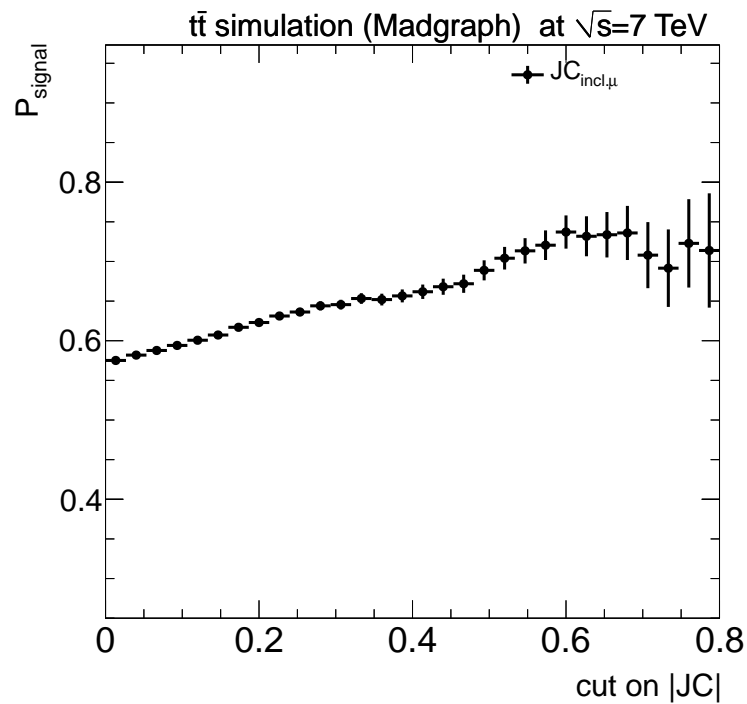
Figure 6.20: P_{signal} as a function of the minimal absolute jet charge cut value.
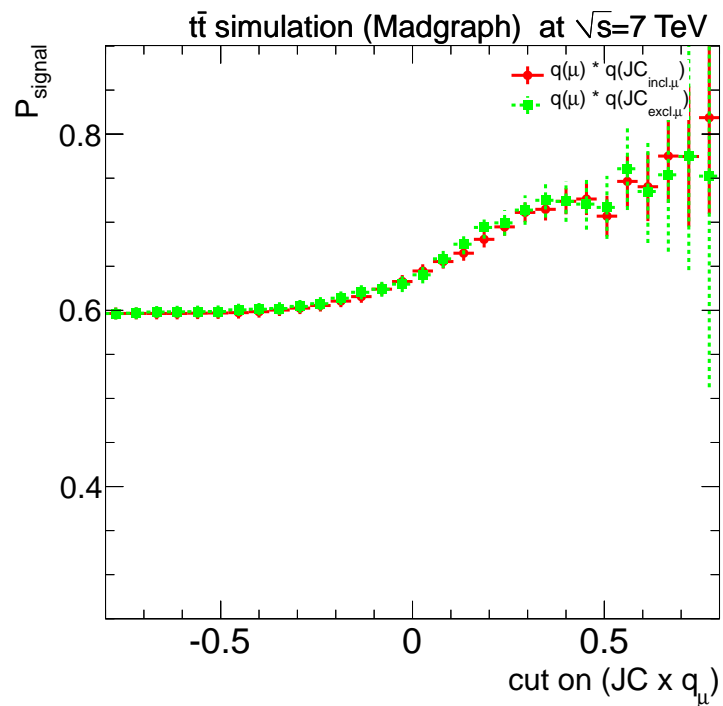

Figure 6.21: P_{signal} as a function of the cut value on $JC \times q_\mu$ for both options described in the text.
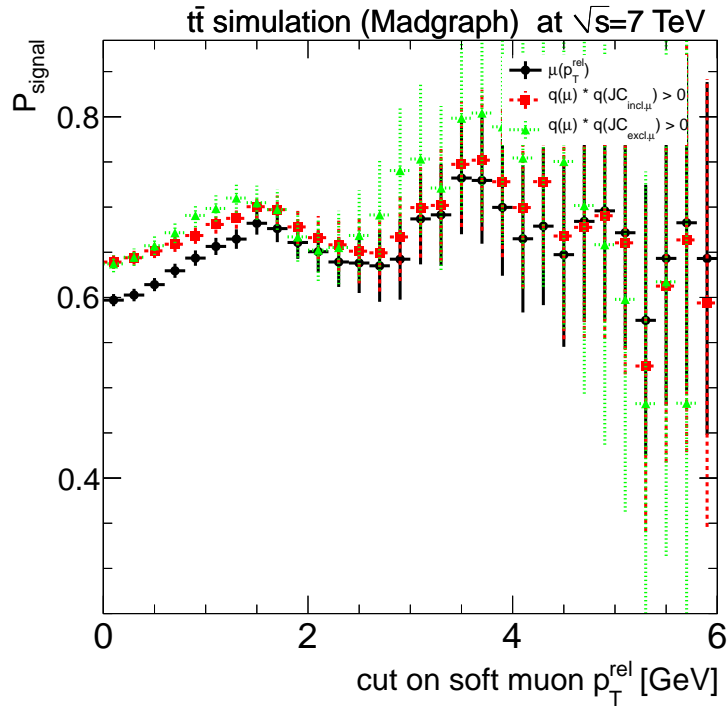
Figure 6.22: P_{signal} as a function of the minimum p_T^{rel} of the soft muon. The three different discriminators described in the text are investigated.
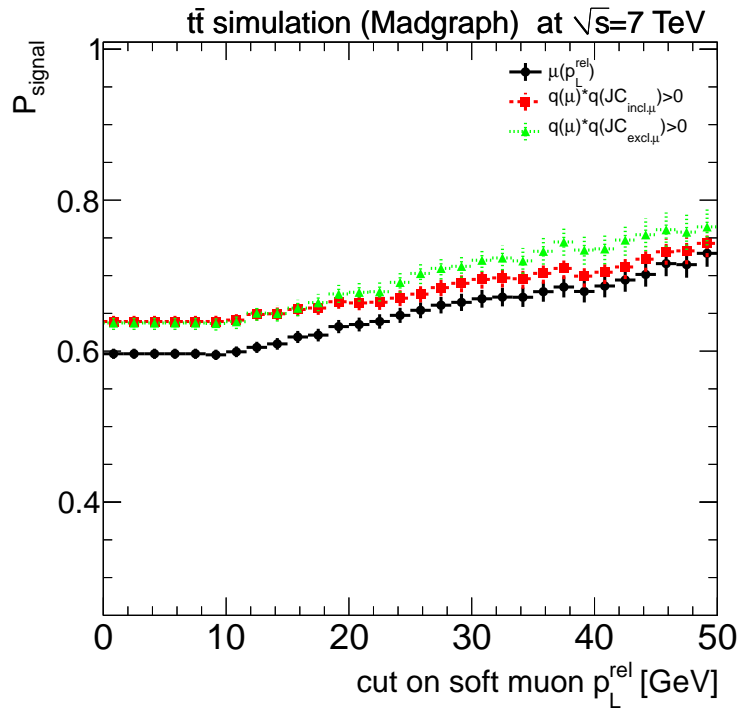

Figure 6.23: P_{signal} as a function of the minimum p_L^{rel} of the soft muon. The three different discriminators described in the text are investigated.

Z^0 bosons which decay into $b\bar{b}$. The triggers for dijet events are highly prescaled and $b\bar{b}$ events from strong interactions are dominant. However, the double-tag technique can be performed on $b\bar{b}$ events from strong interactions, too.

In the dijet MC sample b decays are simulated with the same software and fragmentation model as in the top-pair samples. Therefore, it is possible to use the relative uncertainty of P_{bC} determined on the dijet events for the top-pair analysis. The total value of P_{bC} integrates over the observables on which P_{bC} is sensitive, e.g. p_T^{rel} and p_T. Therefore, a difference in the distribution of these observables causes different total P_{bC} values. Figure 6.25 shows the soft muon p_T^{rel} and p_T distribution for the top pair and dijet sample for events where both b jets are identified. The differences can be caused by the higher boost of the b jets in the top pair events. Since the distributions are different the total value of P_{bC} will not be comparable between both samples. In the following, quantities derived from the multijet sample are denoted by the index '$b\bar{b}$'.

Figure 6.24 (left) shows P_{bC} as function of p_T^{rel} for the top pair and dijet sample where at least one b jet fulfills the charge requirements denoted as *1-tag* in the following. Both samples show comparable results. However, Figure 6.24 (right) shows a significant deviation for the dijet sample when placing the charge requirement on the second b jet as well denoted as *2-tag*. This can be induced by other b charge correlated observables that are affected by the requirement on the second b jet and is investigated in Section 6.4.4. Therefore, the most precise approach is to measure P_{bC} as a function of these observables. Due to a limited number of events this analysis measures only the total charge correlation and compares it to the simulation. The assumption of this approach is that other charge correlated observable or phase space difference are well modeled in simulation. Under this assumption it is justified to take the difference between data and simulation under the same selection criteria as systematic uncertainty. As a cross check P_{bC} is additionally extracted in two bins of p_T^{rel}.
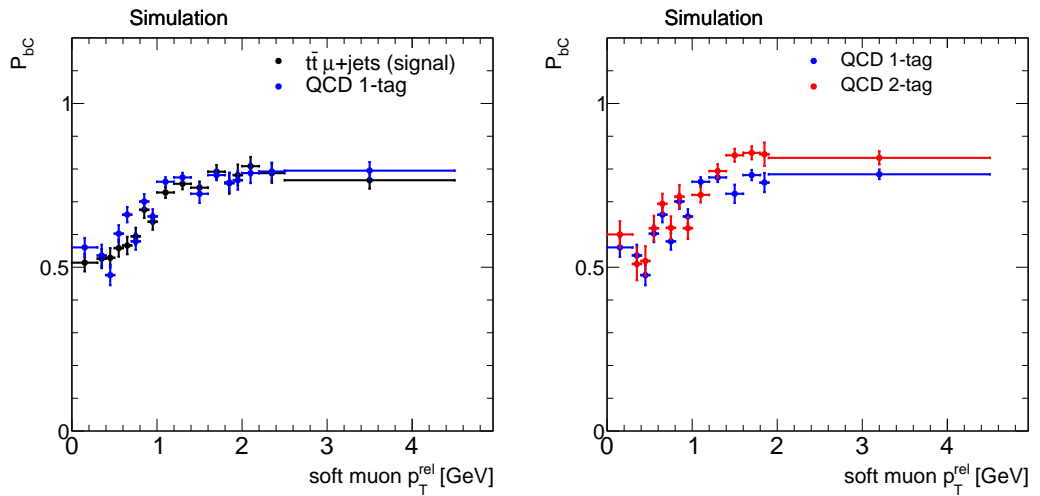

Figure 6.24: P_{bC} as a function of the soft muon's p_T^{rel} for $b\bar{b}$ and $t\bar{t}$ events.

After describing the event selection the relation between the measured quantity and P_{bC} is explained. Finally, the results are given and some systematic studies are presented.
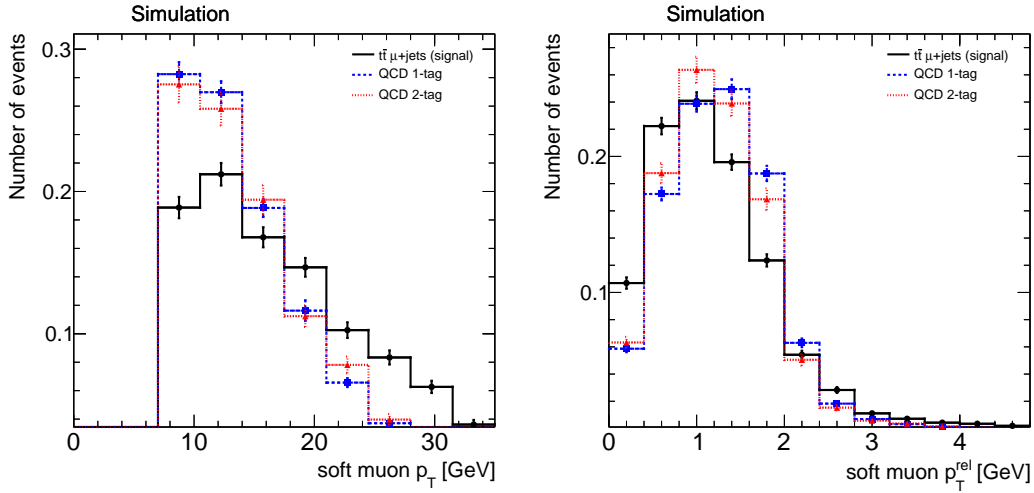
Figure 6.25: Soft muon p_T (left) and p_T^{rel} (right) distribution in dijet and top pair events. Both distributions are normalized to unit area.

6.4.1 Selection of $b\bar{b}$-enriched data

For this measurement data corresponding to an integrated luminosity of $L = 0.2\,\text{fb}^{-1}$ are investigated taken in early 2011. At this stage of LHC running the lower energetic double muon triggers were prescaled marginally. This data taking period is therefore well suited for this analysis. The $b\bar{b}$ selection consists of the following steps:

1 - Trigger : A double muon trigger for muons with $p_T > 6\,\text{GeV}$ is chosen.

2 - Two b Jets : There must be two b jets in the event passing the same quality criteria applied to jets in the top-pair selection (Section 5.4.3).

3 - Veto Third Jet : Events with a third jet passing a lowered p_T threshold of $p_T > 20\,\text{GeV}$ are rejected to enhance back-to-back dijet events and to suppress contributions from initial and final state radiation.

4 - b Charge : Both high-p_T jets must additionally pass the b-charge requirements: contain a soft global muon with $p_T > 8\,\text{GeV}$ to guarantee to be in the trigger plateau and $p_T^{rel} > 0.85\,\text{GeV}$.

The data are compared to simulated multijet events that are enriched with muons on generator level with a p_T requirement of $p_T > 5\,\text{GeV}$. These multijet events are generated in various $\hat{p}_T$ bins and stacked according to their cross sections. The total yield of simulated events is scaled to the yield measured in data. Figures 6.26 to 6.29 show acceptable agreement of the important observables between data and simulation. Using generator information Figure 6.30 shows that the event composition after the selection is dominated by $b\bar{b}$ events. The b identification is performed with a matching of reconstructed b jets to generated b quarks. If both lie within a ΔR cone of $\Delta R < 0.4$ the jet is identified as b jet.
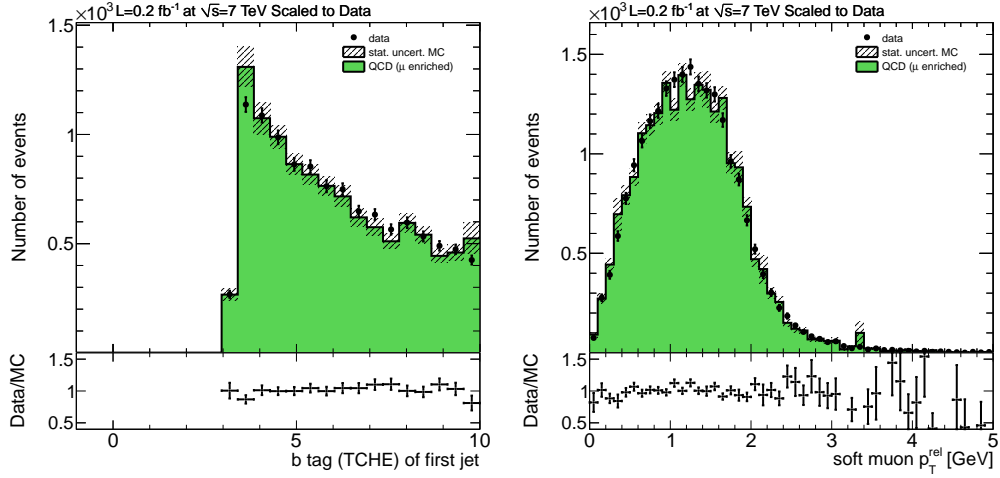
Figure 6.26: Data-MC comparison in multijet events of the b-tag discriminator of the highest-p_T jet (left) and p_T^rel of the soft muons within the two selected b jets (right).

6.4.2 Factorization scheme

The measured quantity of the double-tag method is the fraction S of events with two equally charged b jets. It is related with $P_\text{bC}^{b\bar{b}}$ via

$$S = 2 \cdot P_\text{bC}^{b\bar{b}} \cdot \left(1 - P_\text{bC}^{b\bar{b}}\right) . \tag{6.8}$$

Measuring an equally charged $b\bar{b}$ event implies to determine the charge of the initial quark on one side correctly, this is what the factor $P_\text{bC}^{b\bar{b}}$ accounts for, and to fail on the other side $\left(1 - P_\text{bC}^{b\bar{b}}\right)$. The wrong assignment can happen either for the b or the $\bar{b}$ quark, leading to a factor of 2.

The quadratic Equation 6.10 can be solved.

$$P_\text{bC}^{b\bar{b}} = \frac{1}{2} \mp \sqrt{\frac{1}{4} - \frac{S}{2}} \tag{6.9}$$

Furthermore, Equation 6.10 needs to be factorized to account for the cases $(1 - f_{b\bar{b}})$ where not a real $b\bar{b}$ event is identified.

$$S = 2f_{b\bar{b}} \cdot P_\text{bC}^{b\bar{b}} \cdot \left(1 - P_\text{bC}^{b\bar{b}}\right) + \left(1 - f_{b\bar{b}}\right) \cdot P_\text{fake} \tag{6.10}$$

P_fake is the fraction of same-sign events where an assignment to a $b\bar{b}$-quark pair is not possible.

This factorization assumes that the two quarks evolve independently. This is only an approximation due to the following effects [74]:

- The b hadrons are produced approximately back-to-back, see Figure 6.32. If one jet evolves in a poor acceptance region its likely that also the other one does. The requirement placed on the jets to be in the central region minimizes this effect.
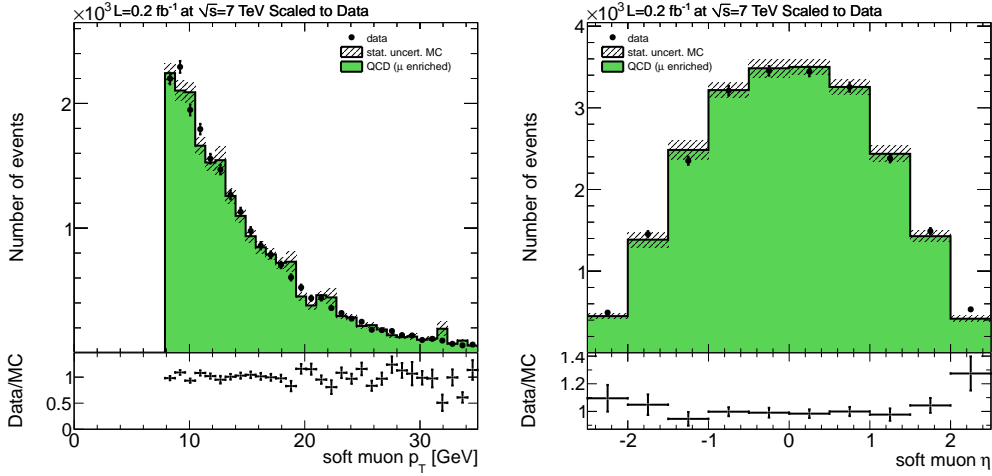
Figure 6.27: Data-MC comparison of p_T (left) and η (right) of the soft muons in the two selected b jets in multijet events.
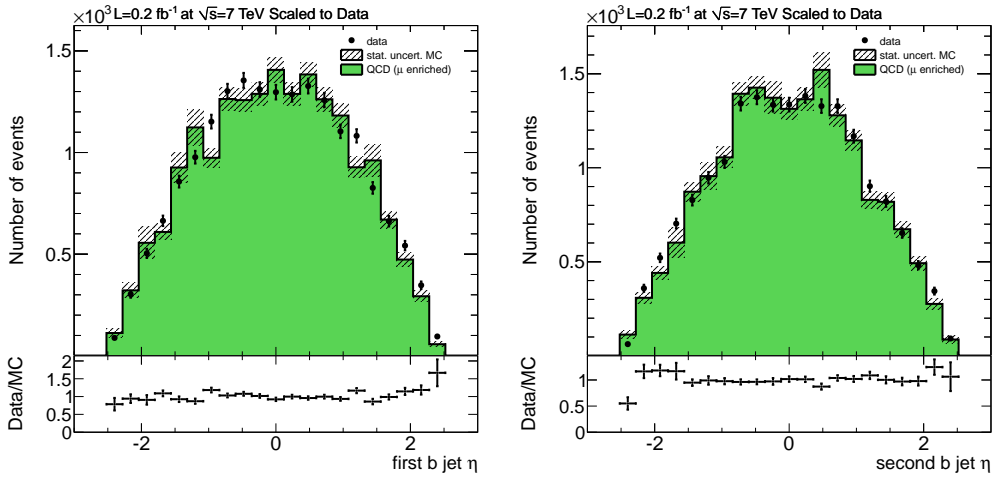

Figure 6.28: Data-MC comparison of the η-distributions of the highest p_T (left), second highest p_T jet (right) in multijet events.

- The initial emission of hard gluons reduces the momentum of the two primary quarks in the same way.

- As described in Section 5.5.1 the track-counting b tagger is used in this analysis. Figure 6.33 shows that the two selected jets are well separated. This minimizes the possible correlation between both sides.

Figure 6.31 shows the p_T and p_T^{rel} distribution of the soft muon from one b jet versus the soft muon from the other b jet for $b\bar{b}$-identified events. P_{bC} depends on these variables. There is no correlation visible. The correlation factors are 0.011 for the p_T distribution and 0.016 for the p_T^{rel} distribution. This confirms the validity of the factorization.
There is a mismatch between data and simulation in Figure 6.32 and Figure 6.33. This emphasizes the importance of determining P_{bC} from data. Since P_{bC} agrees with the expectation within uncertainties this inaccuracy in modeling the angular separation of the jets does not influence the b charge determination performance and therewith the top
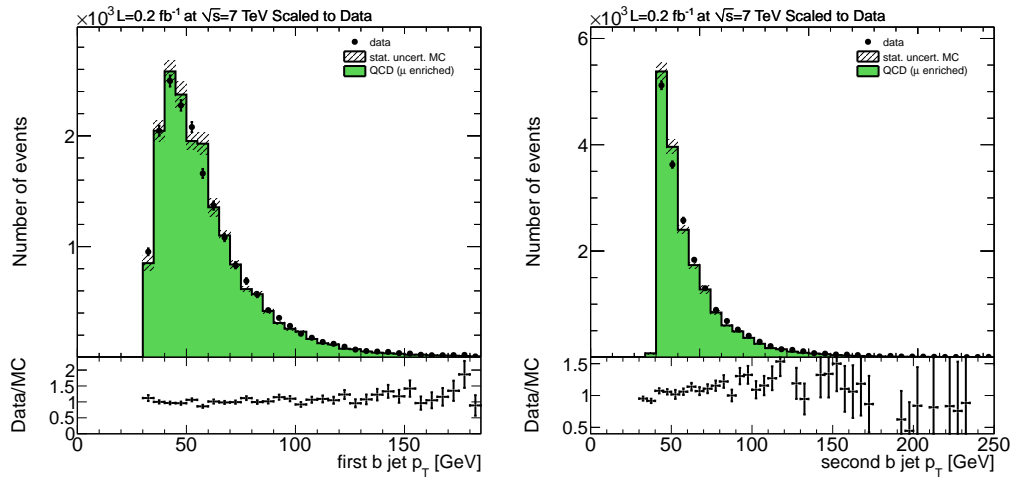
Figure 6.29: Data-MC comparison of the p_T-distributions of the highest p_T (left), second highest p_T jet (right) in multijet events. First and last bin include under- and overflow respectively.
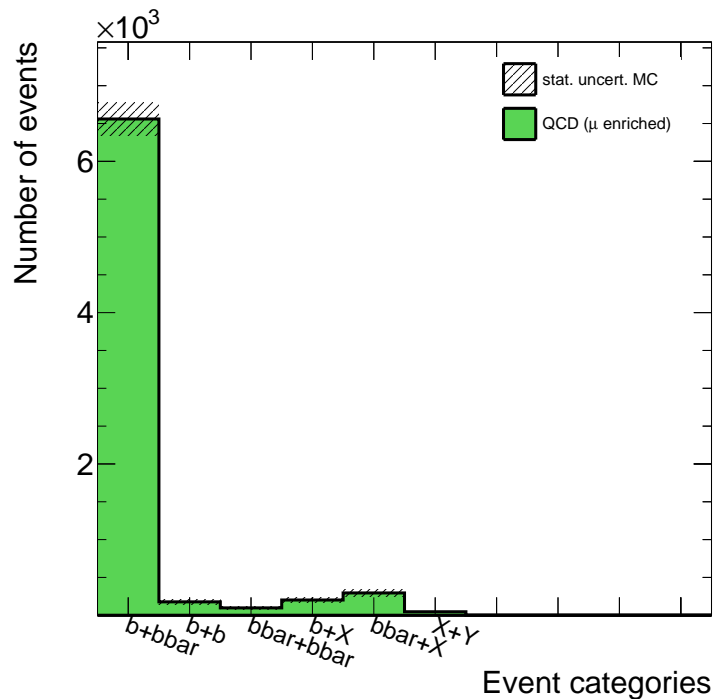

Figure 6.30: Event composition of the selected multijet events. "X" and "Y" denote any other matched particles than b or $\bar{b}$ quarks. There is a clear enrichment in $b\bar{b}$ events visible.
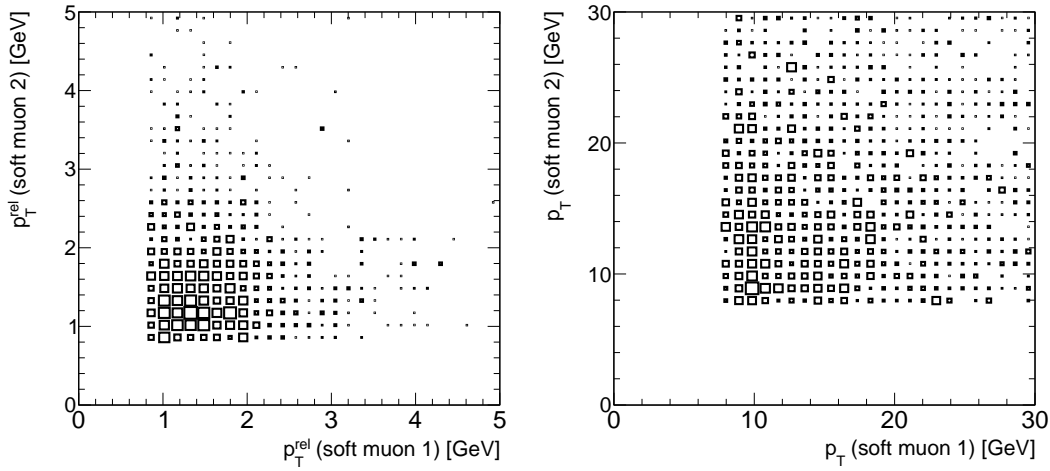
Figure 6.31: p_{T} and $p_{\mathrm{T}}^{\mathrm{rel}}$ distribution of the soft muon from one b jet versus the soft muon from the other b jet for $b\bar{b}$-identified events.

charge analysis result.

6.4.3   Results

By using the factorization scheme, with P_{fake} and $f_{b\bar{b}}$ determined from simulation, $P_{\mathrm{bC}}^{b\bar{b}}$ can be predicted. Table 6.2 summarizes the individual contributions from simulation (MC) and the resulting $P_{\mathrm{bC}}^{b\bar{b}}$ determined from data. Figure 6.34 shows the same in two $p_{\mathrm{T}}^{\mathrm{rel}}$ bins respectively. Only two bins have been chosen due to few selected events in the simulated samples. Figure 6.34 shows the distributions out of which the diagonal bins serve as input for Figure 6.35 $P_{\mathrm{bC}}^{b\bar{b}}$ as a function of $p_{\mathrm{T}}^{\mathrm{rel}}$.

Table 6.2: Results of the data driven b charge measurement and predictions from simulated multijet events. (MC) refers to the parameters that are determined from simulation. The uncertainties on P_{bC} are determined with Gaussian error propagation.

Specification	Factor
Fraction of same-sign events (Data)	$S^{\mathrm{Data}} = 0.35 \pm 0.009(\mathrm{stat.})$
Fraction of same-sign events (MC)	$S^{\mathrm{MC}} = 0.35 \pm 0.02(\mathrm{stat.})$
Fraction of events with two real b jets selected (MC)	$f_{b\bar{b}} = 0.87 \pm 0.02(\mathrm{stat.})$
Fraction of events with same signed fake b jets (MC)	$P_{\mathrm{fake}} = 0.54 \pm 0.08(\mathrm{stat.})$
Probability of correct b charge (MC)	$P_{\mathrm{bC}}^{b\bar{b},\mathrm{MC}} = 0.80 \pm 0.02(\mathrm{stat.})$
Result in data	$P_{\mathrm{bC}}^{b\bar{b},\mathrm{data}} = 0.80 \pm 0.01(\mathrm{stat.})$

Correct b charge determination succeeds on simulated multijet events in $P_{\mathrm{bC}}^{b\bar{b},\mathrm{MC}} = 80\,\%$ of all cases. This value for $P_{\mathrm{bC}}^{b\bar{b}}$ agrees within its errors with the measured value in dijet
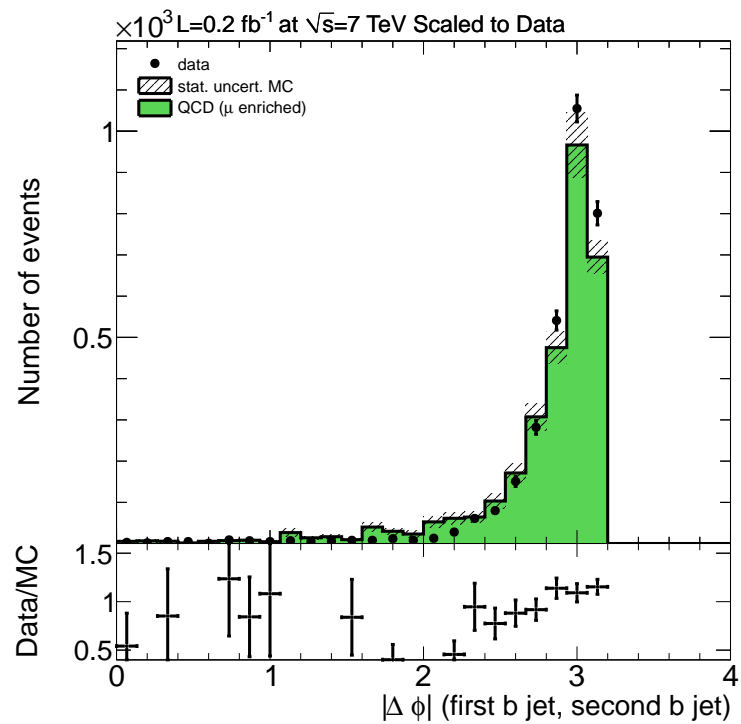
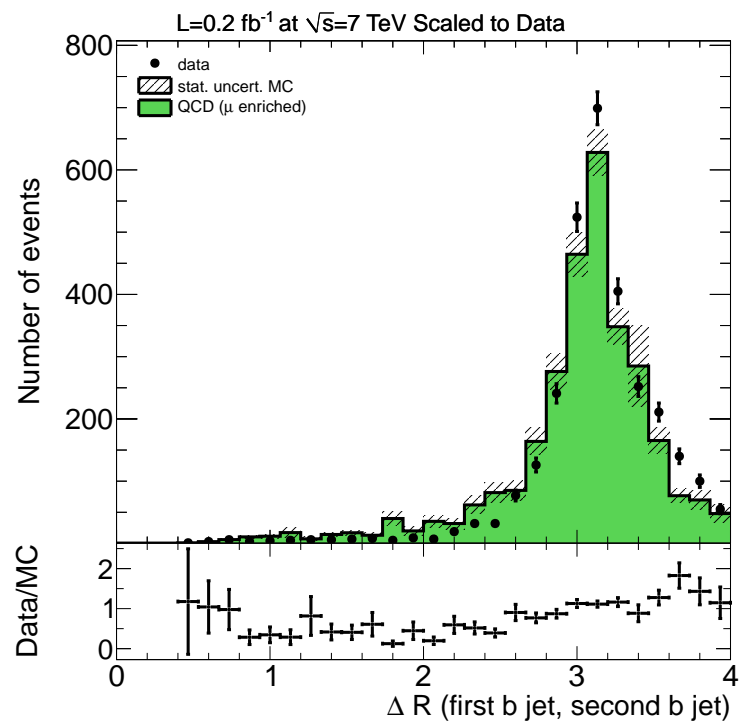**Figure 6.32:** $\Delta\phi$ between the two selected b jets.

Figure 6.33: ΔR between the two selected b jets. The two b are clearly separated.
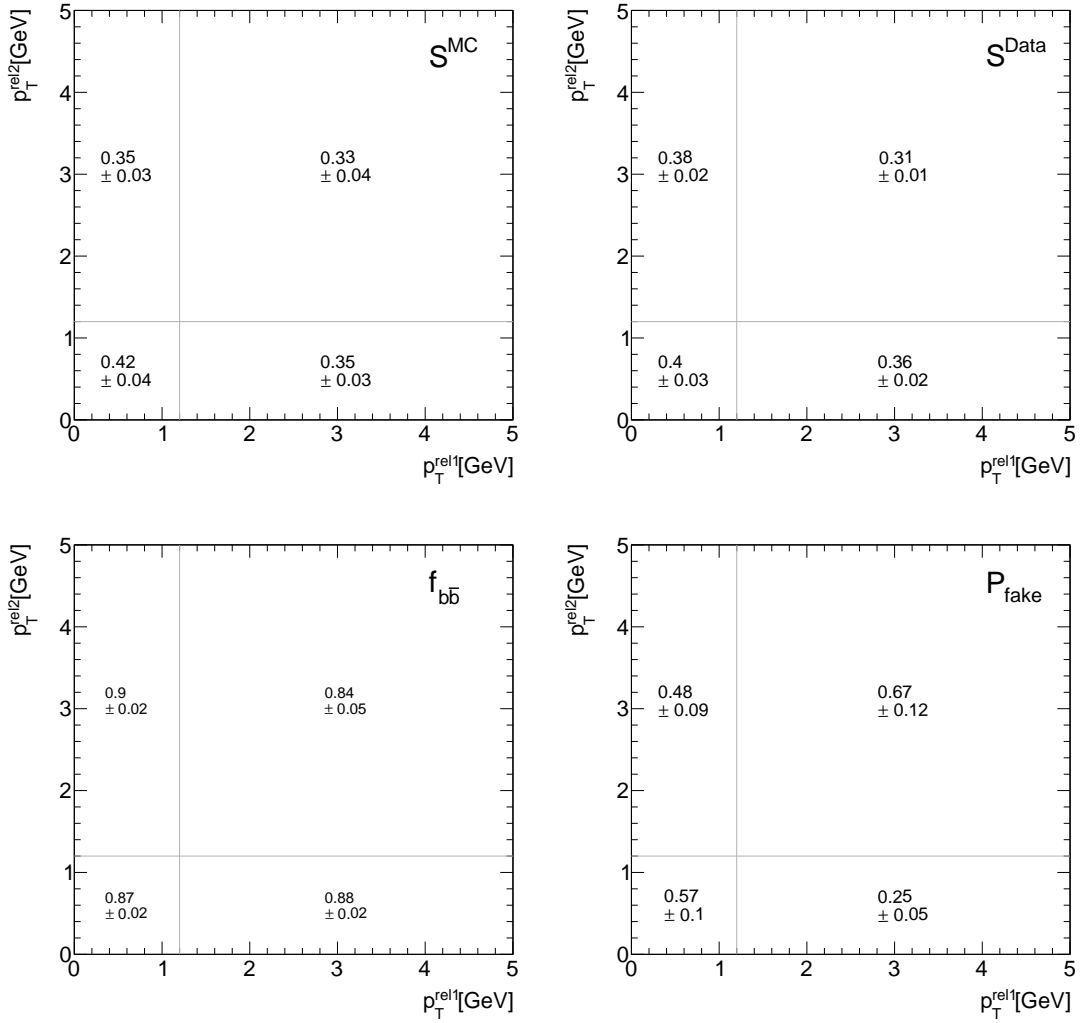
Figure 6.34: Contributions to Equation 6.10 in two p_T^{rel} of the two selected jets.
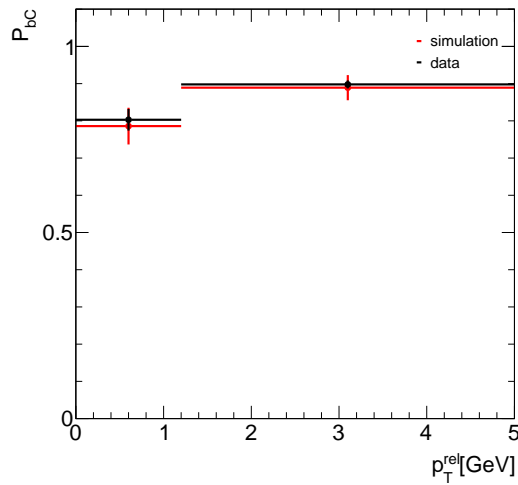

Figure 6.35: P_{bC} as function of p_T^{rel} determined from the diagonal entries from the distributions presented in Figure 6.34.

data of $P_{bC}^{b\bar{b},data} = 80\,\%$. The data driven cross check is successful within statistical errors. The statistical error is larger than the difference between the two values. As conservative assumption this statistical error is taken as systematic uncertainty on P_{bC} in the top charge analysis.

6.4.4 Systematic Studies

The jets in the multijet dataset have different properties than the jets in top-pair events. Figure 6.28 and 6.29 show the phase space of the two b jets in multijet events. The b jets are in the central region and have high momentum. There are some difference compared to the b jet distribution of top pair events shown in Figure 5.12. The b-charge determination is independent from these differences. This guarantees that the comparison between multijet sample and top pair events is adequate. Figure 6.38 shows that P_{bC} does not depend on p_T and η. The multijet sample was recorded in an earlier data taking period. Thus, it has on average fewer pileup vertices. The dependence on the number of pileup vertices and the b-tag value of the jet is studied in Figure 6.39. No dependence is seen.

The difference in P_{bC} as a function of the muon p_T^{rel} with and without the *2-tag* requirement, see Figure 6.24, is investigated by comparing the jet phase space observables and P_{bC}-dependencies on these. No difference in the distributions of the phase space observables is seen. Figure 6.36 shows exemplified the ΔR and $\Delta\phi$ distributions between the two b jets. There is no dependence of P_{bC} on the phase space observables neither. This is exemplified shown in Figure 6.37 for the jet-η.

Either the slight difference in the p_T and p_T^{rel} distributions with and without the *2-tag* requirement shown in Figure 6.25 is responsible for the difference in 6.24 or another charge correlated observables that could be affected by the *2-tag* requirement. The assumption is, that whatever causes the difference is well modeled by simulation.

As additional systematic cross check for P_{bC} a HERWIG sample with a different hadronization model is investigated, see Section 8.2.
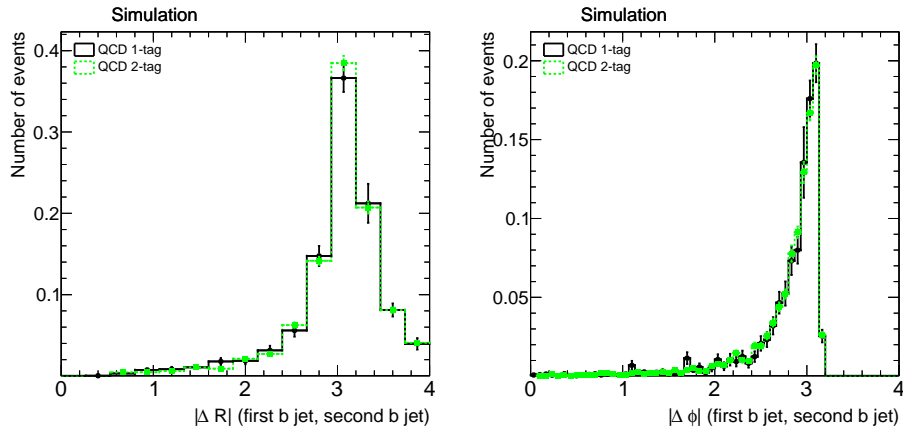

Figure 6.36: ΔR and $\Delta\phi$ distribution between the two b jets with (after) and without (before) the two-b-charged-jets requirement normalized to unit area.
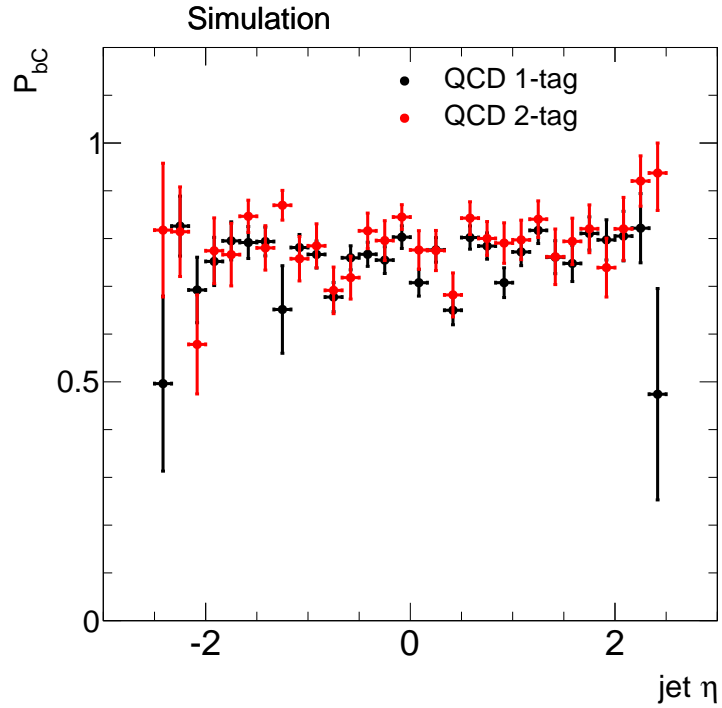
Figure 6.37: P_{bC} as a function of η of the jet with and without the two-b-charged-jets requirement.
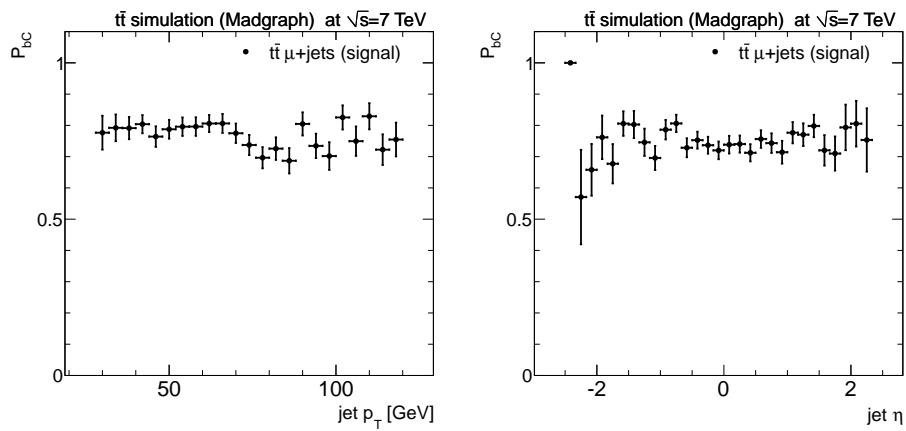

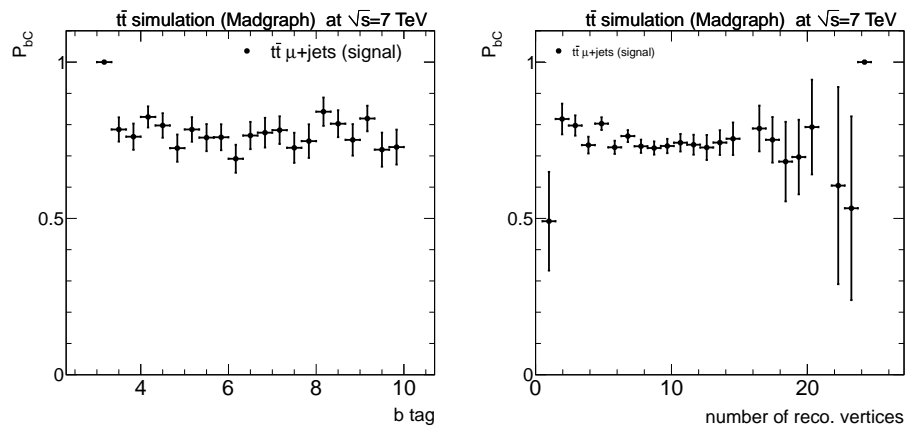**Figure 6.38:** P_{bC} as a function of p_T (left) and η (right) of the jet.

Figure 6.39: P_{bC} as a function of the b-tag value of the jet (left) and the number of reconstructed vertices in the event (right).

Chapter 7

Event Categorization

7.1 Summary of Quality Requirements

In the analysis the events are categorized in standard model events with a top charge of $|q_{\text{top}}| = 2/3$ e and exotic model events with $|q_{\text{top}}| = 4/3$ e. The requirements used to guarantee a good event reconstruction and b charge determination have already been discussed in the previous sections and are summarized here:

B-Jet Identification: There must be at least two b jets in the event (TCHE > 3.3).

Jet Association: The reconstructed invariant mass of the three jets considered to stem from the hadronic side must be less than $250\,\text{GeV}$. The invariant mass of the b jet from the leptonic decay side and the high-p_{T} muon $m(\mu, b_{\text{lep}})$ is restricted by the top mass. There is a veto on events where $m(\mu, b_{\text{lep}})$ exceeds $150\,\text{GeV}$ to suppress wrong assignments.

Charge Assignment: The presence of a muon inside one of the b tagged jets with $p_{\text{T}} > 8\,\text{GeV}$ and $p_{\text{T}}^{\text{rel}} > 0.85\,\text{GeV}$ is required. If both jets fulfill this requirement the jet with the highest b tag is considered for the measurement. The muon out of this jet serves as estimator for the b-quark's charge which happens only on a per cent level of the cases.

The event yield for data and simulated events after each requirement is shown in Figure 7.1. After all requirements events 766 events in data remain for the measurement.

7.2 Categorization Algorithm

Each selected event is classified to be standard model like or exotic model like using the following algorithm:

1. Measure the b charge of the selected b jet using a soft muon inside the jet.

2. Associate this jet with the hadronic or leptonic decay side using the jet-association algorithm.

3. The selected high-p_{T} muon is expected to be the muon from the leptonically decaying W boson.
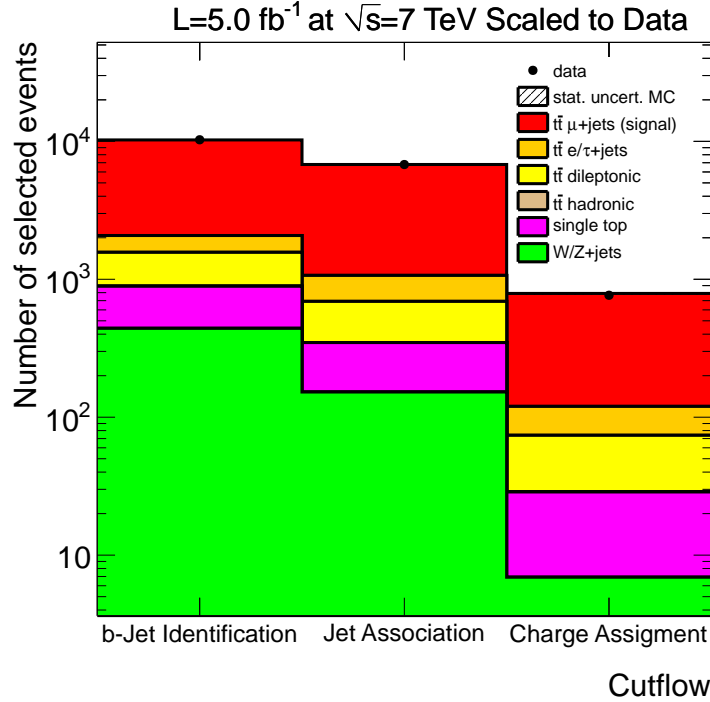
Figure 7.1: The event yield for data and simulated events for each quality requirement consecutively. After the requirement of at least two b jets no simulated events from QCD events remain.

4. Associate the charges of the W bosons by the charge of the high-p_T muon in the event. The leptonically decaying W boson charge equals the charge of the high-p_T muon in the event. The hadronically decaying W boson's charge is opposite to the high-p_T muon's charge.

5. Sum up the charges of the selected b-quark and of the W boson from the same decay side to reconstruct the charge of the top quark. If it is $2/3\,e$, the event contributes to the number of standard model like events N_{SM}. If it is $4/3\,e$, the event contributes to the number of exotic model like events N_{XM}.

The strategy of top-charge assignment is sketched schematically in Figure 7.2.

7.3 Factorization Scheme

The individual effects leading to the event categorization result can be factorized. The individual factors can then be determined independently. This allows to cross check the analysis chain and to investigate systematic uncertainties. Furthermore, one factor (P_{bC}) is measurable on data which allows to validate the result independently from simulation. The probability P to reconstruct a standard model like event is defined as

$$P = \frac{N_{SM}}{N_{SM} + N_{XM}} \quad .$$
(7.1)
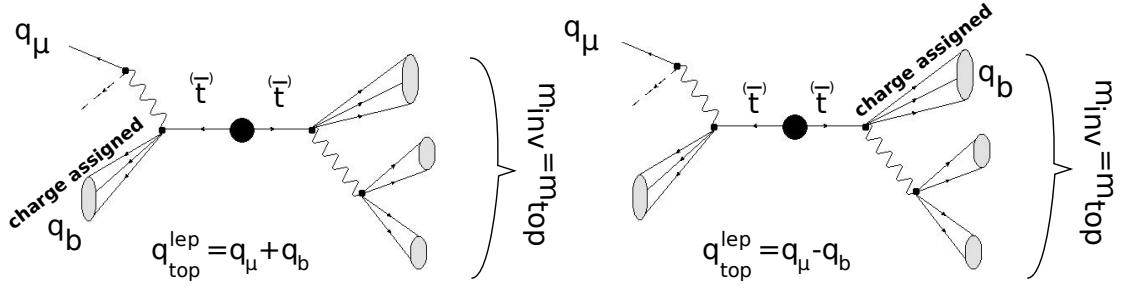
P can be factorized in terms of:

Figure 7.2: Charge reconstruction of the top-quark candidates: The b jet fulfilling the charge requirements is used to determine the charge of the b quark that the jet originates from. If this b jet is associated to the leptonic decay side (left) the top charge is calculated by $q_{\text{top}} = q_\mu + q_b$. If the charge-assigned b jet is associated to the hadronic decay side (right), the other b jet is considered to originate from the leptonic decay side and to have an opposite charge $-q_b$.

f_{BG}: fraction of selected background events.

P_{BG}: probability to reconstruct a standard model like background event.

P_{signal}: probability to reconstruct a standard model like signal event.

P_{btag}: probability to select a real b jet.

P_{pair}: probability for correct association of the selected b jet to the hadronic or leptonic decay side.

P_{bC}: probability to determine the b charge of the selected b jet correctly.

For the factorization of P one has to take into account that a few background events remain after the selection. Applying the categorization on the simulated background events shows that it is equally probable to reconstruct exotic or standard model like top charges within the uncertainties, $P_{BG} \approx 0.5$.

$$P = P_{\text{BG}} \cdot f_{\text{BG}} + (1 - f_{\text{BG}}) \cdot P_{\text{signal}} \qquad (7.2)$$

with the background fraction:

$$f_{\text{BG}} = \frac{N_{\text{BG}}}{N_{\text{signal}} + N_{\text{BG}}} \ . \qquad (7.3)$$

Moreover, the probability P_{signal} to reconstruct a standard model like signal event can be factorized as well. It needs to be taken into account that the b jet is not selected in $(1 - P_{\text{btag}})$ of the cases. Then no correlation to the top quark charge is expected and the probability $P_{\text{!btag, bC}}$ to reconstruct a standard model or exotic model top-quark charge should be 0.5. This is validated on simulated top pair events where the b jet is not selected. If the selected b jet is a real b jet the probability to measure the correct top charge is $P_{\text{pair, bC}}$, leading to:

$$P_{\text{signal}} = (1 - P_{\text{btag}}) \cdot P_{\text{!btag, bC}} + P_{b\,\text{tag}} \cdot P_{\text{pair, bC}} \ . \qquad (7.4)$$

Furthermore, $P_{\text{pair, bC}}$ can be factorized. Reconstructing a standard model event with a real b jet selected implies that the b-quark's charge is determined correctly and that the selected b jet is assigned correctly to the hadronic or leptonic decaying side ($P_{\text{bC}} \cdot P_{\text{pair}}$) or that both fail ($(1 - P_{\text{pair}}) \cdot (1 - P_{\text{bC}})$):

$$P_{\text{pair, bC}} = P_{\text{pair}} \cdot P_{\text{bC}} + (1 - P_{\text{pair}}) \cdot (1 - P_{\text{bC}}) \tag{7.5}$$

The single factors and the resulting prediction for P_{signal} are listed in Table 7.1.

Table 7.1: P_{signal} as calculated in Equation 7.4 and its contributing factors which give a measure of the analysis performance. The difference to the previously determined probabilities is due to a different preselection. The values listed here are determined on the jets selected for the final event categorization. The preselection for the values stated in the previous sections is the selection up to the requirement of at least two b jets. All numbers are determined from simulation. Only the most relevant systematic uncertainties discussed in Section 8.2 are taken into account in the determination of the systematic uncertainties on P_{btag}, P_{bC}, P_{pair} and $P_{\text{!btag, bC}}$ given here since they are listed for illustrating purposes. Only the inclusive systematic uncertainty on P_{signal} is used in the statistical interpretation of the top charge measurement.

Summary of Performance Numbers	
f_{BG}	$0.009 \pm 0.004(\text{stat.}) \pm 0.03(\text{sys.})$
P_{BG}	$0.36 \pm 0.20(\text{stat.}) \pm 0.72(\text{sys.})$
P_{btag}	$0.920 \pm 0.005(\text{stat.}) \pm 0.035(\text{sys.})$
P_{bC}	$0.756 \pm 0.009(\text{stat.}) \pm 0.041(\text{sys.})$
P_{pair}	$0.816 \pm 0.009(\text{stat.}) \pm 0.044(\text{sys.})$
$P_{\text{!btag, bC}}$	$0.56 \pm 0.03(\text{stat.}) \pm 0.22(\text{sys.})$
P_{signal}	$0.650 \pm 0.008(\text{stat.}) \pm 0.05(\text{sys.})$

7.4 Validation of the Factorization Scheme

It is possible to verify the factorization by restricting the categorization on events where the b-jet identification, the jet association or the b-charge determination succeeded. The result of the restricted categorization must then correspond to the remaining factor. This way, additional contributions to the factorization can be excluded. Table 7.2 shows the resulting factors.

In the first line of Table 7.2 the predicted outcome of the categorization of simulated signal events is listed using the factorization. The fraction of standard model like events is predicted correctly with the factorization scheme compared to the values listed in Table 7.1. This confirms that the factorization is complete and successful.

Table 7.2: Results of factorization test. In line 2, a real b jet is selected. In line 3, the b-jet association is correct. In this case an exotic hypothesis is reconstructed only if the b-charge determination fails. Therefore, this value corresponds to P_{bC}. In line 4, the b charge is determined correctly. In this case an exotic hypothesis is reconstructed only if the jet association fails. Therefore, this value corresponds to P_{pair}.

Restriction	Corresponding Factor	$P = N_{\mathsf{SM}}/(N_{\mathsf{SM}} + N_{\mathsf{XM}})$
None	P_{signal}	0.65
b-identification correct	$P_{pair,bC}$	0.75
b-jet association correct	P_{bC}	0.756
b-charge correct	P_{pair}	0.816

Chapter 8

Results and Interpretation

In Figure 8.1 the result of the categorization is shown. An exotic hypothesis would lead to a distribution indicated by the dashed line. The data is in good agreement with the simulated standard model prediction. In the following the test statistics used to determine the statistical significance of the measurement is explained. This clarifies how systematic uncertainties are taken into account which are described in more detail afterwards. Finally the statistical interpretation of the measurement is given.
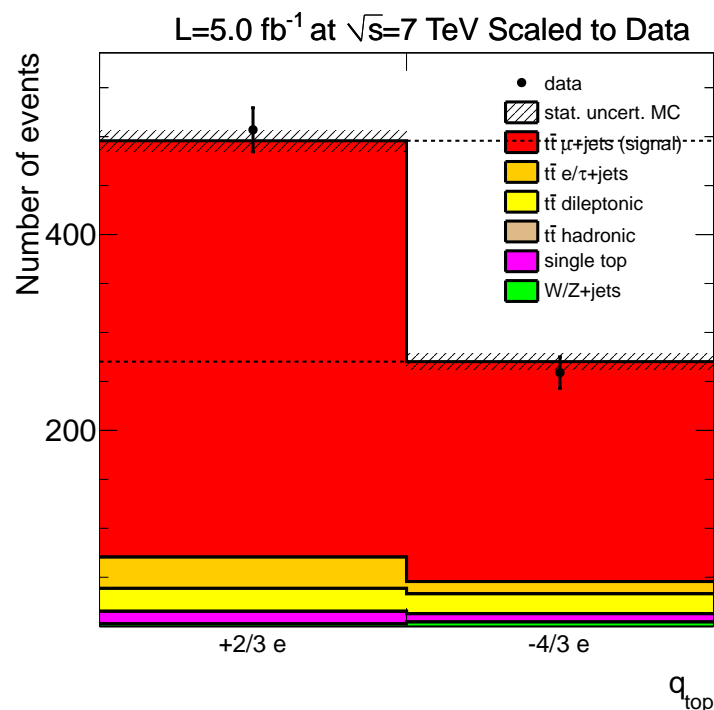

Figure 8.1: Categorized events scaled to the number of selected events in data. The shaded area corresponds to the statistical error of the simulated events. An exotic hypothesis would lead to a distribution indicated by the dashed line.

8.1 Test Statistics

The significance of the measurement is given by the p-value of the test statistics A:

$$A = \frac{N_{\mathrm{SM}} - N_{\mathrm{XM}}}{N_{\mathrm{SM}} + N_{\mathrm{XM}}} \, .$$ (8.1)

N_{i} is the number of events categorized to be standard model like (i=SM) or exotic model like (i=XM).

The asymmetry of Equation 8.1 is normalized in a way that the median expectation of the SM hypothesis is +1 and the expectation of the XM hypothesis is -1:

$$A \;=\; \frac{1}{2P_{\mathrm{signal}} - 1} \frac{(N_{\mathrm{SM}} - N_{\mathrm{BG}}P_{\mathrm{BG}}) - (N_{\mathrm{XM}} - N_{\mathrm{BG}}(1 - P_{\mathrm{BG}}))}{(N_{\mathrm{SM}} - N_{\mathrm{BG}}P_{\mathrm{BG}}) + (N_{\mathrm{XM}} - N_{\mathrm{BG}}(1 - P_{\mathrm{BG}}))} \, .$$ (8.2)

P_{i} is the probability to reconstruct a standard model like event on signal (i=signal) or background (i=BG) events, as defined in Equation (7.1). The expected number of background events N_{BG} is only considering background processes that are not correlated to the top quark's charge. Within statistical uncertainties, the non-correlated backgrounds populate equally the two categories of events, see Section 7.3.

Top pair and single top backgrounds are correlated to the top quark's charge and are treated as signal-like. Among others the correlation is due to the quality veto on the invariant mass of the leptonic decay side applied in the jet association. In this way the two b jets become distinguishable, and thus a correlation is observed.

8.2 Systematic Uncertainties

Choosing an asymmetry as test statistics has the advantage that the absolute normalization cancels and many systematic influences are minimized. Systematic effects which are not correlated to the difference between exotic and standard model like events are for example uncertainties on the luminosity or on scale factors that affect only the total event yield, e.g., trigger efficiency corrections.

Uncertainties that influence the measurement are the ones that affect the individual factors in Equation (8.2): P_{signal}, P_{BG} and f_{BG}. Each systematic effect on them is added up in quadrature and taken into account in the pseudo experiments by broadening their probability density functions. Though being affected by many systematic uncertainties only the most important systematic uncertainties of f_{BG} are investigated since the background contribution is very small. The same is valid for uncertainties on P_{BG}. Uncertainties on P_{signal} are the most relevant. All possible influences that affect the b-jet identification, the jet association and the b-charge assignment must be investigated, see Equation (7.4).

In the following details on the sources of systematic uncertainties and how they are taken into account are given. The results are summarized in Table 8.1 and 8.2.

Some of the systematic samples suffer from low statistics. If the difference of the absolute values is smaller than the statistical uncertainty the statistical uncertainty is taken as systematic uncertainty for the relevant uncertainties. Uncertainties that are not statistical independent because they are determined by variations on the reference sample or if the

contribution of the varied sample to P_{signal}, P_{BG} or f_{BG} is small, the absolute difference is taken as systematic uncertainty. Otherwise the systematic effect would be overestimated. It is indicated by an arrow if the statistical uncertainty or the absolute difference is taken. The statistical uncertainties are calculated using normal approximation with a mean at $passed/total$ and the width $\sigma^{\text{stat.}} = \sqrt{\epsilon(1 - \epsilon)/total}$.

Matching Threshold (*TT/W/DY/T/TbarMatchingUp/Down*): The additional jets in top-pair events are modeled either by higher order matrix elements or by hard emission during shower evolution. The potential double counting of the jet configurations needs to be identified and removed. Therefore, matching algorithms are applied [78]. The jet matching threshold is varied to investigate the effect of this procedure according to the conventions coherently applied in CMS analyses. CMS provides samples for signal and electro-weak background processes. P_{signal} and f_{BG} is affected by this uncertainty. The differences in the result are taken as systematic uncertainty and added in quadrature. It is the largest systematic uncertainty of this analysis.

b-Charge Mis-identification (*BCharge*): A data-driven technique is applied to verify P_{bC} on an independent $b\bar{b}$ sample The simulation describes the data correctly within statistical uncertainties. The data-driven measurement is dominated by the statistical uncertainty on the simulated multijet sample. As discussed in Section 6.4 it is considered as systematic uncertainty leading to a 2.5% uncertainty on P_{bC}.

Factorization/Normalization Scale (*TT/W/DY/T/TbarScaleUp/Down*): The scattering scale Q^2 is varied to investigate effects due to the modeling of the hard interaction process. CMS provides samples for signal and the most background samples. The differences in P_{signal} and f_{BG} between the varied samples and the reference sample are added in quadrature.

Top Mass (*TT169, TT175*): A top pole mass of 172.5 GeV in simulation and 173.5 GeV in data is assumed, see Section 2.3.1. The current uncertainty on the top mass is about 0.9 GeV [21]. This is taken into account by investigating simulated top-pair events with different top masses. P_{signal} agrees within the statistical uncertainty for the sample with a top mass of 169 GeV and the reference sample. The difference in P_{signal} between the 175 GeV mass sample and the reference sample is considered as systematic uncertainty on P_{signal} and added in quadrature.

Pileup (*pileup0p9, pileup1p1, 1BX*): The main uncertainties on the measured pileup distribution are uncertainties on the measured luminosity and on the total inelastic cross section. There are on average eight pileup interactions per event in the considered data, see Figure 4.3. The mean number of interactions is varied by 10% to investigate the effect of pileup modeling with the (*3D*) method. Additionally, an alternative pileup reweighting method (*1BX*) is applied. The differences of all three variations are taken as systematic uncertainties.

Jet-Energy Calibration (*JECUp/Down*): The jet energy calibration is varied within its uncertainties to study effects due to the jet energy measurement.

Jet-Energy Resolution (*JERUp/Down*): The jet energy resolution in simulation is increased for a better description of the resolution measured in data [49]. The uncertainty on the jet-energy resolution is determined by varying the correction within its uncertainties.

Generator (*Powheg*): Instead of MADGRAPH another signal sample generated with POWHEG has been investigated. The resulting difference has been taken into account as systematic uncertainty.

Fragmentation Model (*Herwig*): The data driven b-charge determination is limited by statistical precision and its systematic uncertainties are only investigated qualitatively. Additionally, a robust cross check is performed by estimating the b-charge determination performance on a sample with a different fragmentation scheme, the cluster model implemented in HERWIG. There is no MADGRAPH sample available with HERWIG showering only with the generator MC@NLO. Since the POWHEG sample shows no significant difference in the result the assumption is made that uncertainties due to the different generator and uncertainties due to the different fragmentation model do not compensate. Since the HERWIG sample is only available for another software version than the reference sample it is compared to another MADGRAPH signal sample with the corresponding software version. P_{signal} as well as the b-charge determination performance agrees well with the reference sample. The difference in P_{signal} is taken as systematic uncertainty.

b-Jet Identification (*bTagSFUp/Down*): The b-jet identification performance has direct impact on the asymmetry. The b-tagging performance has been measured as described in [7] and correction factors are provided. Systematic uncertainties are taken into account by varying the correction factors within their uncertainties [7].

Background Contribution (*CrossSection*): The uncertainty on the number of selected W+jets and DY+jets events is taken to be $50\,\%$ respectively. No multijet background event remains in simulation after the top-pair selection. It is expected that the multijet background is negligable after the whole selection or covered in the large systematic uncertainties on f_{BG}. The uncertainties on the event yields are propagated to f_{BG}.

8.3 Statistical Interpretation

By the use of pseudo experiments the probability density function of A can be generated. This approach has already been applied by the CDF experiment [4]. Binomial statistics are approximated by a normal approximation with the mean set to *passed/total* and the width set to $\sigma^{\mathrm{stat.}} = \sqrt{\epsilon(1-\epsilon)/total}$. Pseudo experiments are constructed in the following way:

$f_{\mathbf{BG}}$: Fraction of background events. It is modeled following a Gaussian distribution. The mean of the Gaussian is the expected fraction of background events. σ is set to the systematic uncertainty. If $f_{\mathrm{BG}} < 0$ or $f_{\mathrm{BG}} > 1$ the event is rejected.

N : Total number of selected events. It is modeled following Poisson statistics.

N_{BG} : Total number of background events. It is modeled following binomial statistics with the randomized number of selected events N and the randomized fraction of background events f_{BG}.

N_{signal} : Total number of signal events. Having randomized the number of selected events N and the number of background events N_{BG} it can be calculated as

$$N_{\text{signal}} = N - N_{\text{BG}} \tag{8.3}$$

P_{signal} : Fraction of standard model like signal events. It is modeled following a Gaussian. The mean of the Gaussian is the expected fraction of standard model like signal events. σ is set to the systematic uncertainty.

$N_{\text{SM, signal}}$: Total number of standard model like signal events. It is modeled following binomial statistics with the total number of events equal to the randomized N_{signal} and the probability of the randomized P_{signal}.

$N_{\text{XM, signal}}$: Total number of exotic model like signal events. Having randomized $N_{\text{SM, signal}}$ and N_{signal} is can be calculated as

$$N_{\text{XM, signal}} = N_{\text{signal}} - N_{\text{SM, signal}} \tag{8.4}$$

P_{BG} : Fraction of standard model like background events. It is smeared following a Gaussian with σ set to the systematic uncertainty.

$N_{\text{SM, BG}}$: Total number of standard model like background events. It is modeled following binomial statistics with the total number of randomized N_{BG} and the probability of the randomized P_{BG}.

$N_{\text{XM, BG}}$: Total number of exotic model like background events. Having randomized $N_{\text{SM, BG}}$ and N_{BG} it can be calculated as

$$N_{\text{XM, BG}} = N_{\text{BG}} - N_{\text{SM, BG}} \tag{8.5}$$

N_{SM} and N_{XM} can then be calculated as:

$$N_{\text{SM}} = N_{\text{SM, signal}} + N_{\text{SM, BG}} \tag{8.6}$$

$$N_{\text{XM}} = N_{\text{XM, signal}} + N_{\text{XM, BG}} \tag{8.7}$$

P_{signal} is set to $1 - P_{\text{signal}}$ to model the exotic top quark scenario ($P_{\text{signal}}^{\text{XM}} = 1 - P_{\text{signal}}^{SM}$). Figure 8.2 shows the result of the pseudo experiments including all systematic uncertainties discussed in Section 8.2.

The rejection region (significance of the test) is defined as

$$\alpha = \int_{A_{\text{cut}}}^{\infty} f(X|H_0) dA \tag{8.8}$$

and gives the probability to reject the null hypothesis assuming it is correct (error of the first kind).

$$1 - \beta = 1 - \int_{-\infty}^{A_{\text{cut}}} f(A|H_1)dA \qquad (8.9)$$

is called the power of the test and gives the confidence level to exclude the exotic model. Due to numerical precision and low statistics at the tails of the probability density function of A, since the standard model is probed at a 5σ-CL, no integration is performed but the p-value is determined in the following way. A Gauss fit of the probability density function of the exotic scenario is performed. The distance between the mean value of the distribution and the asymmetry measured from data is determined in units of the fitted standard deviation. The number of standard deviations can be translated into a p-value. It has been checked that the Gauss describes the probability density function of A very good.

The result of the pseudo experiments is shown in Figure 8.2 with and without taking into account the systematic uncertainties discussed in Section 8.2. The error rates of Type I and II as a function of the asymmetry is shown in Figure 8.3. A normalized asymmetry of

$$A = 1.10 \pm 0.12(\text{stat.}) \pm 0.34(\text{sys.}) \qquad (8.10)$$

is measured. The measurement is limited by systematic uncertainties. Due to the sensitivity of the obtained p-value to the modeling of these uncertainties it cannot be used directly to quantify the degree of exclusion of the exotic scenario. Instead the result is interpreted in terms of a p-value test under an exotic null hypothesis. For a direct comparison with already published results [79] the Type I error rate α is chosen a priori at the value where both error rates are equal corresponding to $\alpha = 0.001$. This results in a 99.9% CL exclusion. The recently published ATLAS results [5] compare the measured asymmetry with the asymmetry at a $5\,\sigma$ CL. Figure 8.3 shows that the measured asymmetry is beyond the $5\,\sigma$ CL and therewith nicely confirms the SM expectation.
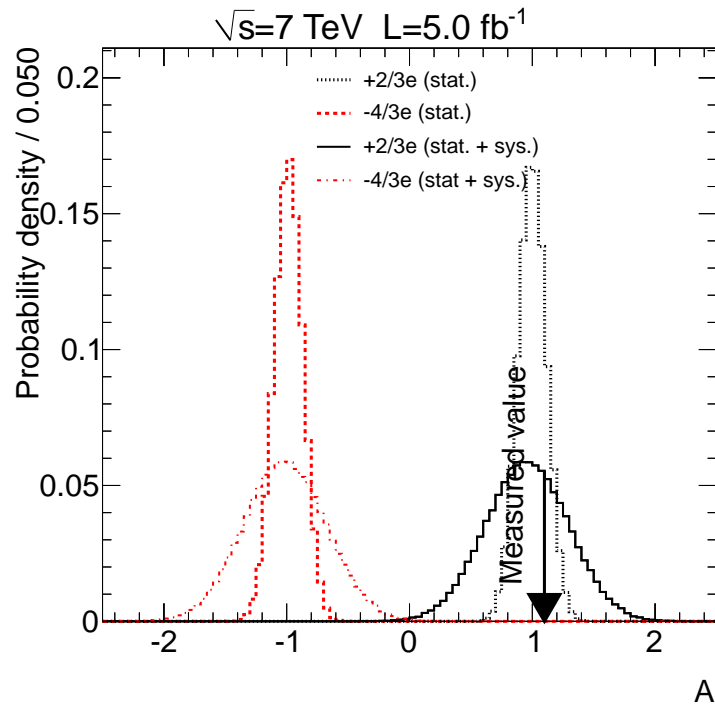
Figure 8.2: Probability density function of the asymmetry obtained from pseudo experiments for the SM and the exotic hypothesis. The vertical arrow corresponds to the measured asymmetry.
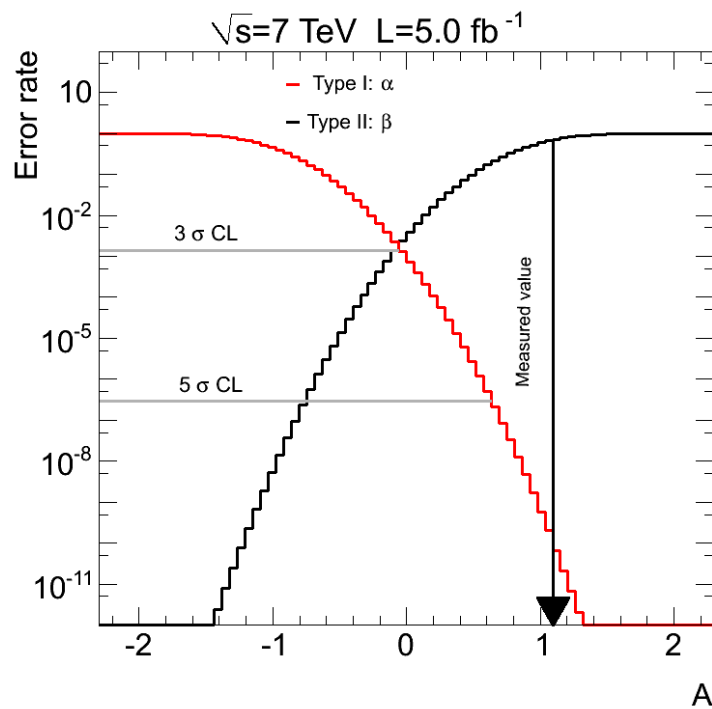

Figure 8.3: Type I (α) and Type II (β) error rate of the hypothesis test as a function of the asymmetry, including all systematic uncertainties. The vertical arrow indicates the measured asymmetry.
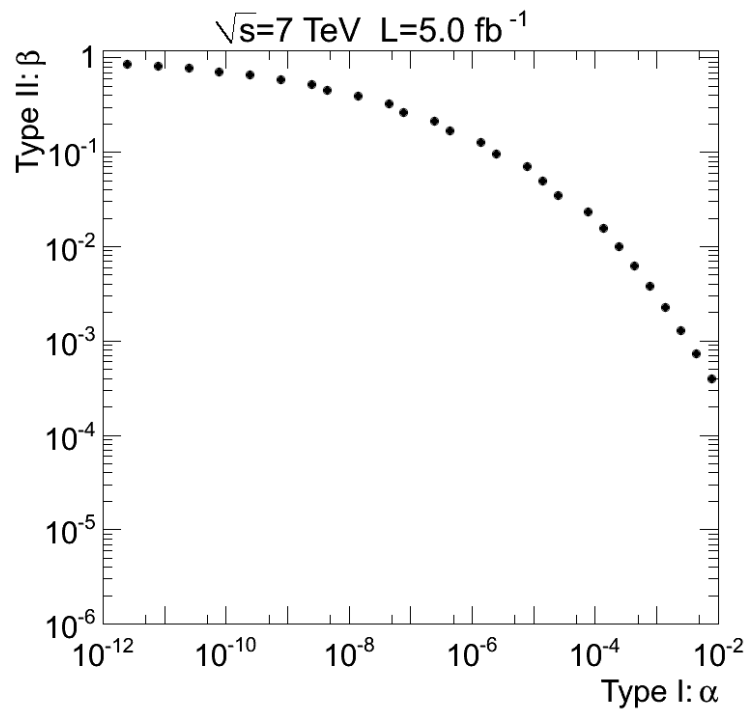
Figure 8.4: Type I (α) error rate as function of Type II (β) error rate of the hypothesis test, including all systematic uncertainties.

Table 8.1: Summary of all considered systematic uncertainties. The values of the three variables P_{signal}, P_{BG} and f_{BG} that affect A and their statistical uncertainties as well as their absolute deviation to the reference value are listed. The statistical uncertainty is determined by binomial statistics. Pseudo experiments have been performed regarding each source of systematic uncertainty separately to determine the impact on the width σ_A of the resulting probability density function of A. The width is extracted by a Gauss fit. The checkmark indicates which uncertainty is taken into account.

Source	σ_A	P_{signal}	$\sigma_{P_{\text{signal}}}^{\text{stat.}}$	$\sigma_{P_{\text{signal}}}^{\text{abs.}}$	f_{BG}	$\sigma_{f_{\text{BG}}}^{\text{stat.}}$	$\sigma_{f_{\text{BG}}}^{\text{abs.}}$	P_{BG}	$\sigma_{P_{\text{BG}}}^{\text{stat.}}$	$\sigma_{P_{\text{BG}}}^{\text{abs.}}$
Reference	0.3362	0.6498	0.0082	0.0471	0.0088	0.0038	0.0227	0.3567	0.1965	0.7163
statUncert	0.1163	–	–	–	–	–	–	–	–	–
TTMatchingUp	0.1146	0.6327	0.0149	0.0171✓	–	–	–	–	–	–
TTMatchingDown	0.1686	0.6246	0.0144	0.0252✓	–	–	–	–	–	–
BCharge	0.1114	0.6665	–	–0.0167✓	–	–	–	–	–	–
TTScaleDown	0.1032	0.6482	0.0154✓	0.0015	–	–	–	–	–	–
TTScaleUp	0.1097	0.6471	0.0164✓	0.0027	–	–	–	–	–	–
TT169	0.083	0.6504	0.0124✓	–0.0006	–	–	–	–	–	–
TT175	0.0942	0.6357	0.0121	0.0141✓	–	–	–	–	–	–
pileup0p9	0.0102	0.6486	0.0081	0.0012✓	–	–	–	–	–	–
pileup1p1	0.0104	0.651	0.0084	–0.0012✓	–	–	–	–	–	–
1BX	0.0119	0.6483	0.0076	0.0015✓	–	–	–	–	–	–
JERUp	0.007	0.6501	0.0082	–0.0003✓	–	–	–	–	–	–
JERDown	0.0121	0.6483	0.0082	0.0015✓	–	–	–	–	–	–
JECUp	0.0158	0.6519	0.0083	–0.0022✓	–	–	–	–	–	–
JECDown	0.0705	0.6393	0.0082	0.0105✓	–	–	–	–	–	–

Table 8.2: Continuation of Table 8.1.

Source	σ_A	P_{signal}	$\sigma_{P_{\text{signal}}}^{\text{stat.}}$	$\sigma_{P_{\text{signal}}}^{\text{abs.}}$	f_{BG}	$\sigma_{f_{\text{BG}}}^{\text{stat.}}$	$\sigma_{f_{\text{BG}}}^{\text{abs.}}$	P_{BG}	$\sigma_{P_{\text{BG}}}^{\text{stat.}}$	$\sigma_{P_{\text{BG}}}^{\text{abs.}}$
Reference	0.3362	0.6498	0.0082	0.0471	0.0088	0.0038	0.0227	0.3567	0.1965	0.7163
T_sScaleDown	0	0.6498	0.0082	0✓	—	—	—	—	—	—
T_sScaleUp	0.0067	0.6497	0.0082	0.0001✓	—	—	—	—	—	—
Tbar_sScaleDown	0.0067	0.6497	0.0082	0.0001✓	—	—	—	—	—	—
Tbar_sScaleUp	0	0.6498	0.0082	0✓	—	—	—	—	—	—
Tbar_tScaleDown	0.0068	0.65	0.0082	−0.0002✓	—	—	—	—	—	—
Tbar_tScaleUp	0.0086	0.6506	0.0082	−0.0008✓	—	—	—	—	—	—
T_tW_DRScaleUp	0.0179	0.6473	0.0082	0.0025✓	—	—	—	—	—	—
T_tW_DRScaleDown	0.0118	0.6483	0.0082	0.0015✓	—	—	—	—	—	—
Tbar_tW_DRScaleUp	0.0138	0.6516	0.0082	−0.0018✓	—	—	—	—	—	—
Tbar_tW_DRScaleDown	0.0087	0.6506	0.0082	−0.0008✓	—	—	—	—	—	—
Herwig	0.0129	0.6481	—	0.0017✓	—	—	—	—	—	—
Powheg	0.0249	0.6462	—	0.0036✓	—	—	—	—	—	—
bTagSFUp	0.0071	0.6502	0.0077	−0.0004✓	—	—	—	—	—	—
bTagSFDown	0.0073	0.6494	0.0089	0.0004✓	—	—	—	—	—	—
WMatchingUp	0.0179	—	—	—	0.004	0.0015	0.0047	0.5539	0.1891	−0.1972✓
WMatchingDown	0.018	—	—	—	0.004	0.0015	0.0047	0.5539	0.1891	−0.1972✓
DYToLLMatchingUp	0.0453	—	—	—	0.0228	0.0151	−0.014	0.0395	0.07	0.3172✓
DYToLLMatchingDown	0.0509	—	—	—	0.0197	0.014	−0.0109	0.8063	0.1889	−0.4496✓
WScaleUp	0.0179	—	—	—	0.004	0.0015	0.0047	0.5539	0.1891	−0.1972✓
WScaleDown	0.0179	—	—	—	0.004	0.0015	0.0047	0.5539	0.1891	−0.1972✓
DYToLLScaleDown	0.0152	—	—	—	0.0048	0.0036	0.004	0.1913	0.2936	0.1654✓
DYToLLScaleUp	0.0153	—	—	—	0.0048	0.0036	0.004	0.1913	0.2936	0.1654✓
CrossSection	0.0157	—	—	—	0.0175	—	−0.0088✓	—	—	—

Chapter 9

Summary and Outlook

The data accumulated by the CMS experiment in 2011 allow for precise tests of the standard model of particle physics. The top quark as the last quark discovered in 1995 plays a special role within the standard model due to its high mass and its consequential prominent properties. Therefore, the top quark is an integral part of new physics searches.

This analysis is dedicated to constrain the top quark charge. The charge of the top quark is classified by investigating the full $5.0\,\text{fb}^{-1}$ of certified proton-proton collision data collected with the CMS detector in 2011. Top pair events are selected by exploiting the distinctive signature of the muon+jets decay channel. The b jets are identified by a robust algorithm that relies on the high impact parameter of the b jet's tracks. b jets are associated to the hadronic and leptonic decay side by the use of a top mass constraint.

The reconstructed charge of a soft muon inside a b jet is proved to be a solid indicator of the charge sign of the original b parton. Alternatively, a jet-charge discriminator is investigated and optimized. Since the result is already dominated by systematic uncertainties the soft muon method is chosen due to its better b charge determination performance despite its lower efficiency. The b charge determination has been investigated both with simulated events, and in a data-driven approach on $b\bar{b}$-enriched events.

The correlation between the soft muon inside the b jet and the high-p_T muon in the event is used to categorize the $t\bar{t}$ events either in SM like events with a top quark charge $q_{\text{top}} = +2/3\,\text{e}$ or in an exotic scenario of 'top like' quarks with a hypothetical charge of $q_{\text{top}} = -4/3\,\text{e}$. The categorization is compared to the expectation for the two scenarios. The exotic scenario which corresponds to an asymmetry of $A = -1$ can be excluded with high significance and the measured asymmetry of $A = 1.10 \pm 0.12(\text{stat.}) \pm 0.34(\text{sys.})$ confirms the SM expectation of $A = +1$.

The result is interpreted in terms of a p-value test under an exotic null hypothesis. For a direct comparison with already published results [79] the Type I error rate α is chosen a priori at the value at which both error rates are equal corresponding to $\alpha = 0.001$. This results in a 99.9% CL exclusion. By comparing the measured asymmetry with the asymmetry at a $5\,\sigma$ CL in Figure 8.3 a top quark charge scenario of $q_{\text{top}} = -4/3\,\text{e}$ can be excluded with more than $5\,\sigma$ CL.

The next step is a direct measurement of the charge by investigating the electromagnetic coupling to photons radiated of the top. This has been already performed by the CDF [80] and ATLAS [81] experiments with low precision. A promising CMS study is on the way [82].

Bibliography

[1] J. BERINGER ET AL. (PARTICLE DATA GROUP). **2012 Review of Particle Physics**. *Phys. Rev. D86, 010001*, (2012). http://pdg.lbl.gov/.

[2] CMS COLLABORATIONS. **Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC**. *Submitted to Physics Letters B*, (2012). arXiv:1207.7235v1.

[3] H. MURAYAMA. **Theory of Neutrino Masses and Mixings**. *Int.J.Mod.Phys.A17:3403-3420*, (2002). doi:10.1142/S0217751X02012818, arXiv:hep-ph/0201022v3.

[4] CDF COLLABORATION. **Exclusion of an Exotic Top Quark with -4/3 Electric Charge Using Soft Lepton Tagging**. *Phys.Rev.Lett. 105 (2010) 101801*, (2010). arXiv:1006.4597v1.

[5] ATLAS COLLABORATION. **Measurement of the top quark charge in pp collisions at sqrt(s) = 7 TeV in the ATLAS experiment**. (2011). ATLAS-CONF-2011-141.

[6] CMS COLLABORATION. **Measurement of ttbar Pair Production Cross Section at sqrt(s)=7 TeV using b-quark Jet Identification Techniques in Lepton + Jet Events**. *CMS Physics Analysis Summary*, (2011). CMS-PAS-TOP-11-003.

[7] CMS COLLABORATION. **Performance of b-jet identification in CMS**. (2011). CMS-PAS-BTV-11-001.

[8] CMS COLLABORATION. **Constraints on the Top-Quark Charge from Top-Pair Events**. *CMS Physics Analysis Summary*, (2012). CMS-PAS-TOP-11-031.

[9] N. KIDONAKIS AND B. D. PECJAK. **Top-quark production and QCD**. *invited review for a special 'Top and flavor physics in the LHC era' issue of the European Physical Journal C*, (2011). arXiv:1108.6063v1.

[10] S. NAUMANN-EMME. **Simultaneous Measurement of Top Quark Mass and Jet Energy Scale Using Template Fits at the CMS Experiment**. *Department of physics, Hamburg University*, (2011). PHD thesis.

[11] N. KIDONAKIS. **Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution**. *Phys.Rev.D82:114030*, (2010). arXiv:1009.4935v2, doi:10.1103/PhysRevD.82.114030.

[12] CMS COLLABORATION. **Combination of top pair production cross sections in pp collisions at 7 TeV and comparisons with theory**. (2011). CMS-PAS-TOP-11-001.

[13] ATLAS COLLABORATION. **Evidence for the associated production of a W boson and a top quark in ATLAS at sqrt(s) = 7 TeV**. *Physics Letters B*, (2012). doi:10.1016/j.physletb.2012.08.011, arXiv:1205.5764v2.

[14] ATLAS COLLABORATION. **Measurement of the t-channel single top-quark production cross section in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector**. *Physics Letters B*, (2012). arXiv:1205.3130v1.

[15] CMS COLLABORATION. **First measurement of $B(t \rightarrow Wb)/B(t \rightarrow Wq)$ in the dilepton channel in pp collisions at sqrt(s)=7 TeV**. *CMS Physics Analysis Summary*, (2012). CMS-PAS-TOP-11-029.

[16] M. C. SMITH AND S. S. WILLENBROCK. **Top-quark pole mass**. *Phys.Rev.Lett.79:3825-3828*, (1997). arXiv:hep-ph/9612329v1, doi:10.1103/PhysRevLett.79.3825.

[17] M. BENEKE AND V. M. BRAUN. **Heavy Quark Effective Theory beyond Perturbation Theory: Renormalons, the Pole Mass and the Residual Mass Term**. *Nucl.Phys. B426 (1994) 301-343*, (1994). arXiv:hep-ph/9402364v2, doi:10.1016/0550-3213(94)90314-X.

[18] W. A. BARDEEN, A. J. BURAS, D. W. DUKE, ET AL. **Deep-inelastic scattering beyond the leading order in asymptotically free gauge theories**. (2000). doi:10.1103/PhysRevD.18.3998.

[19] M. BENEKE ET AL. **Top quark physics**. *CERN-TH-2000-100*, (2000). arXiv:hep-ph/0003033.

[20] K NAKAMURA ET AL. (PARTICLE DATA GROUP). **Review of particle physics**. *J. Phys. G: Nucl. Part. Phys. 37 075021*, (2010). doi:10.1088/0954-3899/37/7A/075021.

[21] TEVATRON ELECTROWEAK WORKING GROUP FOR THE CDF AND D0 COLLABORATIONS. **Combination of CDF and DO results on the mass of the top quark using up to 5.8 fb-1 of data**. (2011). arXiv:1107.5255v3.

[22] I. I. BIGI, Y. L. DOKSHITZER, V. KHOZE, J. KÜHN, AND P. ZERWAS. **Production and Decay Properties of Ultra-Heavy Quarks**. *Phys. Lett. B181, 157*, (1986). doi:10.1016/0370-2693(86)91275-X.

[23] G. MAHLON AND S. J. PARKE. **Spin Correlation Effects in top quark pair production at the LHC**. *Phys.Rev.D81:074024*, (2010). arXiv:1001.3422v2, doi:10.1103/PhysRevD.81.074024.

[24] B. HEGNER. **Studies on Spin-Spin-Correlations in dileptonic top pair-events at LHC**. *Department of physics, Hamburg University*, (2008). PHD thesis.

[25] ATLAS COLLABORATION. **Measurement of spin correlation in $t\bar{t}$ production from pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector**. (2011). ATLAS-CONF-2011-117.

[26] CMS COLLABORATION. **Measurement of Spin Correlations in ttbar production**. *CMS Physics Analysis Summary*, (2012). CMS-PAS-TOP-12-004.

[27] W. BERNREUTHER AND Z.-G. SI. **Distributions and correlations for top quark pair production and decay at the Tevatron and LHC**. (2010). arXiv:1003.3926v1.

[28] H. BACK ET AL. **Search for electron decay mode e+gamma neutrino with prototype of Borexino detector**. *Physics Letters B 525 (1-2): 29-40*, (2002). doi:10.1016/S0370-2693(01)01440-X.

[29] V. EZHELA, S. LUGOVSKY, AND O. ZENIN. **Hadronic Part of the Muon g-2 Estimated on the $\sigma_{tot}^{2003}(e+e- \rightarrow hadrons)$ Evaluated Data Compilation**. (2004). arXiv:hep-ph/0312114v2.

[30] CDF COLLABORATION. **Evidence for $t\bar{t}\gamma$ Production and Measurement of $\sigma_t\bar{t}\gamma/\sigma_t\bar{t}$**. *Phys.Rev.D84:031104*, (2011). arXiv:1106.3970v5,doi:10.1103/PhysRevD.84.031104.

[31] ATLAS COLLABORATION. **Measurement of the inclusive t tbar gamma cross section with the ATLAS detector**. (2011). ATLAS-CONF-2011-153.

[32] O. S. BRÜNING, P. COLLIER, P. LEBRUN, S. MYERS, R. OSTOJIC, J. POOLE, AND P. PROUDLOCK. **LHC Design Report**. *CERN, Geneva*, (2004). CERN-2004-003-V-1.

[33] **ATLAS detector and physics performance: Technical Design Report, 1**. *CERN, Geneva, Technical Design Report ATLAS*, (1999). CERN-LHCC-99-014.

[34] **ATLAS detector and physics performance: Technical Design Report, 2**. *CERN, Geneva, Technical Design Report ATLAS*, (1999). CERN-LHCC-99-015.

[35] CMS COLLABORATION. **CMS Physics Technical Design Report Volume I : Detector Performance and Software**. *CERN, Geneva, Technical Design Report CMS*, (2006). CMS-TDR-008-1.

[36] ALICE COLLABORATION. **ALICE: Physics Performance Report, Volume II**. *J. Phys. G: Nucl. Part. Phys. 32 1295*, (2006). doi:10.1088/0954-3899/32/10/001.

[37] **LHCb reoptimized detector design and performance: Technical Design Report**. *CERN, Geneva, Technical Design Report LHCb*, (2003). CERN-LHCC-2003-030.

[38] L. PERCHALLA. **Kinematic Tau Reconstruction and Search For The Higgs Boson in Hadronic Tau Pair Decays with the CMS Experiment**. *RWTH Aachen University*, (2011). PHD Thesis.

[39] CMS COLLABORATION. (2011). https://twiki.cern.ch/twiki/bin/view/CMSPublic/ LumiPublicResults2011.

[40] P. SAUERLAND. **Kinematic Reconstruction of Tau Leptons and Test for Lepton Universality in Charged Weak Interactions with the CMS Experiment**. *RWTH Aachen University*, (2011). PHD Thesis.

[41] D. CAMPI ET AL. **The CMS magnet project: Technical Design Report**. *CERN, Geneva, Technical Design Report CMS*, (1997). CMS-TDR-001.

[42] V. KARIMÄKI. **The CMS tracker system project: Technical Design Report**. *CERN, Geneva, Technical Design Report CMS*, (1997). CMS-TDR-005.

[43] CMS COLLABORATION. **CMS Tracking Performance Results from Early LHC Operation**. *Eur.Phys.J.C70:1165-1192*, (2010). arXiv:1007.1988v2, doi:10.1140/epjc/s10052-010-1491-3.

[44] LHC EXPERIMENTS COMMITTEE. **The CMS electromagnetic calorimeter project : Technical Design Report**. *CERN, Geneva, Technical Design Report CMS*, (1997). CMS-TDR-004.

[45] CMS COLLABORATION. **The CMS experiment at the CERN LHC**. *JINST 3 S08004*, (2008). doi:10.1088/1748-0221/3/08/S08004.

[46] CMS COLLABORATION. **Electromagnetic calorimeter calibration with 7 TeV data**. (2010). CMS-PAS-EGM-10-003.

[47] LHC EXPERIMENTS COMMITTEE. **The CMS hadron calorimeter project: Technical Design Report**. *CERN, Geneva, Technical Design Report CMS*, (1997). CMS-TDR-002.

[48] CMS COLLABORATION. **Missing transverse energy performance of the CMS detector**. *JINST 6 P09001*, (2011). doi:10.1088/1748-0221/6/09/P09001.

[49] CMS COLLABORATION. **Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS**. *JINST 6 (2011) 11002*, (2011). arXiv:1107.4277v1, doi:10.1088/1748-0221/6/11/P11002.

[50] LHC EXPERIMENTS COMMITTEE. **The CMS muon project: Technical Design Report**. *CERN, Geneva, Technical Design Report CMS*, (1997). CMS-TDR-003.

[51] S. CITTOLIN, A. RÁCZ, AND P. SPHICAS. **CMS trigger and data-acquisition project: Technical Design Report**. *CERN, Geneva, Technical Design Report CMS*, (2002). CMS-TDR-006-add-2.

[52] J. ALWALL AND OTHERS. **MadGraph/MadEvent v4: the new web generation**. *JHEP0709:028*, (2007). arXiv:0706.2334, doi:10.1088/1126-6708/2007/09/028.

[53] P. M. NADOLSKY ET AL. **Implications of CTEQ global analysis for collider observables**. *Phys. Rev. D 78, 013004*, (2008). arXiv:0802.0007v3, doi:10.1103/PhysRevD.78.013004.

[54] Z. WAS. **Precision simulations with TAUOLA and PHOTOS**. *Nucl.Phys.Proc.Suppl.169:16-21*, (2007). arXiv:hep-ph/0610386v2doi:10.1016/j.nuclphysbps.2007.02.113.

[55] S. ALIOLI, P. NASON, C. OLEARI, AND E. RE. **NLO single-top produc-
tion matched with shower in POWHEG: s- and t-channel contributions**.
JHEP 0909:111, (2009). arXiv:0907.4076v2, doi:10.1088/1126-6708/2009/09/111,
http://powhegbox.mib.infn.it/.

[56] E. RE. **Single-top Wt-channel production matched with parton showers us-
ing the POWHEG method**. *Eur.Phys.J.C71:1547*, (2011). arXiv:1009.2450v1,
doi:10.1140/epjc/s10052-011-1547-z, http://powhegbox.mib.infn.it/.

[57] T. SJÖSTRAND, S. MRENNA, AND P. SKANDS. **PYTHIA 6.4 physics and
manual**. *JHEP 0605:026*, (2006). arXiv:hep-ph/0603175v2, doi:10.1088/1126-
6708/2006/05/026.

[58] M. L. MANGANO, M. MORETTI, F. PICCININI, AND M. TRECCAN. **Matching
matrix elements and shower evolution for top-pair production in hadronic col-
lisions**. *JHEP 0701:013*, (2007). arXiv:hep-ph/0611129v1, doi:10.1088/1126-
6708/2007/01/013.

[59] R. FIELD. **Early LHC Underlying Event Data - Findings and Surprises** [1] *Invited
talk at HCP2010, Toronto*, (2010). arXiv:1010.3558v1.

[60] J. ALLISON ET AL. **Geant4 developments and applications**. *IEEE Trans. Nucl.
Sci. 53 (2006) 270*, (2006). doi:10.1109/TNS.2006.869826.

[61] J. M. CAMPBELL AND R. ELLIS. **MCFM for the Tevatron
and the LHC**. *Nucl.Phys.Proc.Suppl.205-206:10-15*, (2010).
doi:10.1016/j.nuclphysbps.2010.08.01, arXiv:1007.3492v1.

[62] K. MELNIKOV AND F. PETRIELLO. **Electroweak gauge boson production at
hadron colliders through** $O(\alpha_s^2)$. *Phys.Rev.D74:114017*, (2006). arXiv:hep-
ph/0609070v1, doi:10.1103/PhysRevD.74.114017.

[63] CMS COLLABORATION. **Particle-Flow Event Reconstruction in CMS and Per-
formance for Jets, Taus, and MET**. (2009). CMS-PAS-PFT-09-001.

[64] CMS COLLABORATION. **Tracking and Primary Vertex Results in First 7 TeV
Collisions**. (2010). CMS-PAS-TRK-10-005.

[65] CMS COLLABORATION. **Performance of muon identification in pp collisions at**
$\sqrt{s} = 7$ **TeV**. (2010). CMS-PAS-MUO-10-002.

[66] S. BAFFIONI ET AL. **Electron reconstruction in CMS**. *Eur. Phys. J. C49 (2007)
1099-1116*, (2007). doi: 10.1140/epjc/s10052-006-0175-5.

[67] M. CACCIARI, G. P. SALAM, AND G. SOYEZ. **The anti-k_t jet clustering
algorithm**. *JHEP 0804:063*, (2008). arXiv:0802.1189v2, doi:10.1088/1126-
6708/2008/04/063.

---

[1] Tune Z2 is identical to the tune Z1 described in [59] except for the PDF. Z2 uses CTEQ6L [53]
while Z1 uses CTEQ5L.

[68] CMS COLLABORATION. **Jet Performance in pp Collisions at** $\sqrt{s} = 7$ **TeV**. (2010). CMS-PAS-JME-10-003.

[69] CMS COLLABORATION. **b-Jet Identification in the CMS Experiment**. (2012). CMS-PAS-BTV-11-004.

[70] A. HERTEN. **Bestimmung der Masse des Top-Quarks anhand der Zerfallslängen von B-Hadronen im CMS-Experiment**. *Diploma Thesis, RWTH Aaachen University*, (2010).

[71] J. LANGE. **b-tagging correction factors for events with at least two b jets**. *Private Communication*, (2011).

[72] TEVATRON ELECTROWEAK WORKING GROUP FROM CDF AND D0 COLLABORATIONS. **Combination of CDF and DO results on the mass of the top quark using up to 5.8˜fb-1 of data**. *FERMILAB-TM-2504-E*, (2011). arXiv:1107.5255v3.

[73] CMS COLLABORATION. **First measurement of** $B(t - Wb)/B(t - Wq)$ **in the dilepton channel in pp collisions at sqrt(s)=7 TeV**. (2012). CMS-PAS-TOP-11-029.

[74] G. J. BARKER. **b-Quark Physics with the LEP Collider: The Development of Experimental Techniques for b-Quark Studies from Z0-Decay**. *Springer Tracts in Modern Physics, Volume 236*, (2010). doi:10.1007/978-3-642-05279-8.

[75] Z. G. UNALAN. **Measurement of the top quark charge**. (2007). PHD thesis.

[76] D. CASADEI. **Estimating the selection efficiency**. (2012). arXiv:0908.0130v7.

[77] B. ANDERSSON, G. GUSTAFSON, G. INGELMAN, AND T. SJÖSTRAND. **Parton fragmentation and string dynamics**. *Physics Reports, Volume 97, Issue 2-3, p. 31-145*, (1983). doi:10.1016/0370-1573(83)90080-7.

[78] M. L. MANGANO, M. MORETTI, F. PICCININI, AND M. TRECCANI. **Matching matrix elements and shower evolution for top- quark production in hadronic collisions**. *JHEP 0701:013*, (2007). doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129v1.

[79] CDF COLLABORATION. **The CDF Measurement of the Top Quark Charge using the Top Decay Products in Lepton+Jet channel**. *CDF note 10460*, (2011).

[80] CDF COLLABORATION. **Evidence for** $t\bar{t}\gamma$ **Production and Measurement of** $\sigma_{t\bar{t}\gamma}/\sigma_{t\bar{t}}$. *Phys.Rev.D84:031104*, (2011). arXiv:1106.3970v5, doi:10.1103/PhysRevD.84.031104.

[81] ATLAS COLLABORATION. **Measurement of the inclusive** $t\bar{t}\gamma$ **cross section at** $\sqrt{s} = 7$ **TeV with the ATLAS detector**. (2011). ATLAS-CONF-2011-153.

[82] H. THOLEN. **Cross Section Measurement of Top Pairs Radiating Photons**. *Master Thesis, RWTH Aachen University*, (2012).

Acknowledgement

Yvonne Küssel

Angaben zur Person

Beruf	Diplom-Physikerin
Familienstand	ledig
Staatsangehörigkeit	deutsch
Geburtsdatum	29. August 1985
Geburtsort	Heidelberg

Beruflicher Werdegang

Mär 2010 - Feb 2013	**Promotion**, Stipendium der RWTH Aachen University Doktorarbeit: *Constraints on the top quarks charge with the CMS experiment*
Seit Dez 2009	**Wissenschaftliche Mitarbeiterin** III. Physikalisches Institut B, RWTH Aachen University

Hochschulstudium

Okt 2004 - Nov 2009	**Physikdiplom**, RWTH Aachen University Diplomarbeit: *Massenbestimmung von Top-Quarks bei semileptonischen Top-Paar-Zerfällen mit dem CMS-Detektor*
Feb 2007 - Jul 2007	**Erasmussemester**, Universitat de Valencia, Spanien

Schulausbildung

Sep 1995 - Jun 2004	**Abitur** , Kurpfalz-Gymnasium Schriesheim