Differential Cross Sections for Top-Quark-Pair Production in the e/μ+Jets Final State at $\sqrt{s} = 8$ TeV in CMS

Dissertation
zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von
Martin Görner
aus Erfurt

Hamburg
2014

Gutachter der Dissertation:	Prof. Dr. Peter Schleper PD Dr. Olaf Behnke
Gutachter der Disputation:	Jun.-Prof. Dr. Christian Sander PD Dr. Andreas Meyer
Datum der Disputation:	22.08.2014
Vorsitzender des Prüfungsausschusses:	Dr. Georg Steinbrück
Vorsitzender des Promotionsausschusses:	Prof. Dr. Daniela Pfannkuche
Leiter des Fachbereichs Physik:	Prof. Dr. Peter Hauschildt
Dekan der Fakultät für Mathematik, Informatik und Naturwissenschaften:	Prof. Dr. Heinrich Graener

Abstract

The large number of top quarks produced in proton-proton (pp) collisions at the Large Hadron Collider facilitates precision measurements of top-quark properties that are crucial for testing the consistency of the Standard Model (SM) of particle physics. In particular, differential measurements of the top-quark pair ($t\bar{t}$) production cross section are essential for comparison with state-of-the-art predictions of Quantum Chromodynamics (QCD) within the SM and for searches for new physics phenomena beyond it. In this thesis, normalised differential $t\bar{t}$ cross sections are measured at a pp centre-of-mass energy of $\sqrt{s} = 8$ TeV. The analysed dataset is recorded by the CMS experiment and corresponds to an integrated luminosity of 19.7 fb^{-1}. The $t\bar{t}$ production cross section is measured differentially as a function of eighteen kinematic quantities of the top quarks and their decay products. The achieved precision of typically $3 - 7\,\%$ allows to distinguish between different QCD predictions for several of the analysed quantities. In particular, the measured $t\bar{t}$ production cross section as a function of the top-quark transverse momentum and its impact on different top-quark related measurements, also outside the scope of this work, is discussed in detail. In addition, the top-quark mass is determined from the normalised differential cross section as a function of the invariant mass of the $t\bar{t}$ system and one additional jet. The obtained result is compatible with other measurements, including the world-average top-quark mass, within the achieved precision of about 1.5 %.

Zusammenfassung

Die große Anzahl an Top-Quarks, die in Proton-Proton (pp) Kollisionen am Large Hadron Collider erzeugt werden, ermöglicht es, die Eigenschaften des Top-Quarks detaillierter zu untersuchen als jemals zuvor und damit die Konsistenz des Standardmodells (SM) der Teilchenphysik zu testen. Insbesondere differentielle Messungen des Wirkungsquerschnittes zur Produktion von Top-Quark-Paaren ($t\bar{t}$) sind hierbei relevant – zum einen für Vergleiche mit aktuellen Vorhersagen der Quantenchromodynamik (QCD) innerhalb des SM und zum anderen für Suchen nach Physik jenseits davon. In der vorliegenden Arbeit werden achtzehn normierte differentielle Wirkungsquerschnitte zur $t\bar{t}$-Produktion bei einer pp-Schwerpunktsenergie von $\sqrt{s} = 8$ TeV gemessen. Der analysierte Datensatz wurde vom CMS-Experiment aufgezeichnet und entspricht einer integrierten Luminosität von 19.7 fb^{-1}. Der $t\bar{t}$-Produktionswirkungsquerschnitt wird differentiell als Funktion von achtzehn kinematischen Größen der Top-Quarks und ihrer Zerfallsprodukte gemessen. Die typische Präzision von $3 - 7\,\%$ erlaubt es, für verschiedene Größen zwischen unterschiedlichen QCD-Vorhersagen zu unterscheiden. Insbesondere die gemessene Verteilung des Transversalimpulses der Top-Quarks und der Einfluss auf andere Messungen im Zusammenhang mit Top-Quark-Paaren, auch außerhalb dieser Arbeit, werden im Detail diskutiert. Weiterhin wird die Masse des Top-Quarks aus dem normierten differentiellen Wirkungsquerschnitt als Funktion der invarianten Masse des $t\bar{t}$-Systems und eines zusätzlichen Jets bestimmt. Die gemessene Top-Quark-Masse ist innerhalb der erreichten Genauigkeit von etwa 1.5 % mit anderen Messungen und dem Weltdurchschnittswert verträglich.

Contents

Chapter 1

Introduction

Humankind has arrived at a very high level of knowledge, skills and possibilities. One of the driving forces behind this overwhelmingly successful evolution has always been curiosity. What are the principles of nature? What are we made of? Where do we come from? These are some of the essential questions – born by curiosity, motivating science and triggering progress.

Particle physics can be seen in exactly this tradition. Trying to answer fundamental questions like "What are the constituents of matter?" and "How do they interact and build up the world as we know it?" has brought modern particle physics to the edge of knowledge and technology. The challenge of particle physics can be compared to a big puzzle game. Standing within a pool of millions of pieces, scientists need a lot of creativity and endurance as well as a bit of luck from time to time to make progress. From the particle physics point of view, high energy experiments are a successful and powerful tool to find new pieces of the puzzle, identify fitting parts and test ideas of a larger picture.
"Daß ich erkenne, was die Welt Im Innersten zusammenhält" [1] was the maxim of Johann Wolfgang von Goethes Faust already more than 200 years ago, summarising nicely the topic of modern particle physics.

The knowledge collected in particle-physics experiments until now manifests itself in the so-called Standard Model (SM) which is able to describe a large part of nature with a fairly small number of fundamental building blocks and principles. Nevertheless, the Standard Model itself cannot be the final answer as several key questions remain unanswered. How does gravity fit into this picture? What is the gravitationally interacting dark matter that we know from astronomy? What is this mysterious dark energy that leads to an accelerating expansion of the universe? All these questions remain unsolved in the SM, suggesting that it is only a step forward but not the final destination.

The latest high energy physics experiments at the Large Hadron Collider (LHC) at CERN, Geneva, were able to extend the SM by another missing piece – the Higgs boson. This could answer the question for the origin of particle masses, embedding the experimental fact that elementary particles possess mass in the theoretical concept of the SM through the so-called "Higgs Mechanism". The discovery of the Higgs boson

was doubtlessly the largest success of the LHC so far. Recognised in media around the world, the eyes of the world were focused at CERN for at least one moment. The discovery of the Higgs Boson facilitated the awarding of the physics nobel prize 2013 to François Englert [2] and Peter W. Higgs [3] *"for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERNs Large Hadron Collider."* [4]

A bit outside the public focus, the LHC produced an incredible number of several thousand publications with many new results, measurements with to date unmatched precision, or world's best exclusion limits for new physics theories. These results are highly relevant to test the SM in all possible ways. It is widely accepted that the SM is not the final word in particle physics. At least an extension or modification is needed. Maybe the SM will reveal itself as an effective theory which is valid only at low energies much smaller than the Planck scale of $\mathcal{O}(10^{19}\,\mathrm{GeV}^{\mathrm{a}})$. The challenge for theoretical particle physics is to develop an improved theory that is able to reproduce the various high precision predictions of the SM.
Several promising concepts have been developed, ranging from Supersymmetry (SUSY) [5] and extra dimensions [6] up to string theory [7, 8]. However, no final answer has been found yet. Therefore, it would help to get some inspiration from experimental particle physics. This could in principle happen in two ways. One possibility is to discover a signature beyond the SM (BSM), either predicted by a new theory (e.g. SUSY) or something completely unexpected (e.g. a new fundamental particle). Another possibility is to probe the SM predictions with very high precision and to identify their limitations. The LHC offers the chance to follow both strategies. With its large centre-of-mass energy, the high integrated luminosity and the high quality of the recorded data, the LHC facilitates the research of the fundamental building blocks of nature, their properties and dynamic interplay in unmatched precision and variety.

One of these fundamental building blocks is to our best knowledge the top quark. Being not incorporated in stable matter, the top quark can only be studied in high energy physics experiments. Understanding top-quark physics is highly relevant in many aspects. Through its fast decay, the top quark gives direct access to the bare quark and its dynamics which within the SM is described by the theory of Quantum Chromodynamics (QCD). Therefore, investigating top-quark production in detail is a precision test for perturbative QCD calculations. The experimental results might give evidence for higher order effects as well as to potential new physics effects modifying the production process. One possible scenario is a hypothetical new heavy resonance [9] decaying into top-quark pairs.
With the discovery of the Higgs boson, the top quark has yet another key role. Due to its large mass of approximately 173 GeV the top quark is expected to possess the largest Yukawa coupling to the Higgs boson. Testing the interplay of Higgs boson and top quark is therefore one of the most important ingredients to reveal whether the

[a] Natural units $c = \hbar = \epsilon_0 = 1$ are used in this thesis. Consequently, energy, mass and momentum possess the dimension of energy and are expressed in units of electron Volts (eV) while the electric charge is given in units of the elementary charge e.

discovered Higgs boson is really the SM Higgs boson or e.g. just one out of several within a BSM scenario. Furthermore, the precise measurement of the top-quark mass provides insight to such fundamental questions as the stability of the electroweak vacuum [10] and therefore the fate of our universe.
Hence, knowing the properties of the top quark with high precision is important to improve our understanding about the essence of nature and is a direct gateway to potential new physics phenomena.

Within the last years, top-quark physics has seen a drastic change. Discovered in 1995 [11, 12] and previously produced only in the Tevatron experiments at Fermilab (USA), the top quark was rediscovered by the LHC already with the first data. Subsequently, the production of top-quark pairs [13] as well as of single top quarks [14,15] has been observed. A large dataset of about 850,000 (4.8M) *top-quark pair events* ($\mathrm{t\bar{t}}$) at $\sqrt{\mathrm{s}} = 7$ TeV (8 TeV) has been recorded by the ATLAS and CMS collaboration respectively. It allows a detailed study of the top quark, its production mechanism and properties. The LHC is therefore a real *Top Factory*, facilitating an important step towards precision measurements in the top-quark sector, which allow to push back the boundaries of knowledge in this field of particle physics further than ever before.

The large available LHC data sample offers the possibility to investigate the $\mathrm{t\bar{t}}$ production cross section differentially as a function of many top-quark related kinematic quantities. Such measurements are precision tests for perturbative QCD predictions, are sensitive to potential new physics phenomena and can be used to extract fundamental parameters as the top-quark mass ($\mathrm{m_{top}}$) or the strong coupling constant (α_S). Moreover, they are useful input in electroweak fits and to improve the precision of Parton Distribution Functions.
In the past, first differential $\mathrm{t\bar{t}}$ cross sections have been measured at the Tevatron by CDF [16, 17] and D0 [18, 19] with large, predominantly statistical uncertainties of typically 20% and a limited set of quantities. After pioneering studies with early LHC data [20,21] (also within this thesis project), first normalised differential $\mathrm{t\bar{t}}$ cross sections are measured by the ATLAS [22] and CMS [23–27] collaborations (also within this work) using the LHC datasets at $\sqrt{\mathrm{s}} = 7$ TeV and 8 TeV. In contrast to the Tevatron results, these measurements are performed for an extended list of quantities and are limited predominantly by systematic uncertainties, achieving a typical precision of $5 - 10\%$.

In this work, semileptonically decaying top quarks with one isolated lepton and at least four jets out of which two jets are associated to b quarks are analysed. Within the work of this thesis, a strategy for differential cross section measurements is developed, contributing largely to first measurements with $\sqrt{\mathrm{s}} = 7$ TeV CMS data, which are published in [25, 28]. Furthermore, a preliminary result using approximately half of the CMS dataset at $\sqrt{\mathrm{s}} = 8$ TeV, which is based on this work, is published in [26]. Improving and extending these previous measurements, the analysis presented in this thesis focuses on the full 2012 LHC proton-proton (pp) collision dataset at a centre-of-mass energy of $\sqrt{\mathrm{s}} = 8$ TeV recorded by the CMS experiment. This dataset corresponds to an integrated luminosity of 19.7 fb^{-1}. The large integrated luminosity and the high quality of the data facilitate a differential cross section analysis with an improved precision and to date unmatched richness of detail. Normalised differential $\mathrm{t\bar{t}}$ cross sections $\frac{1}{\sigma}\frac{\mathrm{d}\sigma}{\mathrm{dX}}$ are

measured as a function of eighteen different kinematic quantities X related to single top quarks, the $t\bar{t}$ system and the observed final-state leptons and jets with a typical precision of $3-7\%$. An overview of all cross section quantities and the corresponding phase space of the measurements is illustrated in Table 1.1.
In addition to the cross section measurements, the determination of the top-quark mass from normalised differential cross sections is studied.

Object	**Kinematic Quantities**	**Phase Space of the Measurement**
$t\bar{t}$ System Top Quarks	$p_T^{t\bar{t}}$, $y^{t\bar{t}}$, $m^{t\bar{t}}$ p_T^t, p_T^t ($t\bar{t}$ com), $p_T^{\text{lead t}}$, $p_T^{\text{sublead t}}$, y^t, $\Delta\phi\,(t,\bar{t}) = \phi^t - \phi^{\bar{t}}$	extrapolated parton level
Lepton (e/μ) Jets B Jets $b\bar{b}$ System Lepton-B Jet System $t\bar{t}$ + Additional Jet System	p_T^l, η^l N_{jets} p_T^b, η^b $m^{b\bar{b}}$, $p_T^{b\bar{b}}$ m^{lb} $\rho_S = \frac{2\cdot 170\,\text{GeV}}{m_{t\bar{t}j}}$	visible particle level

Table 1.1: List of all objects and their kinematic quantities for which normalised differential cross sections are determined

The structure of this thesis is as follows: An overview of the SM is given in Chapter 2 and an introduction to top-quark physics in Chapter 3, followed by a description of the experimental setup of the LHC experiment and CMS detector in Chapter 4. Details on the analysis dataset and simulation tools are listed in Chapter 5 and object definitions, data driven corrections and the event selection to obtain a high purity $t\bar{t}$ sample are detailed in Chapter 6. Subsequently, the kinematic reconstruction of the top quarks from the measured final-state objects with a constrainted kinematic fitting procedure is discussed in Chapter 7. In Chapter 8 the calculation of the cross sections and the definition of the visible particle level phase space and the extrapolated parton level phase space is discussed. Furthermore, the applied regularised unfolding technique that is used to correct for efficiency, acceptance and migration effects is introduced and studied in Chapter 9. All systematic uncertainties are discussed in Chapter 10 and the cross section results are presented in Chapter 11. The transverse momentum of the top quarks reveals itself as not perfectly modelled by all SM predictions and is therefore investigated in detail in Chapter 12. Using a measured normalised differential cross section as input, the extraction of the top-quark mass is studied in Chapter 13. Finally, conclusions are drawn and an outlook is given in Chapter 14.

Chapter 2

The Standard Model of Particle Physics

In our understanding of nature to date, matter is composed of fundamental building blocks. This idea dates back several thousand years to ancient philosophers but is still valid and the key idea of modern particle physics. The interaction between these fundamental building blocks of nature forms the world as we know it. Build up over time and recently extended by the discovery of the Higgs boson, the Standard Model (SM) of particle physics reflects humankinds best approach to describe the structure and dynamics of nature. A short overview will be given in this chapter while a more detailed introduction of the SM can be found e.g. in [23, 29–33].

2.1 Phenomenological Overview and Particle Content

The key idea of the SM is that elementary particles are the fundamental building blocks of nature. They are indivisible, point-like, structureless objects, characterised by their intrinsic properties: mass, spin and the so-called quantum numbers, which describe the dynamics of each particle. The SM is based on fermions with spin $\frac{1}{2}$ and bosons with integer spin.

The twelve fermions are ordered by their mass into three generations. Within each generation, one distinguishes two quarks and the two leptons. Quarks are mainly characterised by their (in contrast to the leptons) non-vanishing *colour charge* C. Three linear independent colour states exist i.e. referred to as red, blue, green. Another quantum number that characterises the particles is the *elementary electrical charge* (Q). Quarks can have $\mathrm{Q} = +\frac{2}{3}$ (up-type: up quark u, charm quark c, top quark t) or $\mathrm{Q} = -\frac{1}{3}$ (down-type: down quark, strange quark, bottom quark). Leptons can have $\mathrm{Q} = +1$ (electron, muon, tauon) or $\mathrm{Q} = 0$ (neutrinos). In addition, an antiparticle exist for each particle. Each antiparticle is characterised by inverted quantum numbers (e.g. $\mathrm{Q}_{\mathrm{antiparticle}} = -1 \cdot \mathrm{Q}_{\mathrm{particle}}$) but besides this identical properties (e.g. mass, spin) with

respect to the particle as listed above. For particles with $C \neq 0$, the colour charge of the corresponding antiparticle is give by the corresponding anticolour (antired, antiblue and antigreen).

The elementary particles[a] are dynamic objects which can interact with each other depending on their quantum numbers. Within these interactions, particles can be converted into different particles or form stable states of matter. These processes follow several rules, e.g. energy, spin, momentum and charge conservation. The dynamics of each particle is described by its properties. The lightest particles are stable while single particles with larger masses are only short-lived. Therefore, the fermions of the first generation are stable. In a simplified picture, they represent the tool kit to build up stable matter. Up- and down-quarks form the proton (p=uud) and the neutron (n=udd). Together with the electron, they are the building blocks of all atoms which form more complex objects.
All particles of the SM, their spin, electrical charge, mass and interactions, which will be detailed in the following, are summarised in Figure 2.1.

The SM describes three fundamental interactions: the electromagnetic, the weak and the strong interaction. These interactions are described by the exchange of spin 1 gauge bosons. The strengths of the interactions depend on the related quantum number of the participating particles, i.e. the specific charges g. Each interaction refers to one specific quantum number.

All particles with an electrical charge ($Q \neq 0$) participate in the *electromagnetic* (EM) interaction which is mediated by the massless and electrically neutral photon. The EM interaction describes e.g. the dynamics of atoms.

The weak interaction is mediated by the heavy W and Z bosons and couples to all particles carrying *weak charge*, i.e. all leptons, quarks and also the W and Z bosons themselves. The W boson exists in two stages with electrical charge $Q = +1$ and $Q = -1$. In contrast, the Z boson carries no electrical charge ($Q = 0$). The weak interaction describes e.g. the decay of free neutrons. Its strength is suppressed by the relatively large masses of the W and Z bosons ($m_Z = 91.2\,\mathrm{GeV}$, $m_W = 80.4\,\mathrm{GeV} \approx 157,000 \cdot m_e$) for energy scales below these masses.

Furthermore, the weak interaction violates parity maximally. This means that its behaviour is not symmetric under spatial point-reflections defined by the chiral symmetry. The W bosons couples only to left-chiral particles and right-chiral antiparticles. The Z boson couples with different strengths to left-chiral and right-chiral particles. Chirality is a fundamental property of the elementary particles. The two different states of *chirality* are for massless particles equal to the helicity, i.e. the projection of the spin onto the momentum.

Another special property of the weak interaction is the coupling of the W bosons to different fermion (flavour) types. The W-bosons couple either to a neutrino and a charged lepton (e.g. $W \to e\nu_e$) or to an up-type and a down-type quark (e.g. $W^- \to \bar{u}d$).

[a] In this thesis, the expression particles refers to particles as well as antiparticles if not specified differently.

Usually, this involves only the flavours of the same generation. For quarks, also flavour-changing couplings involving more than one generation are observed (e.g. $W^- \rightarrow \bar{c}d$). Mass and flavour eigenstates of quarks are not equal. The relation and therefore also the probability for the mixing of different flavours is described by the *Cabibbo-Kobayashi-Maskawa* (CKM) matrix:

$$\mathrm{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{2.1}$$

In general, the CKM matrix can be parametrised by three mixing angles and one

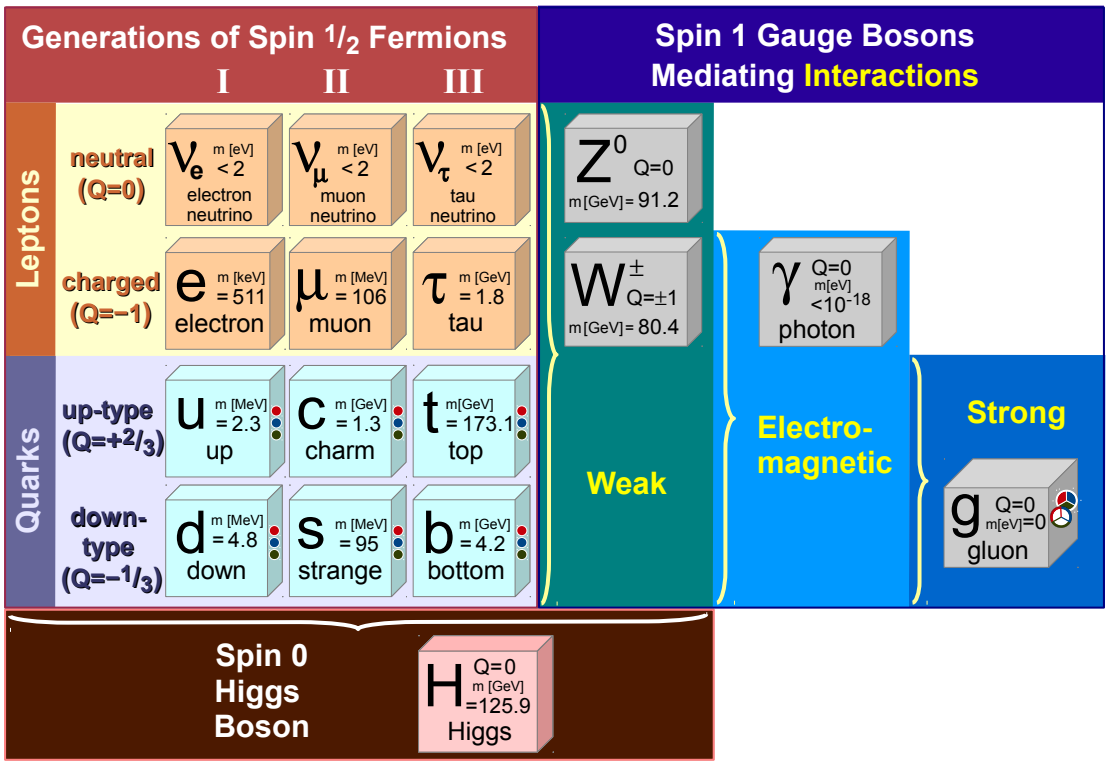

Figure 2.1: Overview of the elementary particles of the SM. The listed properties are their electrical charge (Q), Spin (S) and measured mass (m). The possible colour states of quarks and gluon are illustrated with coloured dots. The measured masses are taken from [32]. The limits on the neutrino masses are concluded from the most precise direct measurement for the electron neutrino and the negligible small mass differences measured in neutrino oscillation experiments. The gluon mass is a theoretical value. The curly brackets indicate the participating particles for the interactions and the coupling to the Higgs boson.

(CP[b]-violating) phase. The magnitudes of the single CKM elements are given by:

$$\mathrm{CKM} = \begin{pmatrix} 0.97427 \pm 0.00015 & 0.22534 \pm 0.00065 & 0.00351^{+0.00015}_{-0.00014} \\ 0.22520 \pm 0.00065 & 0.97344 \pm 0.00016 & 0.0412^{+0.0011}_{-0.0005} \\ 0.00867^{+0.00029}_{-0.00031} & 0.0404^{+0.0011}_{-0.0005} & 0.999146^{+0.000021}_{-0.000046} \end{pmatrix}. \tag{2.2}$$

The quoted values are taken from a SM fit to all available measurements related to the CKM matrix [32]. Large diagonal entries and small but non-zero off-diagonal entries imply the preference of flavour-changing processes within one generation. Especially the mixing of the third generation with the other two generations is suppressed.
The weak and the electromagnetic interaction can be described in the SM within a unified electroweak theory.

The strong interaction between all particles carrying colour charge is mediated by the massless gluon. Besides the quarks (antiquarks) which carry colour (anticolour), the gluon itself carries colour and anticolour and exists in eight different colour states.

At high energies (=small distances), quarks and gluons behave like free particles (*asymptotic freedom*). In contrast, the potential energy of the strong interaction field increases for low energies (=large distances) until the stored energy is large enough to create new particles out of the vacuum. Consequently, bare quarks and gluons cannot be observed. The phenomena of *confinement* leads to the formation of colour neutral states composed of several quarks (hadrons). In analogy to chromatics, colour neutral states are built by combining either a quarks and an antiquark with the respective anticolour or several quarks or antiquarks containing all three colours or anticolours. Hence, free quarks or gluons will generate new colour charged objects until finally only colour neutral hadrons remain. This process is called *hadronisation*. Hadrons composed of three quarks are called baryons and hadrons composed of quark and antiquark are called mesons. The most prominent example for a stable baryon is the proton.

In detail, the structure of hadrons is more complex. The dynamic between the two or three *valence quarks* results in a constant gluon exchange between them. These gluons can produce themselves gluons or virtual pairs of quark and antiquark (*sea quarks*) which will finally annihilate and be reabsorbed by the valence quarks. Each gluon and sea quark carries only a small amount of the total energy of the proton in comparison to the valence quarks. The exact distribution of the proton energy to the underlying elementary particles, which are in this context often called *partons*, is described by the *Parton Distribution Functions* (PDFs). The PDFs are discussed in more detail in the context of top-quark pair production at the LHC (Chapter 3.1).

In contrast to the gauge bosons, the Higgs boson is a spin 0 particle. Being the latest discovered particle of the SM [34,35], it is the physical manifestation of the omnipresent Higgs field that explains the masses of all elementary particles via the Higgs mechanism (see Chapter 2.4). For fermions, the coupling strength for the interaction with the Higgs boson scales linearly with the particle mass. Therefore, the coupling strength of the top quark to the Higgs boson is the largest of all known SM particles.

[b]The CP symmetry is defined as invariance under the combined charge and parity transformation.

The discovery of the beforehand only theoretically motivated and predicted Higgs boson is the latest success of the SM. It complements the extraordinary good description of nature by the SM which has been experimentally verified in many measurements and with high precision.

Nevertheless, the SM has its limitations as it cannot describe everything (yet). Gravity, the fourth fundamental interaction known in nature, is currently not included in the SM. Due to the size of the observed particle masses, the expected strength of gravitational interactions between fundamental particles is negligible in comparison to the other interactions and can be safely neglected at the currently achievable energy scales. Moreover the mysterious dark matter which is observed in astronomy, e.g. via rotation curves of stars, gravitational lensing effects or the movement of galaxy systems, is not explained by the SM. Also other fundamental questions like the origin of the mass hierarchy remain unanswered. As this thesis focuses on top-quark physics within the SM, the limitations of the SM will not be detailed further.

2.2 Lagrangian and Gauge Invariance

The mathematical principles of the SM are quantum field theories. Within this formulation, all fermions are represented by quantised spinor fields and all gauge bosons by quantised vector fields. All information is encoded in the *Lagrangian density* ($\mathcal{L}$), which is a scalar function of all particle fields. In analogy to classical mechanics, the equations of motion are derived assuming a stationary action ($\delta \mathrm{A} = 0$), where the action A is defined as the time integrated Lagrangian density:

$$\mathrm{A} = \int \mathcal{L} \, \mathrm{dt}. \tag{2.3}$$

Interactions are introduced by the principle of local gauge invariance of $\mathcal{L}$ under specific gauge transformations. These gauge transformations (U) are unitary local transformations for the fermion spinors (Ψ):

$$\Psi \rightarrow \Psi' = \mathrm{U}\Psi. \tag{2.4}$$

Local gauge invariance means that the action $\mathrm{A}(\mathcal{L})$ remains unchanged for the transformed fields Ψ' for every space-time point x:

$$\int \mathcal{L}\left(\mathrm{U(x)}\Psi\right) \mathrm{dt} = \int \mathcal{L}\left(\Psi\right) \mathrm{dt}. \tag{2.5}$$

This is the case if the Lagrangian density remains unchanged up to a total derivative. Fulfilling these requirement leads to the introduction of additional terms in $\mathcal{L}$ which are interpreted as interaction terms. The choice of U leads to the different interactions. The coupling constants α for each interaction is related to the corresponding charge g (see Chapter 2.1) of the interactions:

$$\alpha \propto \mathrm{g}^2. \tag{2.6}$$

Finally, the dynamics of the SM arises from the postulation of local gauge invariance under U(1)⊗SU(2)⊗SU(3) transformations.

Within the $\mathrm{U(1)_Y}$⊗$\mathrm{SU(2)_L}$ symmetry, the electromagnetic and weak interaction are described in a unified theory. The $\mathrm{SU(2)_L}$ affects only the left-chiral parts of fermion states. This leads to the observed parity-violating nature of the weak interaction. Therefore, the SM fermions are arranged in left-chiral (index "L") doublets and right-chiral (index "R") singlets of the SU(2) flavour space, e.g. for the first generation fermions:

$$\begin{pmatrix} \nu_\mathrm{e} \\ \mathrm{e} \end{pmatrix}_\mathrm{L}, \ \mathrm{e_R}, \ \begin{pmatrix} \mathrm{u} \\ \mathrm{d} \end{pmatrix}_\mathrm{L}, \ \mathrm{u_R}, \ \mathrm{d_R}. \tag{2.7}$$

The $\mathrm{U(1)_Y}$ symmetry is defined in the space of the *weak hypercharge* (Y), which is constructed from the electromagnetic charge and the third component of the weak isospin quantum number ($\mathrm{T_3}$[c]). From the combination of the $\mathrm{U(1)_Y}$ and $\mathrm{SU(2)_L}$ symmetries, all

[c] For the doublet representation of the fermions, $\mathrm{T_3}$ is $\pm\frac{1}{2}$ while $\mathrm{T_3}$ is zero for all right-chiral fermions.

properties of the weak and electromagnetic interaction are derived, e.g. the physically measurable photon and Z boson are obtained as superpositions of one gauge boson of each gauge symmetry.

Finally, the $SU(3)_C$ symmetry in the colour space (C) is described by the theory of *Quantum Chromodynamics* (QCD) and characterises the strong interaction.

It is one of the large successes of the SM that the at first sight complex and different interactions can be naturally derived from $\mathcal{L}$ by the common and simple assumption of local gauge invariance.

2.3 Perturbation Theory and Renormalisation

An important observable in particle physics is the *cross section* σ. It is a measure for the probability with which a specific physics process (e.g. $gg \to t\bar{t}$) occurs. In principle, σ can be derived interpreting the Lagrangian density. However, the exact calculation involves infinitly many contributions.

As not all contributions are equally relevant and not all contributions can be taken into account, an approximation for the calculation of σ is needed. Using perturbation theory for such calculations, the solution for σ is expanded in orders of the coupling constant α. The single terms of the expansion can be illustrated using Feynman diagrams.

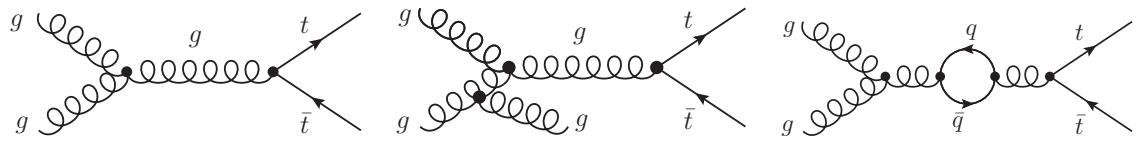

Figure 2.2: Exemplary Feynman diagrams for the process $gg \to t\bar{t}$ in the s-channel: LO contribution (tree level, left) and higher order contributions with a real emission (additional gluon from initial state radiation, middle) and a virtual loop correction ($q\bar{q}$ loop, right).

In this illustration, the leading order (LO) contribution represents the most simple process (e.g. Figure 2.2 left) while higher orders correspond to processes that involve additionally produced particles (e.g. Figure 2.2 middle) or loops (e.g. Figure 2.2 right). For $\alpha \ll 1$ the perturbative expansion of σ holds and calculating only the first terms of the expansion will give a good approximation for the result.

One conceptional problem is that splitting up the calculation in this perturbative expansion leads to divergences appearing in the calculation of separate contributions, i.e. of virtual loop contributions. These divergences would cancel with other contributions if all orders of the expansion would be taken into account. To allow the calculation of finite solutions at any order of perturbation theory, a *renormalisation procedure* is needed.

This procedure involves the introduction of a *renormalisation scale* μ_R which equals a cut-off scale for the considered momenta in the calculation of the loop contributions. The remaining divergences are absorbed in the redefined (*renormalised*) quantities, e.g. coupling constants. As a consequence of this renormalisation, the coupling constants become μ_R depended. The renormalised coupling constants $\alpha(\mu_R)$ are the measurable observables. Their value is finite and divergences occur only in the non-physical, bare parameters of the Lagrangian density. Typically, μ_R is identified with the typical energy or momentum scale of the process, e.g. the momentum transfer Q or the invariant mass m of a produced particle.

The renormalised strong coupling constant ($\alpha_S(\mu_R = Q^2)$) of QCD is given at leading

order by:

$$\alpha_S\left(Q^2\right) = \frac{12\pi}{(33-2n_f)\ln\frac{Q^2}{\Lambda^2_{QCD}}}, \tag{2.8}$$

$$\Lambda^2_{QCD} = \mu^2 \exp\left(\frac{-12\pi}{(33-2n_f)\,\alpha_S(\mu^2)}\right). \tag{2.9}$$

Here, μ denotes a reference scale and $\alpha_S(\mu)$ the value at this reference scale, e.g. $\alpha_S(m_Z^2) \approx 0.118$ [32]. Furthermore, n_f denotes the number of quark flavours contributing to the virtual loops and depends on the energy scale as only quarks with $\left(2m_q\right)^2 \leq Q^2$ contribute.

A common semi-classical interpretation of the running couplings is given by a simple picture. The physical couplings $\alpha(Q^2)$ are not predicted by the underlying theory of the SM and have to be measured. The charge determines the measured coupling strength α. Every particle can create via loop processes additional particles which carry themselves charge. This leads to a screening of the bare charge. When measuring α, the energy scale Q^2 is a measure of the probed distance. The effective charge 'seen' in the process and therefore also the measured value for α depends on the probed distance. For small differences (large Q^2) more of the bare charge is seen. Therefore, different values of α are measured for different values of Q^2.

As n_f is maximally six, α_S decreases for larger values of Q^2. For $Q^2 \to \infty$ this results in the previously discussed asymptotic freedom and the validity of perturbation theory on short scales. In contrast, α_S becomes large for $Q^2 \to \Lambda^2_{QCD}$ ($\mathcal{O}\left(\Lambda_{QCD}\right) \approx 100\,\text{MeV}$) where perturbation theory is not applicable anymore and the process of hadronisation starts.

2.4 Electroweak Symmetry Breaking and Particle Masses

Understanding the mechanism that generates the masses of the known SM particles is one of the most important issues in modern particle physics. The introduced *Higgs mechanism* [2, 3, 36] is vitally discussed as the discovery of the Higgs boson at the LHC [34, 35] allows a deeper insight than ever before.
A more detailed summary of the Higgs mechanism can be found e.g. in [32, 37] and only a short overview of the SM Higgs mechanism will be given in the following.

In the SM without Higgs mechanism, the Z and $W^{\pm}$ bosons would be massless as any other gauge boson, but their masses are measured with high precision, e.g. by the LEP experiment [38]. Simply adding mass terms for the Z and $W^{\pm}$ bosons directly to the Lagrangian density is not possible without violating gauge invariance. Therefore, the Higgs mechanism is introduced by postulating a self-interacting complex doublet of scalar fields Φ. The Higgs potential $V(\Phi)$ is given by:

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi\right)^2 . \tag{2.10}$$

The two parameters μ and λ describe the shape of the potential. For $\mu^2 < 0$ and $\lambda > 0$, the potential has the shape of a Mexican hat and the minimum energy state is not at 0 but has a *vacuum expectation value* of $v = \sqrt{-\mu^2 \cdot \lambda^{-1}}$.
This breaks the $U(1)_Y \otimes SU(2)_L$ electroweak symmetry spontaneously. The non-vanishing vacuum expectation value $v \approx 246\,\text{GeV}$ is the scale of *electroweak symmetry breaking* (EWSB).

As a consequence of the introduced Higgs field, three massless Goldstone bosons are generated, which are absorbed to give masses to the Z and $W^{\pm}$ bosons. Furthermore, the masses of the SM fermions are described by Yukawa interactions with the Higgs doublet. To explain non-zero neutrino masses as concluded from the observed neutrino oscillations [39, 40], it is possible that the Higgs mechanism plays a role but is not entirely responsible for the generation of their masses. Either other mechanisms are needed in addition or right-chiral neutrinos and left-chiral antineutrinos must exist, which cannot be observed due to the parity-violating nature of the weak interaction.

Besides the previously discussed three massless Goldstone bosons, one degree of freedom of the complex doublet Φ remains. It manifests itself in a new fundamental scalar particle, the *Higgs boson*. The mass of the Higgs boson is given as:

$$m_H = \sqrt{\frac{\lambda}{2}} v \tag{2.11}$$

and thus a free parameter of the SM. To date LHC results measure the Higgs boson mass to be 125.9 ± 0.4 GeV [32]. The coupling of the Higgs boson to other fermions is described by Yukawa interactions and the coupling strength is proportional to the fermion masses. Furthermore, the Higgs boson couples also to the SM bosons and to itself. The coupling strength between Higgs boson and other bosons V is proportional

to m_V^2. The coupling strength of the Higgs boson self coupling is proportional to m_H^2 and therefore to λ, which is consequently called the Higgs self-coupling parameter.

Chapter 3

Top-Quark Phenomenology

Among all known particles of the SM, the top quark plays a special role. The main reason is its mass of $m_{top} \approx 173\,\text{GeV}$ which is larger than for all other elementary particles currently known.

Within the SM, the existence of a third quark generation was theoretically motivated to explain CP violation in 1973 [41]. After its postulation the direct experimental discovery of the top quark took 20 years. The a priori unknown and large mass of the top quark necessitated a collider experiment with at this time unmatched high centre-of-mass energy ($\sqrt{s}$).
In 1995, the Tevatron experiments CDF [11] and D0 [12] at Fermilab (USA) managed to observe the direct production of top-quark pairs in *proton-antiproton* ($p\bar{p}$) collisions at $\sqrt{s} = 1.8\,\text{TeV}$ with a five sigma significance. After the discovery of the top quark, the Tevatron collider was the only tool allowing the study of top quarks for more than 15 years.

Since 2010, top quarks are also produced in pp collisions at the LHC. The LHC datasets recorded at a centre-of-mass energies of $\sqrt{s} = 7$ TeV (2010-2011) and $\sqrt{s} = 8$ TeV (2012) contain about $850,000$ and $5,000,000$ top-quark pairs per experiment respectively.
The analysis presented in this thesis is based on pp LHC collision data at $\sqrt{s} = 8$ TeV. An overview of top-quark phenomenology will be given in the following. A more detailed overview on top-quark physics can be found e.g. in [42, 43].

Top-quark production in hadron collider machines with focus on the production of top-quark pairs in pp collider experiments is discussed in Chapter 3.1. The decay of the top quark and the observable final states of top-quark pairs are detailed in Chapter 3.2. The top-quark mass is discussed in Chapter 3.3 while the relevance of the top quark in consideration of the Higgs discovery and physics beyond the SM is detailed in Chapter 3.4.

3.1 Production Mechanisms at Hadron Colliders

The majority of the top quarks which are studied in hadron collider experiments is produced in pairs (one top quark and one antitop quark) but also the production of single top quarks can be observed. Other processes like the production of two top-quark pairs is very rare and will not be discussed further.

3.1.1 Top-Quark Pair Production

The production of top-quark pairs in pp and $p\bar{p}$ collider experiments is characterised by the strong interaction. The initial state allows for *quark-antiquark* ($q\bar{q}$) annihilation and *gluon-gluon* (gg) fusion production modes in LO. The corresponding Feynman diagrams are shown in Figure 3.1. At NLO, top-quark pairs can also be produced from quark-gluon initial states. For higher order processes with additionally produced real quarks or gluons one differentiates between *initial state radiation* (ISR) and *final state radiation* (FSR). ISR involves the initial state partons and FSR the final state top quarks or their decay products.

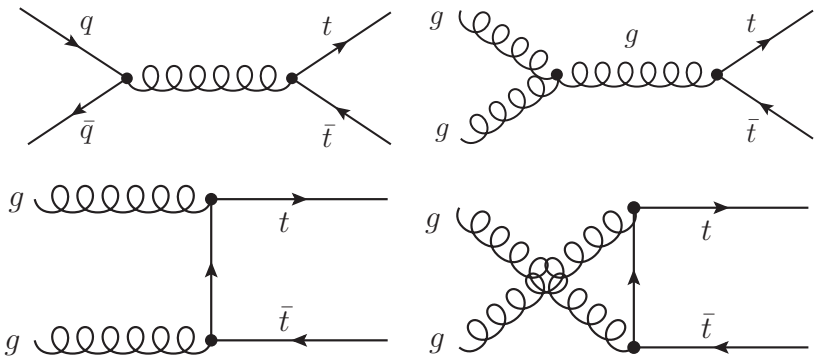

Figure 3.1: Feynman diagrams of LO top-quark pair production from the quark-antiquark initial state (top left) and from the gluon initial state (top right, bottom).

The total $t\bar{t}$ production rate for a pp collider experiment is given by the total inclusive cross section $\sigma_{pp\to t\bar{t}}$. In the regime of asymptotic freedom, $\sigma_{pp\to t\bar{t}}$ can be factorised in the partonic cross section $\hat{\sigma}_{ij\to t\bar{t}}$ and the PDFs. The partonic cross section describes the perturbatively calculated hard process (e.g. $ij = gg \to t\bar{t}$) at small distances (see Chapter 2.3) while the long-distance effects of the partons inside the hadrons are described by the PDFs. Both energy regimes are separated by the *factorisation scale* μ_F. A typical choice of μ_F is the *energy transfer of the hard process* (Q), which will be used in the following.

Using the factorisation theorem, $\sigma_{pp\to t\bar{t}}$ is given by the convolution of the partonic cross section with the PDFs $f_i(x,Q)$, involving an integration over all momentum fractions x_1 and x_2 of the two partons of the initial state and a sum over all possible

initial state partons i and j:

$$\sigma_{p_1 p_2 \to t\bar{t}} = \sum_{(i,j)\in(g,q,\bar{q})} \int_0^1 \int_0^1 \left(\hat{\sigma}_{ij\to t\bar{t}}\right) \cdot f_i^{p_1}(x_1, Q) \cdot f_j^{p_2}(x_2, Q) \cdot dx_1 dx_2. \tag{3.1}$$

In general, the total inclusive cross section $\sigma_{pp\to t\bar{t}}$ depends on the factorisation scale μ_F, the renormalisation scale μ_R (see Chapter 2.3), the strong coupling constant $\alpha_S(\mu_R)$, the partonic centre-of-mass energy $\sqrt{\hat{s}}$ and the top-quark mass (m_{top}). For $\sqrt{s} = 8$ TeV, the total inclusive cross section is in the order of $\sigma_{pp\to t\bar{t}} = 245.8$ pb [44], mainly depending on the choice of the scales (here $\mu_F = \mu_R = m_{top}$), the top-quark mass (here $m_{top} = 173.3\,\text{GeV}$), and the PDFs (here MSTW2008nnlo).

The PDFs describe the number density of gluons and all specific quark types as function of the proton momentum fraction x they carry for a given energy scale Q. In Figure 3.2, the PDFs for an energy scale in the order of top-quark production are shown. In a simplified picture, the three valence quarks of the proton (uud) carry $x = \frac{1}{3}$ of the proton momentum. The presence of additional gluons and quarks (see Chapter 2.1) leads to a more complicated situation. In processes of large energy transfers Q, more gluons and sea quarks can be resolved (see Chapter 2.3) and dominate the PDFs for low values of x. PDFs have been measured experimentally e.g. in deep-inelastic scattering processes at the electron-proton collider HERA [45].

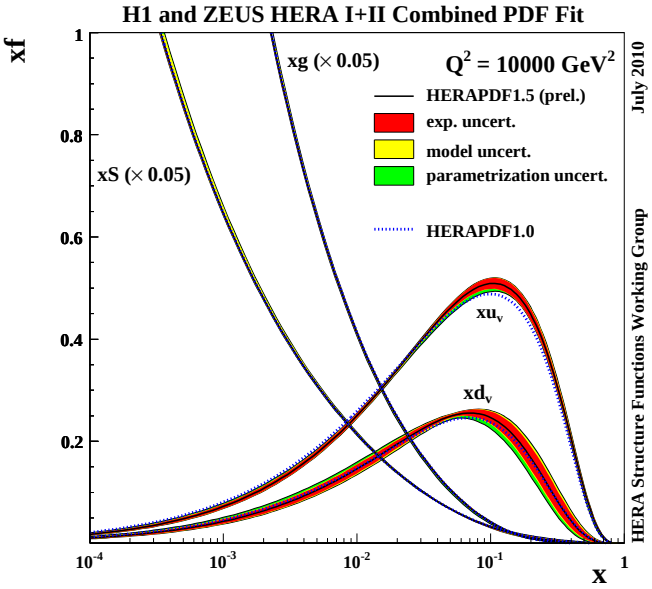

Figure 3.2: Parton Distribution Functions (PDFs) of the proton. The PDFs for the valence quarks (u_v, d_v), sea quarks (S) and the gluons (g) are shown as function of the proton momentum fraction x for an energy scale of $Q = 100\,\text{GeV}$. The values correspond to the HERAPDF1.5 NLO PDF set. The figure is taken from [46].

For pp collider experiments like the LHC, the predominant $t\bar{t}$ production mechanism is via gg fusion (about 90%). The $q\bar{q}$ annihilation production mode is suppressed in

comparison to $p\bar{p}$ collider experiments because antiquarks occur only as sea quarks.

As the top quark is the heaviest known SM particle, a higher minimal centre-of-mass energy $\sqrt{s}$ for its production is needed in comparison to other particles. The minimal energy needed for the production of a top-quark pair is $\sqrt{\hat{s}} = 2 \cdot m_{top} \approx 345\,\text{GeV}$. The centre-of-mass energy of the colliding hadrons (e.g. protons) is related to the partonic centre-of-mass energy $\sqrt{\hat{s}}$ via the proton momentum fractions x_1 and x_2 of the two initial state partons:

$$\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}. \tag{3.2}$$

3.1.2 Single Top-Quark Production

In addition to the previously discussed production of top-quark pairs, also single top quarks can be produced in hadron colliders. The production cross section for these processes is lower because the production mechanism involves the weak interaction, i.e. Wtb vertices. The possible LO Feynman diagrams (s-channel, t-channel and tW-channel) are shown in Figure 3.3.

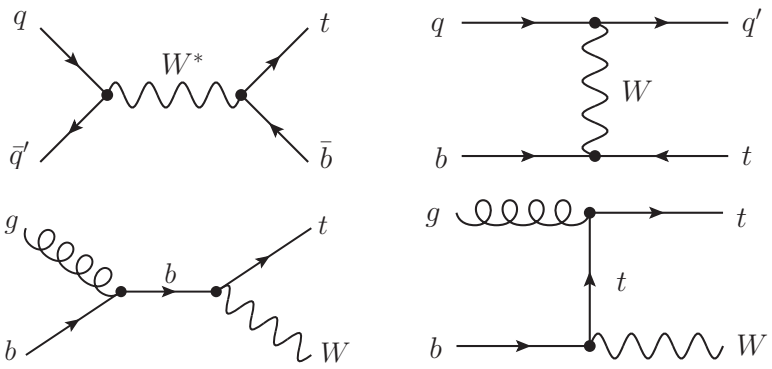

Figure 3.3: Feynman diagrams of the LO production of single top quarks in hadron colliders: s-channel (top,right), t-channel (top,left) and tW-channel (bottom).

In the t-channel, the top quark is created by a fusion of a b quark and a virtual W boson. This channel is the predominant production mode for single top quarks at the LHC with $\sqrt{s} = 8$ TeV. The charge of the initial state quark determines if a top quark or an antitop quark is produced. For a pp collider, up quarks dominate the PDFs of the valence quark. Therefore, the production of top quarks ($\sigma_{pp\to t,t-channel} \approx 56.4$ pb) is preferred over the production of antitop quarks ($\sigma_{pp\to \bar{t},t-channel} \approx 30.7$ pb) [47].

In the s-channel, a W boson creates either a top and an antibottom quark or an antitop and a bottom quark. This depends on the charge of the W boson, which is determined by the charge of the initial state quarks. Therefore, the same argumentation as for the t-channel holds and the production of top quarks ($\sigma_{pp\to t,s-channel} \approx 3.8$ pb) is preferred over the production of antitop quarks ($\sigma_{pp\to \bar{t},s-channel} \approx 1.8$ pb) [47].

In the tW-channel, the top (antitop) quark is produced in association with a W boson.

This production mode involves a gluon and a sea bottom (anti)quark in the initial state. Therefore, the cross section is charge-symmetric ($\sigma_{\mathrm{pp}\to\bar{\mathrm{t}},\mathrm{tW-channel}} = \sigma_{\mathrm{pp}\to\mathrm{t},\mathrm{tW-channel}} \approx$ 11.1 pb) [47].

3.1.3 Kinematic Quantities for Leading Order $\mathrm{t\bar{t}}$ Production

The kinematic quantities of the top quarks and the $\mathrm{t\bar{t}}$ system from top-quark pair production are measured within this thesis. Therefore, a short overview of these quantities and their connection to the underlying production mechanism in pp collisions is given in the following. A more detailed discussion can be found in [23]. For the production of top-quark pairs at LO, no additional partons occur in the partonic final state. For simplicity, only this case will be discussed further.

The invariant mass of the $\mathrm{t\bar{t}}$ system ($\mathrm{m}^{\mathrm{t\bar{t}}}$) is at LO equal to the centre-of-mass energy of the partonic interaction:

$$\mathrm{m}^{\mathrm{t\bar{t}}}_{\mathrm{LO}} = \sqrt{\hat{\mathrm{s}}} \geq 2 \cdot \mathrm{m}_{\mathrm{top}}. \tag{3.3}$$

Therefore, $\mathrm{m}^{\mathrm{t\bar{t}}}$ is a measure of the momentum fractions of the initial state partons and the pp centre-of-mass energy.

As the exact energy and momentum of the initial state partons from the proton is unknown, it is beneficial for all momenta to differentiate between a longitudinal and transverse component. The longitudinal component is chosen along the axis of the initial proton flight direction (z-axis). The initial momentum value depends on the PDFs of the initial state partons.
In contrast, the momentum of the initial state partons in the plane transverse to the z-axis can be considered as zero. Therefore, momentum conservation can be applied for this component and particle detectors are optimised for the measurement of transverse momenta.
The longitudinal momentum $\mathrm{p_z}$ is typically expressed by the rapidity y using the total Energy E:

$$\mathrm{y} = \frac{1}{2} \ln \left(\frac{\mathrm{E} + \mathrm{p_z}}{\mathrm{E} - \mathrm{p_z}} \right). \tag{3.4}$$

Rapidity differences are invariant under Lorentz transformations.

The rapidity of the $\mathrm{t\bar{t}}$ system ($\mathrm{y}^{\mathrm{t\bar{t}}}$) describes its boost in longitudinal direction and is at LO determined by the ratio of the proton momentum fractions $(\mathrm{x_1}, \mathrm{x_2})$ of the initial state partons.

$$\begin{aligned} \mathrm{E}^{\mathrm{t\bar{t}}}_{\mathrm{LO}} &= \mathrm{E}^{\mathrm{parton_1}} + \mathrm{E}^{\mathrm{parton_2}} \\ &= \sqrt{s}\,(\mathrm{x_1} + \mathrm{x_2}) \end{aligned} \tag{3.5}$$

$$\begin{aligned} \mathrm{p}^{\mathrm{t\bar{t}}}_{\mathrm{z,LO}} &= \mathrm{p}^{\mathrm{parton_1}} - \mathrm{p}^{\mathrm{parton_2}} \\ &= \sqrt{s}\,(\mathrm{x_1} - \mathrm{x_2}) \end{aligned} \tag{3.6}$$

Using Equation 3.4 to 3.6 results in:

$$\begin{aligned} y_{LO}^{t\bar{t}} &= \frac{1}{2}\ln\left(\frac{E_{LO}^{t\bar{t}} + p_{z,LO}^{t\bar{t}}}{E_{LO}^{t\bar{t}} - p_{z,LO}^{t\bar{t}}}\right) \\ &= \frac{1}{2}\ln\left(\frac{x_1}{x_2}\right). \end{aligned} \tag{3.7}$$

The transverse momentum of the $t\bar{t}$ system ($p_T^{t\bar{t}}$) is a measure of the recoil against additionally produced particles and therefore vanishes at LO:

$$p_{T,LO}^{t\bar{t}} = 0. \tag{3.8}$$

Consequently, the produced top quarks are at LO back-to-back in the plane that is transverse to the axis of the initial pp collision. Consequently, the difference in azimuthal angle of the two top quarks in this plane ($\Delta\phi\,(t,\bar{t})$) equals at LO π:

$$\Delta\phi\,(t,\bar{t})_{LO} = \pi. \tag{3.9}$$

The transverse momentum of the top quarks in the $t\bar{t}$ rest frame (p_T^t ($t\bar{t}$ com)) depends on the scattering angle (θ^*) in the $t\bar{t}$ rest frame and the magnitude of the top-quark momentum (p^{top}):

$$p_T^t\ (t\bar{t}\ \text{com}) = p^{top} \cdot \sin\theta^*. \tag{3.10}$$

For the LO process the magnitude of the top-quark momentum is given by the top-quark mass and the partonic centre-of-mass energy:

$$\begin{aligned} p_{LO}^{top} &= \sqrt{(E_{LO}^{top})^2 - (m^{top})^2} \\ E_{LO}^{top} &= \frac{1}{2}\sqrt{\hat{s}} \\ \longrightarrow p_{LO}^{top} &= \sqrt{\frac{1}{4}(\sqrt{\hat{s}})^2 - (m^{top})^2}. \end{aligned} \tag{3.11}$$

The top-quark transverse momentum in the detector rest frame (p_T^t) is a convolution of p_T^t ($t\bar{t}$ com) and $p_T^{t\bar{t}}$. For $t\bar{t}$ production at LO p_T^t and p_T^t ($t\bar{t}$ com) are equal.

3.2 Top-Quark Decay

3.2.1 Decay of Single Top Quarks

The decay of the top quark is characterised by the weak interaction. According to the CKM matrix (see Equation 2.1 and 2.2), the top quark decays almost exclusively into a W boson and a bottom quark ($V_{tb} \approx 1$).

The top-quark decay is special due to the large top quark mass. The top quark is besides the Higgs boson the only (currently known) particle which can decay into a real W boson. Furthermore, the large mass leads to a large value of the decay width (Γ_t):

$$\begin{aligned} \Gamma_t &\propto m_{top}^3 \\ &\approx 1.3\,\text{GeV} \end{aligned} \tag{3.12}$$

which results in a very short lifetime (τ_t) of:

$$\begin{aligned} \tau_t &= \frac{1}{\Gamma_t} \\ &\approx 5 \cdot 10^{-25}\text{s}. \end{aligned} \tag{3.13}$$

This is even a factor of approximately ten smaller than the time scale at which hadronisation takes place (see Chapter 2.3):

$$\begin{aligned} \tau_{\text{Had}} &\propto \frac{1}{\Lambda_{\text{QCD}}} \\ &\approx 3 \cdot 10^{-24}\text{s}. \end{aligned} \tag{3.14}$$

Thus, top quarks decay too fast to build colour neutral objects and offer therefore the unique possibility to study bare quarks. Hence, top-quark properties are perturbatively calculable and can be directly studied experimentally. For example, all spin information of the top quark is transmitted to its decay products and therefore experimentally accessible.

3.2.2 Top-Quark Pair Final States

The different $t\bar{t}$ final states are characterised by the decay of the two W bosons. Each W boson can decay either leptonically into a lepton and the corresponding antineutrino or hadronically into a quark and an antiquark. Considering the mass of the W boson and neglecting the flavour mixing, two different hadronic final states and three leptonic final states are possible for each W-boson.

The final state of the $t\bar{t}$ system is then named fully hadronic (both W bosons decay hadronically), dileptonic (both W bosons decay leptonically) and semileptonic (one W boson decays hadronically and one leptonically). The possible flavour compositions of the final state are illustrated in Table 3.1. The measured numbers for the relative

abundance (called *branching ratios* (BRs)) are taken from [32] and assume lepton universality. Counting all possible final states for the LO picture, neglecting flavour mixing but considering the factor of three for quarks because of the three possible colour states, the probability is $\frac{6}{9}$ for a hadronic W-boson decay and $\frac{1}{9}$ for each of the three possible leptonic W-boson decays. This results at LO in an expected BR of $\frac{36}{81}$ for the fully hadronic and the semileptonic final state and a BR of $\frac{9}{81}$ for the dileptonic final state.

W^--boson decay \ W^+-boson decay	$u\bar{d}$, $c\bar{s}$	$e^+\nu_e$	$\mu^+\nu_\mu$	$\tau^+\nu_\tau$
$\bar{u}d$, $\bar{c}s$	45.7 %	7.3 %	7.3 %	7.3 %
$e^-\bar{\nu}_e$	7.3 %	1.2 %	1.2 %	1.2 %
$\mu^-\bar{\nu}_\mu$	7.3 %	1.2 %	1.2 %	1.2 %
$\tau^-\bar{\nu}_\tau$	7.3 %	1.2 %	1.2 %	1.2 %

Table 3.1: BR of all possible $t\bar{t}$ final states, characterised by the decay of the two W bosons. The quoted BRs correspond to the measured values in [32], assuming lepton universality. The blue entries correspond to the e/μ+jets final state studied in this thesis.

The decay of the top quarks can be calculated using the same techniques as for the partonic production cross section. In principle, production and decay should be treated simultaneously. Although the absolute value of the top-quark width is large, its relative size with respect to the top-quark mass is small:

$$\frac{\Gamma_t}{m_{top}} \propto \frac{1}{m_{top}^2} \approx 0.0075. \tag{3.15}$$

This allows the application of the *narrow width approximation*, i.e. the separation of the process into an on-shell top-quark production and the subsequent decay.

3.2.3 Characteristics of the Semileptonic $t\bar{t}$ Final State

The final state investigated within this thesis is the semileptonic final state with one muon or electron (in the following called ℓ+jets, e/μ+jets or simply signal). About 30% of all top-quark pairs are decaying into this final state. A Feynman diagram for the semileptonic decay without additional radiation is shown in Figure 3.4.

The final state objects of the LO decay are one lepton (e or μ) and one corresponding neutrino, two b quarks and two light quarks (u and d or s and c flavour). In higher orders, additional particles can be produced, e.g. via FSR of the light and b quarks. For simplicity, this case will not be considered further and only the LO scenario is discussed in the following. The invariant mass of the two light jets and the invariant mass of the

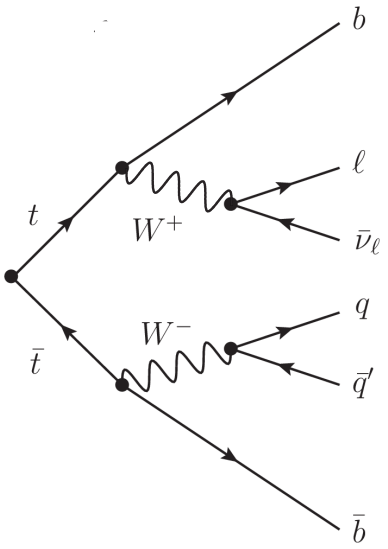

Figure 3.4: LO Feynman diagram for a semileptonic $t\bar{t}$ decay.

lepton and the neutrino form the W-boson mass of 80.4 GeV. Moreover, the invariant mass of the W boson and the associated b-flavoured quark form the top-quark mass. Invariant masses are used experimentally as constraints for the reconstruction of the top quarks from the measured detector objects.

The final state muon or electron is directly measured by the particle detectors. In contrast, the neutrino is only detected indirectly via a momentum imbalance in the transverse plane (E_T^{miss}). As discussed in Chapter 3.1.3 the initial state momentum in the plane transverse to the initial flight direction of the protons is zero. Therefore, also the sum of all transverse momenta of the final state objects is zero if all particles are correctly detected. In the case of one neutrino, the measured transverse momentum imbalance (E_T^{miss}) equals the transverse momentum of the neutrino.

Due to confinement (see Chapter 2.1) all final state quarks result in a bunch of colour neutral hadrons. Experimentally, the kinematics of each quark can be determined by clustering all its decay products into one object (called jet). The concept of a jet is illustrated in Figure 3.5. Furthermore, jets stemming from b quarks can be identified using the flight distance of the B-hadrons before their further decay, which is in the order of $c\tau = 500\mu m$.

QCD processes beyond LO in perturbation theory can lead to additional partons in the final state, which result in additional jets. One kinematic quantity involving these additional jets is the distribution of the invariant mass of the $t\bar{t}$ pair and an additionally produced jet ($m_{t\bar{t}j}$). From this, the dimensionless observable ρ_S is constructed using a scale m_0 in the order of m_{top}:

$$\rho_S = \frac{2 \cdot m_0}{m_{t\bar{t}j}}, \quad m_0 = 170 \text{ GeV}. \tag{3.16}$$

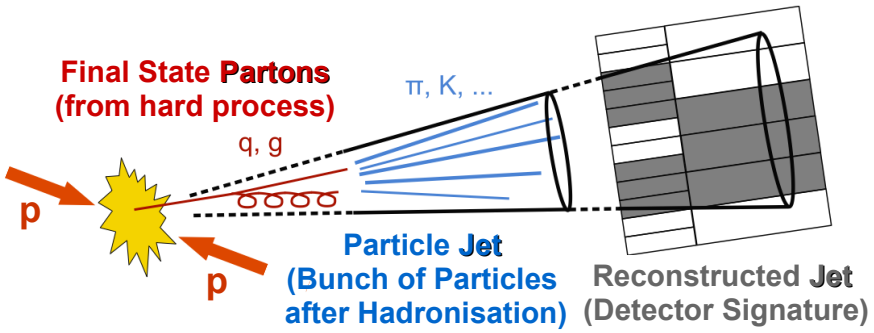

Figure 3.5: Sketch of proton-proton collision with a parton (quark or gluon) in the final state, resulting in a collimated spray of particles, a jet. Taken from [48] and modified.

The measured distribution of ρ_S is expected to be sensitive to the top-quark mass [49].

Other physics processes that can mimic the e/μ+jets final state are the production of leptonically decaying W or Z bosons accompanied by additional jets, the production of QCD multi-jet final states or the production of single top quarks. Another background process are $t\bar{t}$ events that are decaying into a different final state.

The e/μ+jets final state is a good compromise between BR and experimentally manageable background processes.
In comparison, the dileptonic final state has an intrinsic lower background contribution but a smaller BR and the fully hadronic final state has a larger BR but suffers from a large QCD multi-jet background.
Semileptonic final states with a τ-lepton are experimentally challenging because in contrast to a muon or electron the τ-lepton decays fast and has to be reconstructed from its decay products. Moreover, the decay of τ-leptons result in additionally produced neutrinos that make this final state particularly complicated.

Another advantage of the e/μ +jets final state in contrast to the dileptonic final state is that the $t\bar{t}$ event topology can be reconstructed from the measured final state objects. The method for the event reconstruction used in this thesis is discussed in Chapter 7. Besides the z-component of the neutrino momentum, all final state objects can be directly or indirectly measured. The missing component can be compensated by the introduction of kinematic constraints, e.g. for the W-boson mass ($m(\ell\nu) \equiv m_W$).
In comparison, all final state objects are directly measured in the fully hadronic final state but the presence of four indistinguishable light quarks results in an intrinsically higher number of possible jet assignments to the underlying top quarks.
The difficulty for an event reconstruction in the dileptonic final state is that the kinematic system is kinematically underconstrained due to the two neutrinos for which only the sum of their transverse momenta is measured as E_T^{miss}. Therefore, additional constraints need to be introduced for the reconstruction of the top quarks in the dileptonic $t\bar{t}$ final state.

3.3 The Top-Quark Mass

Its large mass leads to many unique properties of the top quark like the short lifetime or the large coupling to the Higgs boson. Within this thesis, the measurement of the top-quark mass from differential cross sections is studied and therefore, a short overview about the definition of the top-quark mass and its relevance for the stability of the electroweak vacuum will be given in the following. A more detailed overview can be found in [43, 50, 51].

3.3.1 Definitions

Different concepts for the definition of the top-quark mass exist and especially the relation of the measured top-quark mass to the theoretical mass definitions is frequently under discussion.

Pole Mass

Using the Feynman Rules, an internal fermion line within a Feynman diagram is described by a propagator. For massive fermions, this propagator has a pole at $m - \frac{i}{2\Gamma}$. The real part of this complex pole is called the *pole mass* (or on-shell mass).

$\overline{\text{MS}}$ Mass

Within perturbation theory (see Chapter 2.3), particle masses can be accurately defined for high energies (short distances) involving a renormalisation procedure. The most common renormalisation scheme is the modified minimal subtraction ($\overline{\text{MS}}$) scheme [52]. The $\overline{MS}$ *mass* (or short-distance mass) defined within the perturbative expansion, is a well-behaved parameter from the theory point of view.

The relation between the top-quark $\overline{\text{MS}}$ mass and the top-quark pole mass can be exactly calculated and is in the order of 6%:

$$m_{\text{top}}^{\overline{\text{MS}}} \approx m_{\text{top}}^{\text{pole}} - 10\,\text{GeV}. \tag{3.17}$$

Experimental and Monte-Carlo Mass

Within the most precise top-quark mass measurements, the *(experimental) mass* of the top quark is constructed from the invariant mass of its decay products. In first approximation, the experimental mass can be identified with the pole mass. The difference between the two definitions arise from non-perturbative QCD effects within the hadronisation process and are expected to be in the order of 1 GeV. A more detailed discussion can be found in [53].

Typically, these top-quark mass analyses use predictions of Monte-Carlo (MC) event

generators, which are detailed in Chapter 5.3. Therefore, the measured mass equals in principle the top-quark mass parameter as defined in the used MC event generator.

3.3.2 Stability of the Electroweak Vacuum

As described in Chapter 2.4, the SM Higgs mechanism involves a Mexican Hat potential with a non-vanishing vacuum expectation value for the ground state. The minimum of this potential characterises the stability of the electroweak vacuum. The key question is if this minimum is global for all energy scales. If this is not the case, the universe can possibly change its state from the local to the global minimum via a phase transition through quantum tunnelling. In this context, three scenarios are commonly discussed:

- **stable**: the minimum of the Higgs potential is global
- **metastable**: the probability for quantum tunnelling is larger than the age of the universe
- **unstable**: the probability for quantum tunnelling is smaller than the age of the universe

The scenario of an unstable electroweak vacuum is hard to understand as the existence of our universe would be really by chance. Therefore, this scenario is commonly interpreted as indicator for new physics beyond the SM that is needed to modify the prediction and therefore also the interpretation.

The stability of the electroweak vacuum can be parametrised by the experimentally accessible masses of the top quark and the Higgs boson.

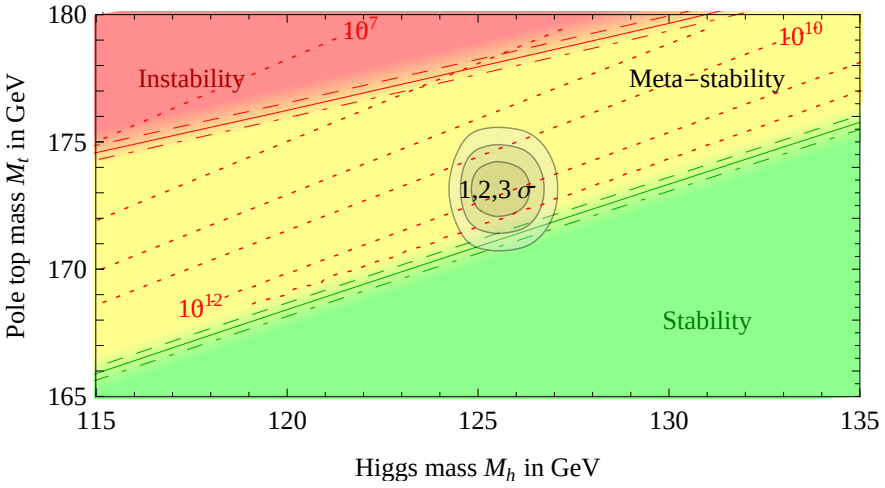

Figure 3.6: Illustration of the stability of the electroweak vacuum as a function of the top-quark mass, the W-boson mass and the Higgs-boson mass. Taken from [10].

As illustrated in Figure 3.6, the measured values correspond to a metastable scenario close to the stability region and no final conclusion can be drawn with the current experimental precision. A stable electroweak vacuum allowing the validity of the SM up to the Planck scale of $\mathcal{O}(10^{19}\,\text{GeV})$ is possible, albeit the measured masses differ from this region by more than two standard deviations.

3.4 The Top Quark and its Relevance for Higgs Physics and Physics Beyond the Standard Model

After the discovery of the Higgs boson, top-quark physics is even more relevant than before. Besides the already discussed role of the top quark for the stability of the electroweak vacuum, the large mass of the top quark implies the largest yukawa coupling to the Higgs boson for all (known) SM particles. The Higgs boson production via gg fusion involves e.g. a top-quark loop. Moreover, measuring the coupling between the Higgs boson and the top quark is an important step to test if the Higgs boson has exactly the properties as predicted by the Higgs mechanism within the SM.

New physics can manifest itself directly in $t\bar{t}$ final states. A common scenario is a hypothetical new particle decaying into a pair of top quarks. Such a scenario would e.g. result in a resonance of the invariant mass of the top-quark pair. Possible resonances are predicted by technicolour models [54–56], axigluon models [57] or Kaluza-Klein excitations [58–60] in extra-dimensional models [6, 61]. Furthermore, different BSM scenarios can lead to changes of the $t\bar{t}$ kinematics, e.g. for $\Delta\phi\,(t,\bar{t})$ [62].

Due to their complex final state, top-quark pairs are an important SM background process for many BSM searches. Event topologies with several leptons, high E_T^{miss} or high jet activity are possible. Therefore, the precise knowledge of the $t\bar{t}$ production process and the properties of the experimentally observed final states is important.
Especially supersymmetric (SUSY) extensions of the SM [63] involve the top quark in many ways. In some SUSY models, the top quark can decay into a charged Higgs boson instead of the W boson. In other scenarios, its Supersymmetric partner decays into the top quark and an undetectable stable neutralino, resulting in a $t\bar{t}$ final state with additional E_T^{miss}.
Consequently, differential $t\bar{t}$ cross section measurements are an important tool to understand this SM process and constrain possible implications on BSM searches.

Chapter 4

Experimental Apparatus

All measurements presented in this thesis are performed with pp collision data of the Large Hadron Collider (LHC) recorded by the Compact Muon Solenoid (CMS) detector. Detailed descriptions of the LHC and the CMS detector can be found in [64] and [65], respectively. A short overview about the LHC is given in Chapter 4.1 and the CMS detector is briefly introduced in Chapter 4.2.

4.1 The Large Hadron Collider

The LHC is the world's most powerful (human-built) accelerator to date. It offers the possibility to study proton as well as heavy ion collisions at high centre-of-mass energies and high rates.
Being located near Geneva in the border region of France and Switzerland, the LHC is operated by the European Organization for Nuclear Research CERN[a]. The LHC collider ring of 27 km circumference is located in a tunnel 100 m below the ground and embedded in the CERN accelerator complex (see Figure 4.1).

In the following, only the pp operation mode will be discussed. Protons are created by ionising Hydrogen. They are pre-accelerated to an energy of 450 GeV in several steps before they are injected into the LHC, where they are accelerated further and finally brought to collision. For the pre-acceleration, linear accelerators (LINACs), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) are used.

Within the LHC ring, the proton beams revolve clockwise and anticlockwise in two evacuated beam pipes. Superconducting magnets are used to focus and guide the beams. Within the beams, protons are grouped in bunches. Moreover, superconducting radio-frequency cavities are used for further acceleration. Liquid helium cooling is used to achieve temperatures of 1.9 K, which are needed for superconductivity.
When the final beam energy is reached, the protons are brought to collision. Being designed for a centre-of-mass energy of $\sqrt{s}$ =14 TeV for pp collisions, the LHC was

[a]The acronym CERN is derived from its previous name Conseil Européen pour la Recherche Nucléaire.

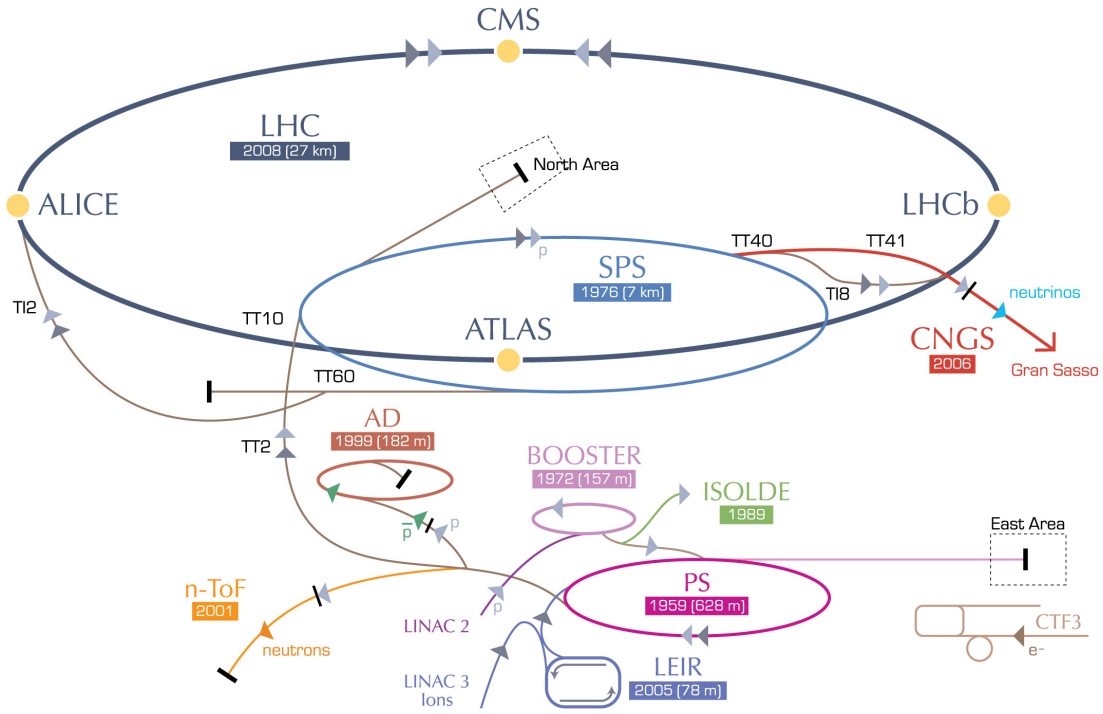

Figure 4.1: Overview of the CERN accelerator complex with the LHC, its pre-accelerators and main experiments. Taken from [66].

operated at centre-of mass energies of up to $\sqrt{s} = 7$ TeV in 2010 and 2011 and up to $\sqrt{s} = 8$ TeV in 2012. Higher centre-of-mass energies ($\sqrt{s} \geq 13\,\text{TeV}$) are envisaged for the next operation periods scheduled for 2015.

Six main experiments are located around the LHC ring. Each experiment consists of a particle detector located at one of the four collision points:

- ALICE (A Large Ion Collider Experiment) [67]
- ATLAS (A Toroidal LHC ApparatuS) [68]
- CMS (Compact Muon Solenoid) [69]
- LHCb (Large Hadron Collider beauty experiment) [70]
- LHCf (LHC forward) [71]
- TOTEM (TOTal Elastic and diffractive cross section Measurement) [72]

The four major experiments are ALICE, ATLAS, CMS and LHCb. ALICE is dedicated to heavy-ion physics and the investigation of the quark-gluon plasma. LHCb is specialised on physics with hadrons originating from bottom quarks. ATLAS and CMS are multi-purpose experiments, designed for SM physics measurements, the Higgs discovery and Higgs physics, as well as for various new physics searches (e.g. SUSY). The two smaller experiments LHCf and TOTEM are dedicated to forward physics and the measurement of the total inelastic cross section of pp collisions.

The most important parameter of the LHC besides the centre-of-mass energy is the instantaneous luminosity ($\mathcal{L}$), which is a measure of the pp interaction rate. $\mathcal{L}$ is determined by the parameters of the machine:

$$\mathcal{L} = \frac{N_b f N_{p,1} N_{p,2}}{A}. \tag{4.1}$$

Here, N_b is the number of proton bunches, f is the bunch collision frequency (11 kHz), $N_{p,1}$ and $N_{p,2}$ are the number of protons within the colliding bunches, and A is the so-called beam profile. A depends e.g. on the transverse and longitudinal bunch sizes and the beam crossing angle.

The time-integrated luminosity ($\mathscr{L}_{int} = \int \mathcal{L} dt$) is a measure of the size of the recorded dataset and correlates the number of expected events (N_X) from a physics process X to its total inclusive cross section $\sigma_{pp\to X}$:

$$N_X = \mathscr{L}_{int} \cdot \sigma_{pp\to X}. \tag{4.2}$$

The LHC design values are $\mathcal{L} = 10^{34} cm^{-2} s^{-1}$ with $N_p = 1.15 \cdot 10^{11}$ protons per bunch and $N_b = 2808$ bunches per beam, separated by 25 ns, which corresponds to a bunch crossing rate of 40 MHz. In the 2012 pp data-taking period at $\sqrt{s} = 8$ TeV, a maximum instantaneous luminosity of $\mathcal{L}(2012, max) = 7.7 \cdot 10^{33} cm^{-2} s^{-1}$ was reached using a setup with 1374 bunches per beam separated by 50 ns [73] and $1.6 - 1.7 \cdot 10^{11}$ protons per bunch. The integrated luminosity delivered by the LHC is $23.3\,fb^{-1}$ from which $21.8\,fb^{-1}$ have been recorded by the CMS detector. The evolution of the integrated luminosity over time for 2012 is shown in Figure 4.2.

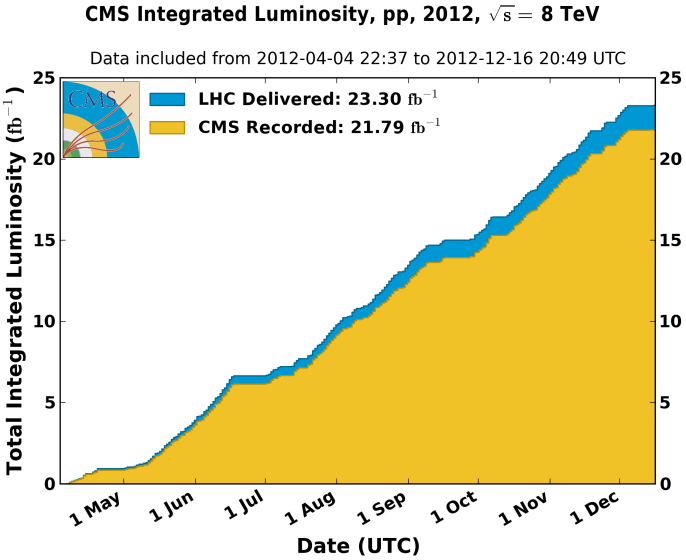

Figure 4.2: Evolution of the integrated luminosity over time for the 2012 proton-proton data-tacking period at a centre-of-mass energy of $\sqrt{s} = 8$ TeV. Shown are the integrated luminosity as delivered by the LHC and the integrated luminosity recorded by the CMS detector. Taken from [74].

4.2 The Compact Muon Solenoid Experiment

The Compact Muon Solenoid (CMS) detector is located 100 m underground. It has the typical cylindrical, onion-like structure of high energy physics particle detector with several subdetectors placed around one nominal interaction point where the two proton beams of the LHC are brought to collision. The CMS detector is radially symmetric around the direction of the beams (beamline) and symmetric along the beamline in both directions of the interaction point.
Different technological concepts are used to identify the final state particles resulting from the (pp) collisions (hadrons, photons, muons, electrons) and measure their energies, charges and trajectories.

A schematic overview of the CMS detector is shown in Figure 4.3.
The subdetectors are arranged such that a good coverage of the solid angle is achieved.

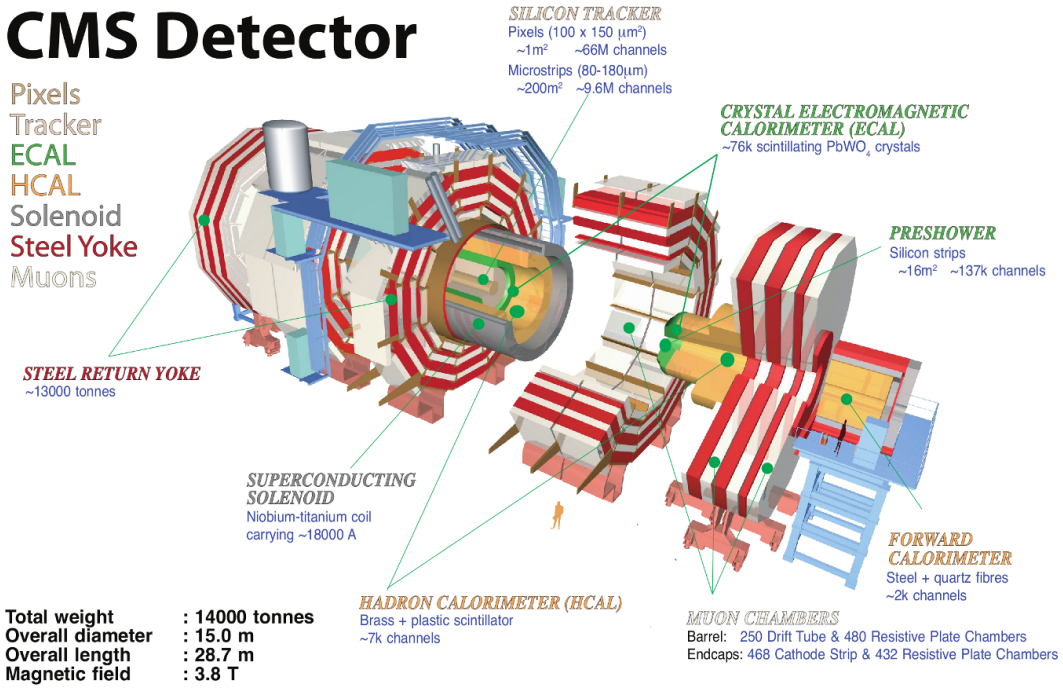

Figure 4.3: Overview of the subdetectors of the CMS-detector. Taken from [75].

Closest to the nominal interaction point, the silicon-tracking system with an inner pixel detector and several strip detectors is installed to measure the trajectories of charged particles. The electromagnetic calorimeter (ECAL) is intended to absorb electrons and photons and measures their energies. Similar, the hadronic calorimeter (HCAL) is designed to absorb and measure the energy of hadrons. Surrounding the inner tracking system and the calorimeters, a superconducting solenoid generates a magnetic field of 3.8 T and bends the particle tracks so that their charge and momentum can be derived from the trajectory. Around the solenoid, several muon chambers are installed

to distinguish muons, which are not absorbed by the calorimeters, from other particles. Altogether, the CMS detector has a length of 28.7 m, a diameter of 15.0 m and a total weight of about 14,000 t. A brief survey for the single subdetectors is given in the Chapters 4.2.1 to 4.2.4 while a more detailed description of the CMS detector can be found elsewhere [76].

The right-handed coordinate system used for physics analysis is determined by the radial design of the detector. The origin of the coordinate system is at the centre of the detector, i.e. at the nominal interaction point. The x axis points radially inwards to the centre of the LHC ring, the y axis vertically upwards and the z direction is chosen along the anticlockwise beamline.
The transverse (x,y or r,ϕ) plane is (as discussed in Chapter 3.1.3) used to express many momentum related quantities such as p_T or the total momentum imbalance E_T^{miss}. The azimuthal angle ϕ is measured in the transverse plane with respect to the x axis. Furthermore, the polar angle θ, which is measured in the (y,z) plane with respect to the z axis, is usually expressed by the pseudorapidity η:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{4.3}$$

The pseudorapidity is for massless particles (for which E=p) equal to the rapidity discussed in Chapter 3.1.3:

$$\begin{aligned} \eta &= \frac{1}{2}\ln\frac{p+p_z}{p-p_z} \\ &\overset{(m=0)}{=} \frac{1}{2}\ln\frac{E+p_z}{E-p_z} \\ &\overset{(Eq.3.4)}{=} y. \end{aligned} \tag{4.4}$$

Furthermore, angular distances are usually expressed as $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$.

4.2.1 Tracking System

The tracking system [77, 78], often referred to as tracker, is intended to measure the trajectories of charged particles as precise as possible. It is one of the most important parts of the CMS detector and consists of semi-conducting silicon as active material. Due to ionisation, traversing charged particles create electron-hole pairs in the (pixel or strip) semiconductor device, resulting in an electronic signal which is interpreted as hit. From several hits the particle trajectory is reconstructed. Extrapolating the reconstructed tracks to the nominal interaction point, primary and potentially also secondary vertices are reconstructed. In addition, the bending magnetic field of the solenoid allows for the determination of the particle momenta and the signs of the particle charges.

The whole tracker has a sensitive area of approximately 210 m^2 and is therefore the to date largest silicon tracker of all particle physics experiments. It has a length of

5.8 m with a radius of 1.1 m and covers the range of $|\eta| \leq 2.5$. Due to the decreasing particle flux with increasing distance from the nominal interaction point, the tracker is divided into different parts. An overview of the tracker is shown in Figure 4.4.

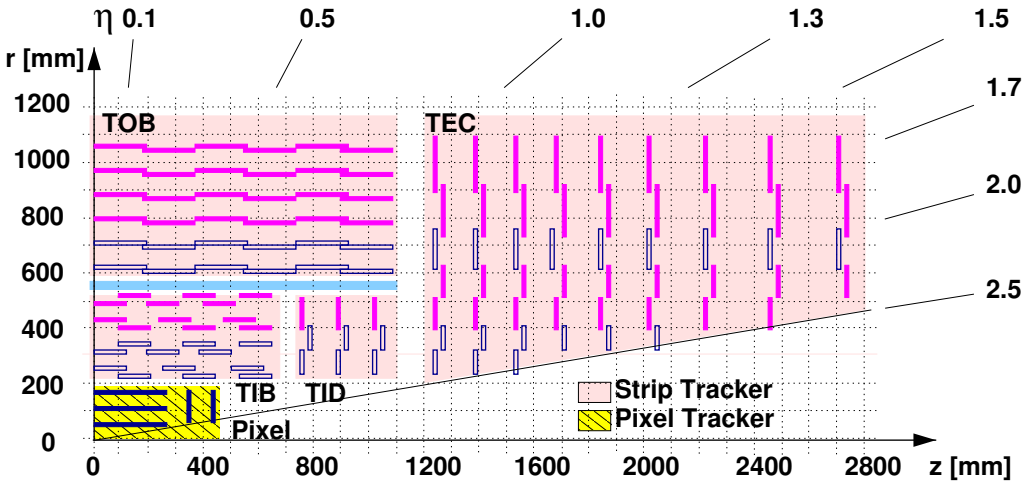

Figure 4.4: Schematic cross section through the CMS tracker in the (r,z) plane. Taken from [79] and modified.

Silicon Pixel Detector

The pixel detector is the inner-most subdetector of CMS, constructed from 66 million silicon pixels of 100×150 μm^2 size and arranged in 1440 modules. Three concentric cylindrical barrel layers surround the nominal interaction point and four discs close the barrel ends. The 53 cm long barrel layers are located at 4.4, 7.3 and 10.2 cm radial distance from the nominal interaction point while the endcap discs are located at longitudinal distances of 34.5 cm and 46.5 cm. The high granularity of the pixel design results in a high hit-reconstruction efficiency of typically above 99% and a typical spatial hit resolution of about 10 μm in transverse and $20 - 40$ μm in longitudinal direction [80].

Silicon Strip Detectors

The pixel detector is surrounded by the strip detector which consists of 9.3 million silicon strips, arranged in 15,148 modules. The strip detector is subdivided into the *tracker inner barrel* and *discs* (TIB,TID) and the *tracker outer barrel* and *endcaps* (TOB,TEC).
The TID and TIB are composed of four radially arranged layers of strips parallel to the beamline and three endcap discs at each side. These strips have a pitch size of 80 to 141 μm and a length of 10 cm. The occupancy is typically about 3% and the (r,ϕ) resolution is typically $13 - 38$ μm in the TIB [80].
The TOB consists of six layers, surrounding TIB and TID and the TEC is composed of

nine discs at each side, bordering the other tracking subdetectors. Here, strips with 97 to 184 μm pitch size and 25 cm length are used. An occupancy at the percent level and a typical (r,ϕ) resolution of 18 to 47 μm is achieved in the TOB [80].
The inner two layers of TIB and TOB, the inner two rings of TID and TEC and the fifth ring of the TEC have in addition a strip module mounted back-to-back with a stereo angle of 100 mrad. This allows measuring the longitudinal coordinate in the barrel and the radial coordinate in the discs. The achieved single-point resolutions of this measurements are about one order of magnitude worse than the one in the transverse direction [80].

4.2.2 Calorimeters

Electromagnetic Calorimeter

The CMS *electromagnetic calorimeter* (ECAL) [81] is built from lead tungstate ($PbWO_4$) crystals with a short radiation length[b] of $X_0 = 0.89$ cm where electromagnetic showers are induced. As the CMS ECAL is a homogeneous calorimeter, the crystals act as absorber and scintillator. The corresponding scintillation process is fast. About 80% of the light is emitted within 25 ns with wavelengths of 420-430 nm. This scintillation light is detected by avalanche photodiodes (APDs) or vacuum phototriodes (VPTs). A cooling system enables the operation at a stable temperature to ensure equal rates of the detected photo electrons. On average, a signal of 4.5 photo electrons is detected for a particle depositing an energy of 1 MeV.

The ECAL is subdivided into the ECAL barrel (EB) calorimeter, the ECAL endcap (EE) calorimeters and the ECAL preshower (ES) detectors. The cylindrical EB calorimeter is located at a distance of 1.3 m from the beam pipe and covers the range of $|\eta| < 1.479$. An EE calorimeter disc is located on both sides of the EB calorimeter extending the range to $|\eta| < 3.0$. Furthermore, an ES detector is located in front of each EE calorimeter disc, covering the range $1.653 < |\eta| < 2.6$. An overview of the layout of the CMS ECAL is shown in Figure 4.5.

The EB calorimeter consists of 61,200 crystals with a high granularity of 0.0174×0.0174 in (η,ϕ). This corresponds to 22×22 mm^2 at the front face. The length of the EB calorimeter crystals is 230 mm corresponding to 25.8 X_0. In order to avoid particles passing trough the cracks between the crystals, they are tilted by 3° with respect to the radial direction towards the nominal interaction point. APDs are used for the light detection.

Each EE calorimeter has 7,324 crystals with a front-face cross section of 28.6×28.6 mm^2 and a length of 220 mm corresponding to 24.7 X_0. Similar to the EB calorimeter, the crystals are slightly tilted with respect to the direction of the nominal interaction point. VPTs are used for the light detection.

The ES detector in front of each EE is a two-layer sampling calorimeter with high

[b] The radiation length X_0 is defined as the length after which a traversing electron has reduced its initial energy E_0 to $\frac{1}{e}E_0$ due to bremsstrahlung.

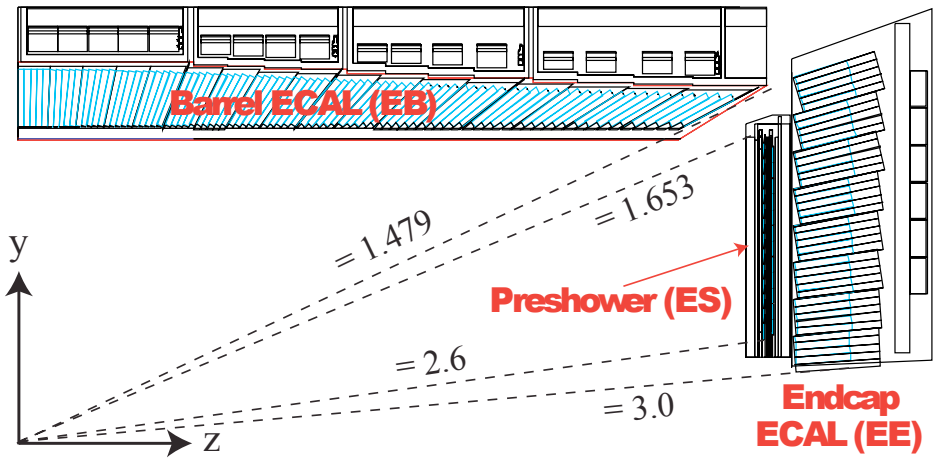

Figure 4.5: Schematic cross section of one quarter of the electromagnetic calorimeter of the CMS detector in the (y,z) plane. Taken from [65].

spatial resolution. It is used for an improved discrimination between single photons and neutral pions decaying into two photons. Each ES detector consists of a 3 X_0 thick lead absorber and silicon strip sensors.

The relative energy resolution for electrons (as measured with test beams [82]) is:

$$\left(\frac{\sigma(\mathrm{E})}{\mathrm{E}}\right)^2 = \left(\frac{2.8\%}{\sqrt{\mathrm{E[GeV]}}}\right)^2 + \left(\frac{12\%}{E[\mathrm{GeV}]}\right)^2 + (0.3\%)^2\,. \tag{4.5}$$

The first term describes stochastic effects of the shower development, the second term parametrises electronic noise and the last term contains remaining effects such as inter-channel miscalibration.
Consequently, the achieved energy resolution for 30 GeV (80 GeV) electrons is 0.72% (0.46%).

Hadronic Calorimeter

The design of the CMS *hadronic calorimeter* (HCAL) [83] is driven by the space restrictions between the ECAL and the solenoid.
The HCAL is subdivided into the hadronic barrel (HB), hadronic endcap (HE), hadronic outer (HO) and hadronic forward (HF) calorimeters. The HB calorimeter surrounds all inner detector systems previously discussed and fills up the space between EB calorimeter and the solenoid. An HE calorimeter is located at both ends of the HB calorimeter between the EE calorimeter and the endcaps of the return yoke. The HB calorimeter covers a range of $|\eta| < 1.4$ while the HE calorimeter covers the region $1.3 < |\eta| < 3.0$. In addition, the HF is located close to the beam pipe at a distance of 11.2 m from the nominal interaction point and covers the range $3.0 < |\eta| < 5.2$. Furthermore, the HO calorimeter is located between solenoid and muon system to catch

particles not absorbed by the HB calorimeter.
An overview of the CMS HCAL layout is shown in Figure 4.6.

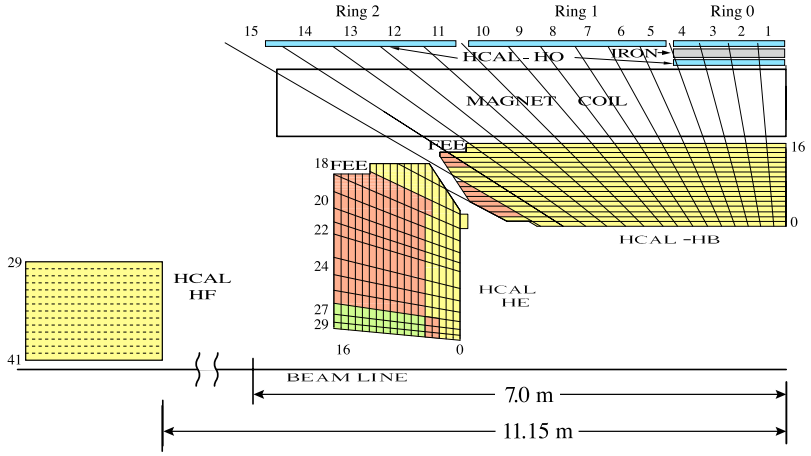

Figure 4.6: Schematic cross section of one quarter of the hadronic calorimeter of the CMS detector in the (y,z) plane. Taken from [84].

The HB and HE calorimeters are sampling calorimeters built of alternating layers of non-magnetic brass as absorber (the first and last layer is made of steel) and plastic scintillator as active material. The brass absorber plates have a thickness of about 5 cm in the HB calorimeter and 8 cm in the HE calorimeter and an interaction length[c] of $\Lambda_I = 16.4$ cm. The scintillation layers have a thickness of 3.7 mm (9 mm for the most inner layer). Within the scintillator modules, wavelength-shifting fibers are embedded. Hybrid photodiods (HPDs) are used to detect the shifted scintillation light.
The calorimeters are segmented in towers of size 0.087×0.087 in (η, ϕ), pointing with their axes radially to the interaction point. One HCAL tower matches 5×5 ECAL crystals. For $|\eta| \geq 1.6$ a lower granularity (larger tower size) is chosen. In all cases, one HCAL tower matches a specific array of ECAL crystals, e.g. 5×5 ECAL crystals in the most inner region.
The HE calorimeter has a sufficient effective depth of about 10 Λ_I. Due to the limited space for the HB calorimeter, an effective depth of only 5.8 Λ_I is achieved for $|\eta| = 0$. The ECAL adds about 1 Λ_I.

The HO calorimeter uses the solenoid itself and partly additional steel as absorber in combination with some scintillator layers. It extends the effective depth in the central detector region to at least 11.8 Λ_I (for $|\eta| = 0$).

The HF calorimeter is an iron and quartz-fiber calorimeter. Quartz fibers with a diameter of about $600 - 800$ μm are aligned parallel to the beamline and embedded in a steel absorber structure. Charged shower particles generate Cherenkov light[d] in

[c] The nuclear interaction length Λ_I is defined as the mean distance travelled by a hadronic particle before undergoing an inelastic nuclear interaction.

[d] Cherenkov radiation is electromagnetic radiation emitted by charged particles traversing a dielectric

the fibers which is detected by photomultipliers. Half of the fibers run through the full length of the iron absorber (165 cm) while the other half starts at a depth of 22 cm from the front face. This allows the distinction between hadronic and electromagnetic (EM) showers because EM showers typically deposit most of their energies fast, i.e. within the first centimeters of the absorber.
The fibres are bundled in towers of 0.175×0.175 in (η,ϕ) except for the towers with highest $|\eta|$ where a lower granularity is chosen.

The performance of the HCAL is measured with test beams. For example, the energy resolution for single pions is found to be approximately $\frac{120\%}{\sqrt{\mathrm{E[GeV]}}}$ in the HB [85].

4.2.3 Solenoid Magnet and Return Yoke

A central feature of the CMS detector is the world's largest superconducting solenoid to date, providing a homogeneous magnetic field of 3.8 T. This is about 100,000 times stronger than the magnetic field of the earth. This results in a bending of all charged particles that travel through CMS (with $\mathrm{p_T} \neq 0$), enabling the determination of the particle momenta from the measured tracks.
Furthermore, the tracker and the main parts of the calorimeters are placed inside the magnetic coil to minimise the absorbing material. Consequently, it is relatively large (6 m inner diameter, 12.5 m length).
The magnetic flux is returned using a massive iron yoke. This steel return yoke makes up the largest part of the total weight and consists of five wheels around the solenoid and two endcaps.

4.2.4 Muon System

In contrast to other particle types, muons traverse the tracker and calorimeters without significant energy loss. Therefore, the CMS muon system [86] is placed outside the solenoid for the identification of muons. The additional information of the muon system improves also the measurement of the tracker-based muon momentum. Furthermore, the unique detector signature of muons is often used for triggering (see Chapter 4.2.5).

The CMS muon system consists of three different types of gaseous muon detectors which are embedded in the return yoke. Depending on the expected particle flux, drift tube chambers (DTs) or cathode strip chambers (CSCs) are used in the endcap or barrel region complemented by resistive plate chambers (RPCs). The DTs cover the range $|\eta| < 1.2$ while the CSCs are used at $0.9 \leq |\eta| < 2.4$. In addition, the RPCs cover the range $|\eta| < 1.6$. An overview of the CMS muon system is shown in Figure 4.7.

The expected particle flux in the barrel region facilitates the usage of DTs. DTs and RPCs are combined in muon stations. Four layers of muon stations are located in the barrel region. Each station has eight DTs which are aligned parallel to the beam line and provide a measurement in the (r,ϕ) plane. The three inner stations have in addition

medium at a speed greater than the phase velocity of light in this medium.

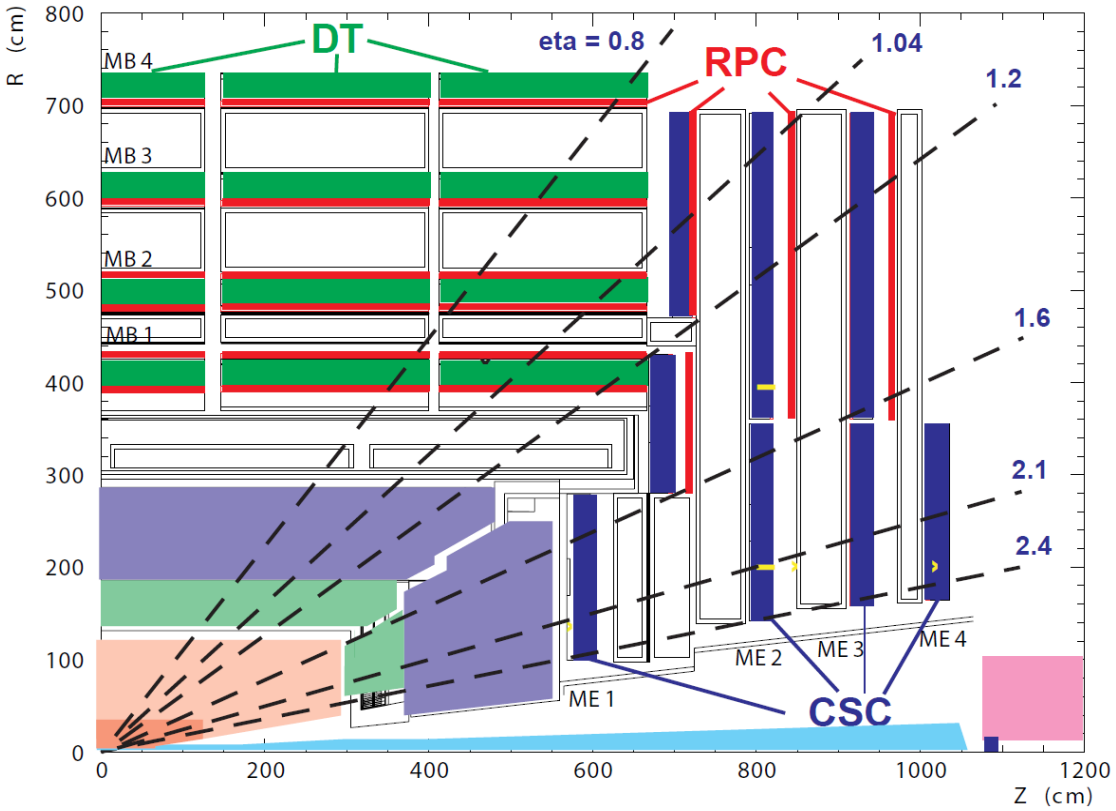

Figure 4.7: Schematic cross section of one quarter of the muon system of the CMS detector in the (r,z) plane. Taken from [65].

four radial aligned DTs, providing a measurement in the longitudinal z direction.

The higher particle flux in the endcap regions necessitates the usage of CSCs. Parallel and radial oriented strips with respect to the beam line provide measurements in (r,ϕ) and η. Up to four muon stations with CSCs and RPCs are located in the endcap regions.

In CMS, muon trajectories are typically reconstructed combining the information of the tracker and the muon system (see Chapter 6.1 and 6.4). The related measurement of the muon momentum is significantly improved by the long lever arm of the muon system for high p_T values. The achieved relative muon resolution using tracker and muon system is typically about $1-3\%$ [87] for $20\,\text{GeV} \leq p_T \leq 100\,\text{GeV}$, worsening for larger values of $|\eta|$ and p_T.

4.2.5 Particle Identification and Trigger System

Combining the information of several subsystems, physics objects like muons or electrons are reconstructed. The simplified signature for all detectable particle types are illustrated in Figure 4.8 while their exact definition and reconstruction procedure is discussed in Chapter 6.

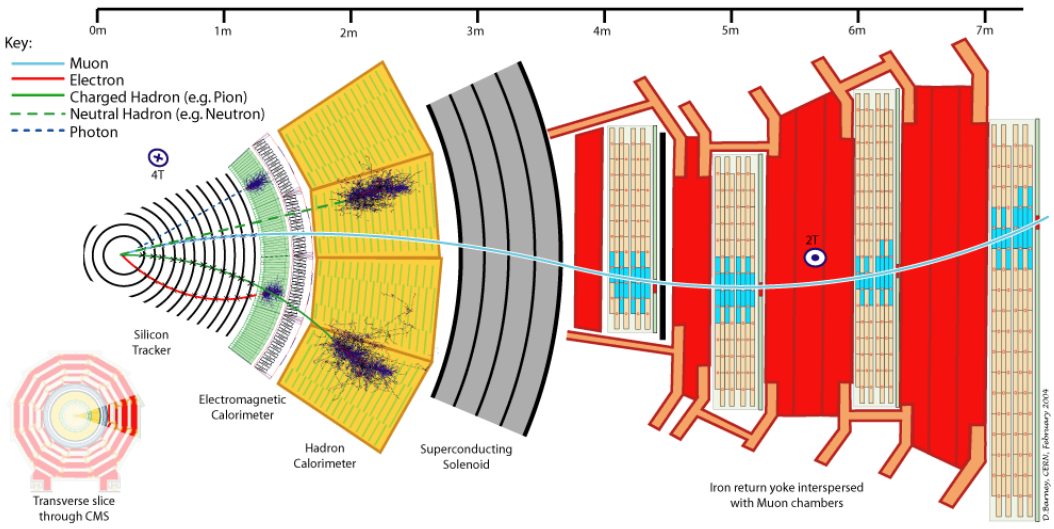

Figure 4.8: Principle of particle identification in CMS using all subdetectors. A schematic cross section through the CMS detector in the (r,ϕ) plane is shown. Taken from [88].

Reconstructed particles are e.g. used to identify events of interest used for the analysis input. In the first step, it has to be decided if the event is recorded. The high granularity of the CMS detector results in an amount of about 1 MB of raw information per event. Given the high bunch-crossing rate of 20 MHz (for 2012 pp collisions), the resulting data from all collisions is too large to be stored. Only data in the order of 100-300 Hz can be recorded. Therefore, CMS uses a two-step trigger system [89] for a fast and efficient online selection. Different triggers are developed to identify event signatures of interest, e.g. with high activity in the central detector region, reconstructed leptons or missing transverse energy, which are finally stored. These triggers are developed for specific analysis types (Higgs analysis, top-quark analysis, etc).

The first trigger step are hardware-based *level-1* (L1) *triggers*. L1 triggers involve only the fast processable information of the calorimeters and the muon system (with reduced granularity) to limit the time of the trigger decision to 3.2 μs. Simplified and therefore fast algorithms are used to build primitive particle candidates and simple event quantities like e.g. the global energy sum. This reconstruction procedure is done first locally and regionally in the individual detector systems. Subsequently, the information from all subsystems is merged and an event is selected if certain thresholds, e.g. related to the event activity or high energetic objects, are passed.

In the second trigger step software-based *high-level triggers* (HLT) are used. In contrast to the L1 triggers, the more complicated information is used in this final trigger step. This involves also the reconstruction of more complex (physics) objects like muons or jets using more sophisticated algorithms. The involved physics objects are similar to the ones used for the offline analysis. For a fast ($<$ 50 ms per event) and efficient performance, the reconstruction of more complex objects, that involve time consuming algorithms, is limited to detector regions marked as interesting by the

L1 input triggers. Typically, HLT requirements involve one or several physics objects (electrons, muons, jets) with minimum p_T, η and identification or isolation requirements. Also HLT triggers with global event requirements like minimum E_T^{miss} or minimum $H_T = \sum_{\text{jet i=0}}^{N_{jets}} p_T(\text{jet i})$ exist. Events fulfilling the requirement of at least one HLT path are recorded for further offline analysis. To reduce the stored information further, some trigger paths with high rates are prescaled and only a fraction on the events passing these trigger criteria are recorded.

Although the total event rate is reduced by a factor of about 10^5 by the trigger system, the total amount of data which needs to be stored is still high. For design luminosity, the data of all LHC experiments which has to be stored is about 700 MB per second, summing up to about 10^7 GB within one year. Therefore, a complex and highly advanced *data acquisition system* (DAQ) [90] as well as a high-performant computing infrastructure [91] is needed for the further data analyses.

Chapter 5

Data and Simulation

5.1 Software Framework

The official *CMS software* (CMSSW), version 5.3.11 [92], is used for the analysis of data and simulated samples. In particular, the *Top-Quark Analysis Framework* (TQAF) [93] is used for the kinematic reconstruction of the $t\bar{t}$ events.
Detector conditions and object calibrations like e.g. alignment constants or *Jet Energy Corrections* (JEC) are summarised within the *Global Tag* (GT). For this thesis, the following GTs are used:

- Data: FT_53_V21_AN5
- Simulation: START53_V27

All analysis specific software is bundled in the common software package of the top-quark analysis groups of DESY and the University of Hamburg called "TopAnalysis" [94] which is documented in [95].

5.2 Dataset and Trigger Choice

For this thesis, the full dataset of pp collisions at $\sqrt{s} = 8$ TeV centre-of-mass energy recorded in 2012 by the CMS detector is considered. This dataset corresponds to an integrated luminosity of $L = 19.7 \pm 0.5\,\mathrm{fb}^{-1}$ [96].
To speed up the analysis process, sub-datasets with complete offline reconstruction, muon and electron signatures in the detector and reduced event content are used. More precisely, the following single-lepton sub-datasets from the data-reprocessing campaign of 22$^{\mathrm{nd}}$ January 2013 for the offline reconstruction are considered for the analysis:

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

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

To ensure the full functionality of all detector components, all CMS datasets are checked within a two-step (online and offline) *Data Quality Monitoring* (DQM) validation process [97]. The certified list of good runs from the DQM process are collected in the following JSON file [98] which is used for the analysis:

- Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON

As first, basic selection step the lepton-related trigger paths are chosen as analysis input:

- SingleMu datasets: HLT_IsoMu24_eta2p1_v∗
- SingleElectron datasets: HLT_Ele27_WP80_v∗

The chosen trigger paths are unprescaled for the complete data-taking period and contain only a loose online selection applied to trigger objects to identify an electron with $p_T > 27\,\text{GeV}$ and some basic identification characteristics or a muon with $p_T > 24\,\text{GeV}$, $|\eta| < 2.1$ and a minimal isolation requirement. Therefore, the desired events with semileptonically decaying $t\bar{t}$ pairs with an electron or muon in the final state are selected with high efficiency and other processes are suppressed already from the beginning on.

5.3 Monte-Carlo Simulations

Monte-Carlo (MC) simulations model the complete chain of a physics process from the production over the decay, fragmentation and hadronisation process up to the finally measured signal in the detector. Therefore, they are a powerful tool for all data analyses to test different analysis setups, study the expected performance and finally compare the MC expectation to the measured result in data – either at reconstruction level or within a theoretically well defined generator level phase space.

5.3.1 Principles of Event Generation

The chain of MC simulation starts with the hard-scattering process where the desired final state is produced from the initial partons, e.g. $gg \to t\bar{t}$. This calculation is factorised into the perturbatively calculated *Matrix Element* (ME) of the process and the integration over the PDFs to address all possible initial partonic states. In general, the number of additionally produced partons from the calculation of the hard-scattering process is limited due to the computational speed and the complexity of the calculation. The next step of the simulation is the *Parton-Shower* (PS) process. Within the PS, additional *Initial- and Final-State Radiation* (ISR/FSR) is modelled to approximate higher order corrections not addressed in the ME. Furthermore, the PS models the evolution of the partons down to scales of $\alpha_S \approx 1$ using the *DGLAP* equations [99–101]. At these scales, the process of non-perturbative hadronisation starts.

If different tools are used to simulate ME and PS, a matching procedure is applied to avoid double counting of parton emission in the overlapping phase space between ME calculation and PS simulation.
Within the hadronisation and fragmentation process, the formation of colour neutral baryons and mesons is described using phenomenological models which are tuned to data. Additionally, also the decay of short lived-particles and resonances is simulated. The treatment of the *Underling-Event* (UE) activity, which is caused by the so called beam remnants [a], is done within data driven tunes. Finally, the physics final state is passed to a detailed detector simulation describing the interaction of the particles with the detector material and resulting in the ultimately measured digital signature which is used as input for the object reconstruction for every physics analysis. The whole chain of MC simulation is illustrated in Figure 5.1.

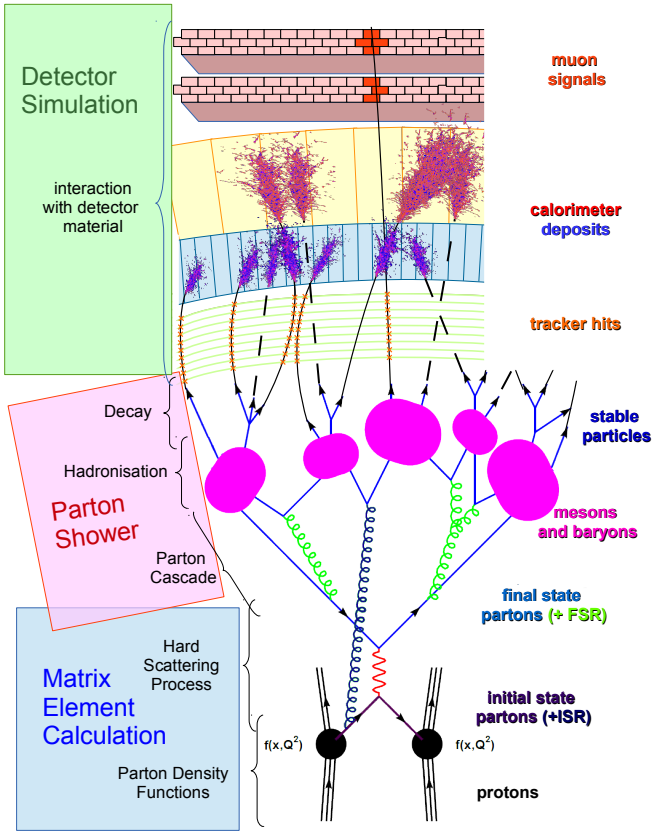

Figure 5.1: Illustration of the different steps of a physics process modelled within a Monte-Carlo event simulation for a QCD process. Taken from [88, 102, 103], merged and modified.

[a] Beam remnants are the remaining constituents of the initially colliding particles (here protons) which are not involved in the initial partonic state of the considered physics process and result in additional final-state particles or can lead to multi-parton interactions.

5.3.2 Simulated Samples

Several MC predictions are used within this thesis to optimise the performance of the analysis, address systematic uncertainties and compare the obtained data result with different theoretical approaches and approximations of the underlying perturbative QCD calculations. All simulated samples are generated within the official CMS *Summer12* MC production campaign [104].

Events with top-quark pairs as well as the production of W and Z bosons with additional jets and events with $t\bar{t}$ pairs accompanied by an additional vector boson (W,Z,γ) are simulated with the MADGRAPH event generator [105] (v5.1.5.11), involving MEs for $2 \rightarrow$ n processes at tree-level with up to n=9 particles in the final state. Therefore, additional hard radiation (dominantly from gluons)is already treated in the matrix element of the hard-scattering process and for $t\bar{t}$ events the production of up to three additional real partons is considered. These events are interfaced to PYTHIA [106] (v6.426) for the PS, involving the simulation of ISR and FSR, a transverse-momentum-ordered shower evolution and the modelling of the fragmentation and hadronisation process using the *Lund String model*. The k_T based MLM matching algorithm [107] is used to avoid double counting of parton emission in ME and PS.
In the following, the expression "$t\bar{t}$ (events)" refers only to the simulated $t\bar{t}$ processes without an additional vector boson. For these $t\bar{t}$ events, the MADSPIN [108] package was used to incorporate spin correlation effects. Additionally, tau decays are handled by TAUOLA [109] (v27.121.5).
Moreover, the production of single-top-quark and antitop quark events is generated in the s-, t- and tW-channel with the POWHEG event generator [110–112] (v1.0 r1380), which provides MEs at full NLO perturbative QCD precision and passed to PYTHIA for parton showering.
Events with two vector bosons (VV, i.e. WW, WZ, ZZ) are simulated with PYTHIA only, providing a ME description at LO and additional PS. Also the process of QCD multijet production is simulated using only PYTHIA. For the muon final state, multi-jet events with $\hat{p}_t > 20\,\text{GeV}$[b] are simulated. A pre-selection requires the presence of at least one muon on generator level with $p_T(\mu) > 15\,\text{GeV}$. This cut on generator level has an efficiency of $\epsilon_{gen} = 0.00037$. For the electron final state, two different filters have been used to enrich the QCD multi-jet samples with events including electrons. On the one hand the *BCtoE* filter preselects events with electrons from the decay of hadrons with b or c quarks. On the other hand the *EMEnriched* filter preselects events in which final-state particles might be reconstructed as electrons. In both cases there are three different samples for $\hat{p}_t$ from 20 to 30, 30 to 80 and 80 to 170 GeV, which all have different pre-selection efficiencies ($\epsilon_{gen}^{EM} = 0.0101,\ 0.0621,\ 0.1539$ and $\epsilon_{gen}^{BCE} = 0.00058,\ 0.00225,\ 0.0109$).

In addition, for certain systematic variations (e.g. the matching threshold between ME and PS, the choice of the hard-scattering scale (Q^2) etc.) specific samples from the *Summer12* production cycle are used.

[b] The variable $\hat{p}_t$ refers to the transverse momentum of the outgoing partons of the 2→2 hard-scattering process in their rest frame simulated by PYTHIA before parton showering.

In particular, additional $t\bar{t}$ signal samples are used for comparison with the data, involving also the MC@NLO [113, 114] (v3.41) and the POWHEG-BOX (v1.0) event generators for ME calculation at full NLO perturbative QCD precision. Moreover, HERWIG [115] (v6.520) is alternatively used for parton showering, the simulation of ISR and FSR with a transverse-momentum-ordered shower evolution and the modelling of the fragmentation and hadronisation process using the *cluster model*. Moreover, the low-statistics MADGRAPH+PYTHIA samples without MADSPIN are used in several parts of the analysis as independent test sample.
An overview about all details for the used generators can be found in Table 5.1.
To model the underlying event activity, the Z2* tune [116] is used for PYTHIA and the AUET2 tune [117] for HERWIG. Moreover, the JIMMY tool (v4.31) [118, 119] is used for HERWIG to generate multiple parton scattering.
For the *Parton-Density Functions* (PDFs), the CTEQ6 [120] and CT10 [121] PDF sets are used.

Generator	**Version used**	**PDF set for matrix-element calculation**	**Underlying-event tune**
MADGRAPH (+MADSPIN)	5.1.5.11	CTEQ6L1	Z2*
POWHEG-BOX ($t\bar{t}$)	1.0	CTEQ6M (with PYTHIA) CT10 (with HERWIG)	-
POWHEG-BOX (single-top-quarks)	1.0 r1380	CT10	-
MC@NLO	3.41	CTEQ6M	-
PYTHIA	6.426	-	Z2*
+TAUOLA	27.121.5		
HERWIG	6.520	-	AUET2
+JIMMY	4.31		

Table 5.1: Details of the Monte-Carlo event generators and simulation tools used for the generation of the official CMS *Summer12* Monte-Carlo production campaign samples, which are used within the analysis.

For all considered physics processes, the assumed pp production cross section at $\sqrt{s} = 8$ TeV, the corresponding MC event generators for ME calculation as well as for the PS and the number of generated events are listed in Table 5.2 for the central sample and in Table 5.3 and Table 5.4 for all samples addressing systematic variations.

All event yields from the MC prediction are normalised to the integrated luminosity in the first step. For this normalisation the expected pp production cross sections at $\sqrt{s} = 8$ TeV from theory calculations are used.
The cross-section for $t\bar{t}$ production corresponds to the latest perturbative QCD calculation at NNLO+NNLL accuracy [44] while the ones for single-top-quark and antitop-quark production correspond to a perturbative QCD calculation at approximate NNLO

Central MC Samples

Physics Process / MC sample	Matrix-Element Generator	Parton-Shower Generator	$N^{\text{MC Events}}_{\text{Generated}}$	σ [pb]
$t\bar{t}$	MadGraph +MadSpin	pythia	62,131,965	245.8
$t\bar{t} + \gamma$	MadGraph	pythia	1,719,954	1.8
$t\bar{t}$ + W	MadGraph	pythia	196,046	0.2
$t\bar{t}$ + Z	MadGraph	pythia	210,160	0.2
W$\rightarrow l\nu$	MadGraph	pythia	57,709,905	36,257.0
$Z/\gamma^* \rightarrow l^+l^-$ ($m_{ll} > 50\,\text{GeV}$)	MadGraph	pythia	30,459,503	3,503.0
t (s-channel)	powheg	pythia	259,961	3.8
t (t-channel)	powheg	pythia	3,758,227	56.4
t (tW-channel)	powheg	pythia	497,658	11.1
$\bar{t}$ (s-channel)	powheg	pythia	139,974	1.8
$\bar{t}$ (t-channel)	powheg	pythia	1,935,072	30.7
$\bar{t}$ (tW-channel)	powheg	pythia	493,460	11.1
QCD mu enr.	pythia	pythia	21,484,602	134,680.0
QCD EM enr., $\hat{p}_t$ 20-30	pythia	pythia	35,040,695	2,914,860.0
QCD EM enr., $\hat{p}_t$ 30-80	pythia	pythia	33,088,888	4,615,893.0
QCD EM enr., $\hat{p}_t$ 80-170	pythia	pythia	34,542,763	183,295.0
QCD BCtoE , $\hat{p}_t$ 20-30	pythia	pythia	1,740,229	167,388.0
QCD BCtoE , $\hat{p}_t$ 30-80	pythia	pythia	2,048,152	167,040.0
QCD BCtoE , $\hat{p}_t$ 80-170	pythia	pythia	1,945,525	12,982.0
WW	pythia	pythia	10,000,431	54.8
WZ ($m_{ll} > 12\,\text{GeV}$)	pythia	pythia	10,000,283	33.2
ZZ ($m_{ll} > 40\,\text{GeV}$)	pythia	pythia	9,799,908	8.1
Z'$\longrightarrow t\bar{t}$ ($m_{Z'} = 10^3\,\text{GeV}$, $\Gamma_{Z'} = 10^2\,\text{GeV}$)	MadGraph	pythia	104,043	5.0

Table 5.2: Detailed list of the considered physics processes (first column), the corresponding Monte-Carlo simulation (second to fourth column) and the assumed pp-production cross section σ for $\sqrt{s} = 8$ TeV (last column). Moreover, $m_{Z'}$ refers to the assumed mass and $\Gamma_{Z'}$ to the assumed width of the resonance Z'. All MC samples are generated within the official *Summer12* production cycle of CMS.

precision [47]. In addition, the cross sections for the production of a $t\bar{t}$ pair and a W boson [122] or Z boson [123] were calculated at NLO QCD accuracy while the one for the production of a $t\bar{t}$ pair and an additional photon [124] was calculated by multiplying the result obtained from WHIZARD [125] with an NLO k-factor [126].

Furthermore, the production cross sections of the QCD multijet samples are derived at LO precision by PYTHIA and include the pre-selection efficiency on generator level as described above.
Moreover, the production cross sections for vector bosons (accompanied by additional jets) have been derived at (N)NLO precision using MCFM [127] or FEWZ [128,129] and are taken from [130]. Additionally, the quoted production cross section for $Z/\gamma^* \to l^+l^-$ corresponds to $m_{ll} > 50\,\text{GeV}$ and the ones for WZ and ZZ to $m_{ll} > 10\,\text{GeV}$. For the $Z' \longrightarrow t\bar{t}$ process, which is used for pseudo data closure tests only, a relatively large cross section of 5 pb is used for a conservative estimation of the potential impact of such an event signature.
Finally, the simulation of the CMS detector is done by Geant4 [131,132] followed by the standard chain of data reconstruction: simulation of detector hits, signal digitisation and object reconstruction.

$t\bar{t}$ Systematic MC Samples

$t\bar{t}$ MC Sample	**Matrix-Element Generator**	**Parton-Shower Generator**	**$N^{\text{MC Events}}_{\text{Generated}}$**
alternative generator	POWHEG	PYTHIA	21,675,970
alternative generator	MC@NLO	HERWIG	32,852,589
alternative generator	POWHEG	HERWIG	27,684,235
increased ME-PS matching threshold	MADGRAPH	PYTHIA	37,083,003
decreased ME-PS matching threshold	MADGRAPH	PYTHIA	34,053,113
increased Q^2 scale	MADGRAPH	PYTHIA	41,908,271
decreased Q^2 scale	MADGRAPH	PYTHIA	39,286,663
m_{top} =178.5 GeV	MADGRAPH	PYTHIA	24,359,161
m_{top} =175.5 GeV	MADGRAPH	PYTHIA	40,244,328
m_{top} =173.5 GeV	MADGRAPH	PYTHIA	26,489,020
m_{top} =171.5 GeV	MADGRAPH	PYTHIA	24,439,341
m_{top} =169.5 GeV	MADGRAPH	PYTHIA	39,518,234
m_{top} =166.5 GeV	MADGRAPH	PYTHIA	27,078,777

Table 5.3: Detailed list of all $t\bar{t}$ Monte-Carlo samples incorporated to address systematic uncertainties within the analysis. All Monte-Carlo samples are generated within the official *Summer12* production cycle of CMS. All listed samples generated with MADGRAPH use the MADSPIN package.

Single-Top-Quark Production, Systematic MC Samples
– with increased Q^2 scale

MC Sample	Matrix-Element Generator	Parton-Shower Generator	$N^{\text{MC Events}}_{\text{Generated}}$
t (s-channel)	POWHEG	PYTHIA	1,998,845
$\bar{t}$ (s-channel)	POWHEG	PYTHIA	999,812
t (t-channel)	POWHEG	PYTHIA	1,945,116
$\bar{t}$ (t-channel)	POWHEG	PYTHIA	979,898
t (tW1-channel)	POWHEG	PYTHIA	1,492,816
$\bar{t}$(tW1-channel)	POWHEG	PYTHIA	1,492,534
t (tW2-channel)	POWHEG	PYTHIA	455,270
$\bar{t}$ (tW2-channel)	POWHEG	PYTHIA	497,676
t (tW3-channel)	POWHEG	PYTHIA	442,237
$\bar{t}$ (tW3-channel)	POWHEG	PYTHIA	497,376

– with decreased Q^2 scale

MC Sample	Marix-Element Generator	Parton-Shower Generator	$N^{\text{MC Events}}_{\text{Generated}}$
t (s-channel)	POWHEG	PYTHIA	1,945,749
$\bar{t}$ (s-channel)	POWHEG	PYTHIA	999,849
t (t-channel)	POWHEG	PYTHIA	1,951,907
$\bar{t}$ (t-channel)	POWHEG	PYTHIA	979,359
t (tW1-channel)	POWHEG	PYTHIA	1,493,130
$\bar{t}$ (tW1-channel)	POWHEG	PYTHIA	1,493,101
t (tW2-channel)	POWHEG	PYTHIA	496,818
$\bar{t}$ (tW2-channel)	POWHEG	PYTHIA	497,682
t (tW3-channel)	POWHEG	PYTHIA	453,233
$\bar{t}$ (tW3-channel)	POWHEG	PYTHIA	497,674

Table 5.4: Detailed list of all Monte-Carlo samples for single-top-quark production incorporated to address systematic uncertainties within the analysis. The tW-production channel is slitted in three subsamples (tW1-tW3). All Monte-Carlo samples are generated within the official *Summer12* production cycle of CMS.

5.3.3 Pile-Up Modelling

At the high LHC luminosity, there is a high probability that one single bunch crossing may produce several separate events. These events are called *Pile-Up* (PU) events. As the detector a priori cannot distinguish the signal-like event which fired the trigger and the additionally produced pile-up events, the measured signal in the detector is a combination of both. Especially within reconstructed jets, additional particles produced in PU events are clustered and a special treatment is therefore necessary (see Chapter 6.4.2).
For the 2012 pp collision data at $\sqrt{s} = 8$ TeV used for this analysis an average number of about 20 PU interactions per event is expected. Therefore, the correct modelling of the underlying PU distribution as well as the correction of the measured objects for pile-up effects are essential for every LHC data analysis.

While generating the MC predictions, PU events are added following the description in [133]. However, this admixture is only generic and the PU distribution differs from the one actually estimated from collision data[c] which is shown in Figure 5.2.
Thus, individual PU event weights are calculated taking the generic PU distribution from simulation and the measured PU distribution from pp collision data as input [136]. Following the official PU correction prescription allows to weight each MC event such that the finally obtained PU distribution of the MC prediction and the one from data coincide. The simulated PU distribution is shown in Figure 5.3 before (left) and after (right) applying the correction. In order to verify the correctness of the PU

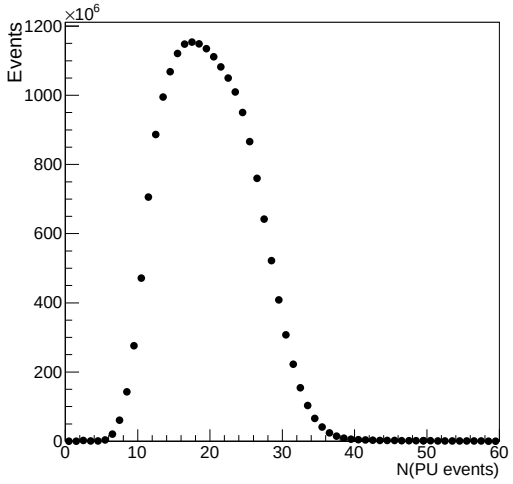

Figure 5.2: Reference pile-up distribution extracted from 2012 proton-proton collision data at $\sqrt{s} = 8$ TeV.

correction procedure the distribution of reconstructed primary vertices between data and MC prediction is compared. This quantity is directly related to the number of PU

[c]The PU distribution from data can be estimated following the method and tools described in [134, 135].

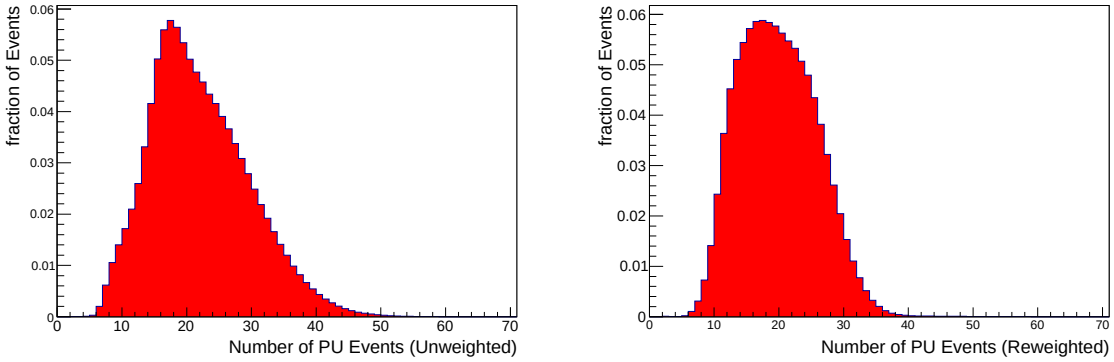

Figure 5.3: Initial pile-up distribution for Monte-Carlo samples before (left) and after (right) applying the official PU correction prescription. Both distributions are normalised by their integral.

interactions. As illustrated in Figure 5.4 the distributions agree reasonably for both tested selection steps of the analysis– before and after requiring the events to contain b-tagged jets. The distributions are individually normalised by their integral to focus on the description of the shape.
The hatched uncertainty band in Figure 5.4 illustrates the assumed uncertainty on the PU modelling. Following the official recommendation [137] a ±5% variation of the total inelastic pp cross-section which CMS determines to be 69.4 mb is performed to address this uncertainty.

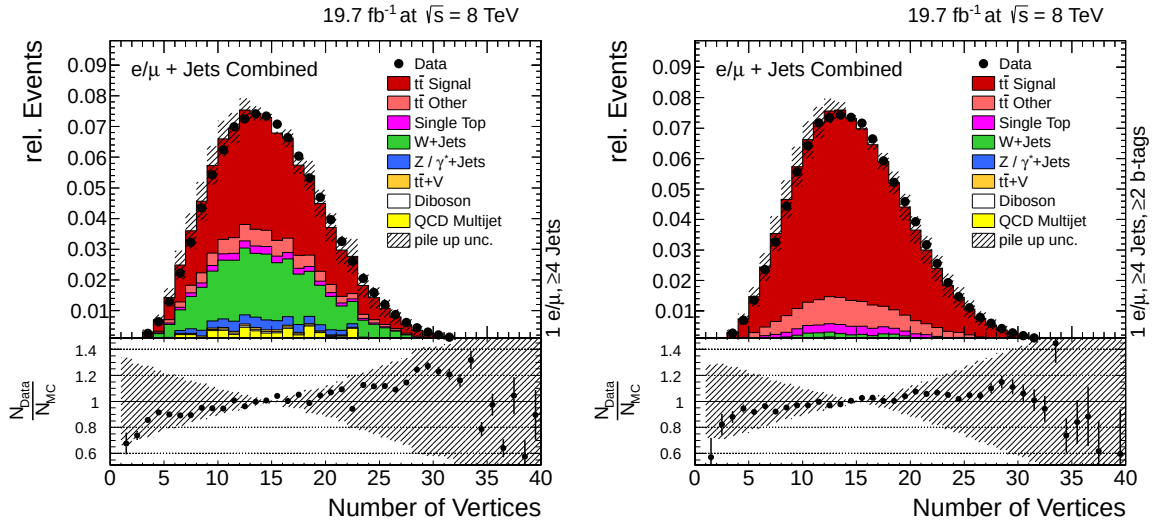

Figure 5.4: Normalised distribution of the reconstructed primary vertices for the combined e/μ+jets $t\bar{t}$ final state with (left) and without (right) requiring the events to contain b-tagged jets. The PU correction prescription is applied to the Monte-Carlo prediction. The hatched band corresponds to the uncertainty of the pile-up modelling.

Chapter 6

Object Definition and Event Selection

As discussed in Chapter 3, the $t\bar{t}$ events desired for the analysis are initially accompanied by a large amount of other physics processes with partially much larger production cross sections (see Table 5.2).
The expected signature for semileptonic $t\bar{t}$ final states within the detector is one lepton, missing transverse energy from one neutrino and four jets stemming from two light quarks and two b quarks. Therefore, the following event selection is applied to suppress other physics processes and select a high purity $t\bar{t}$ sample:

- exactly one isolated[a] high energetic muon or electron
  $\longrightarrow$ to suppress QCD multijet events
- no other loosely isolated muons or electrons
  $\longrightarrow$ to suppress Z/γ^*+jets events
- at least four high energetic jets
  $\longrightarrow$ to suppress W-boson+jets events
  - at least two jets associated to b quarks
    $\longrightarrow$ to suppress all background processes with light flavour content
  - at least two jets not associated to b quarks
    $\longrightarrow$ needed for the full event reconstruction

The applied selection of top-quark pairs follows with the common reference selection for the semileptonic decay channel as recommended by the *Top Physics Analysis Group* (TopPAG) described in [138]. Minor modifications ensure equal kinematic acceptances for the selected muon and electron samples. For the kinematic event reconstruction of the top quarks at least two jets not associated to b quarks are needed to reconstruct the hadronically decaying W boson.

[a] Muons and electrons are phenomenologically named "isolated" if they are produced by the decay of a Z or W boson and not within the cascade of a jet. Experimentally, the isolation of a muon or electron is defined by the amount of energy deposited around the lepton track.

6.1 Particle Flow and General Concepts

In order to reconstruct physics objects from the measured detector signals, several reconstruction steps have to be applied.

Within the track reconstruction procedure the various hits in the different parts of the CMS tracker system are combined using a combinatorial track finder [139] which is based on a Kalman Filter [140]. The result of this iterative procedure is a collection of reconstructed tracks which is used further on e.g. to identify vertices.
For each event, a *Primary Vertex* (PV) corresponding to the pp interaction point is demanded, which is reconstructed from a minimum of four tracks within a longitudinal distance of $|z| < 24\,\text{cm}$ along the beam direction and a distance of $\rho < 2\,\text{cm}$ to the nominal interaction point (beam spot) in the transverse interaction plane. Because of the spread of the individual protons in the colliding bunches, beamspot and PV are not identical. In case of several reconstructed vertices, the one with the highest p_T^2 sum of all associated tracks is considered as PV. Secondary vertices are e.g. used to identify jets originating from b quarks.
Tracks not originating from the selected primary vertex or its associated secondary vertices are considered as PU events (see Chapter 5.3.3) and are discarded from the analysis. This step is called *Charged Hadron Subtraction* (CHS).

Single particles are then reconstructed combining reconstructed tracks and vertices, energy deposits in the calorimeters and signals in the muon chambers. All objects used throughout this analysis are reconstructed using the *Particle-Flow* (PF) algorithm for event reconstruction [141]. The PF technique attempts to maximise the amount of detector information used for the object reconstruction and treats the event as a whole. The aim of the PF algorithm is the individual reconstruction of every stable particle, considering muons, electrons, photons, charged and neutral hadrons. These particles are called *PF candidates*. An illustration of the PF concept can be found in Figure 6.1. In general, the PF algorithm consists of the following steps [142]:

- clustering of energy deposits in the calorimeters
- track reconstruction and extrapolation of the tracks to the calorimeters
- muon identification
- electron pre-identification
- linking of topologically connected elements
- particle identification and reconstruction

The CMS detector is especially suitable for the application of the PF technique due to its high-performance silicon tracker, the good separation power of the high magnetic field, the high granularity of the ECAL and the hermeticity of ECAL and HCAL.
Moreover, the list of individual particles from the PF procedure is then successively used to build jets, to determine the missing transverse energy, to reconstruct and identify tau leptons from their decay products and to identify b jets. It has been proven that especially the performance of jets, E_T^{miss} and the isolation of charged leptons can be significantly improved using the PF algorithm.

To avoid the overlap of jets and leptons, so-called *Top Projections* are used, i.e.

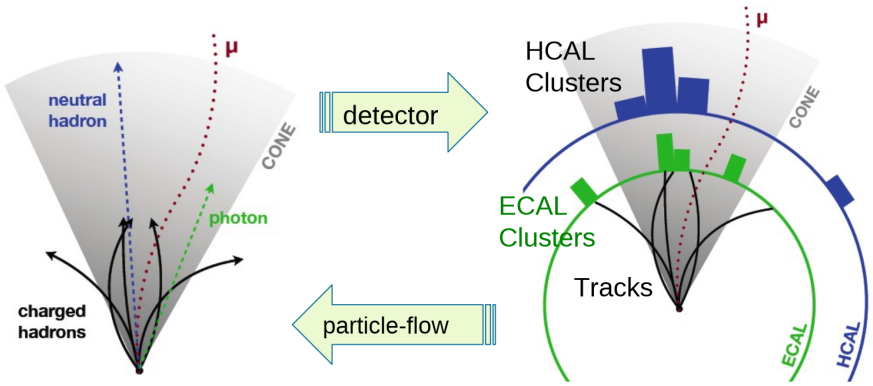

Figure 6.1: Illustration of the particle-flow concept to reconstruct every individual muon, electron, photon, charged and neutral hadron from a maximum of detector information. Taken from [143].

loosely isolated PF electrons and muons are excluded from the jet clustering step. This follows the prescription of the *PF2PAT* algorithm described in [144].

6.2 Muons

In general, muons in CMS are reconstructed from a track in the inner tracker and/or in the muon system as detailed in [87].
Due to the unique signature in the muon system, the low interaction with the detector material and the superior resolution of the tracking system, the muon is one of the objects that can be reconstructed best in CMS.

6.2.1 Signal Muon Selection

For the $t\bar{t}$ muon+jets final state, every event is required to contain exactly one offline reconstructed muon with $p_T > 33\,\text{GeV}$ and $|\eta| < 2.1$. The cut on $|\eta|$ is chosen to stay within the plateau region of trigger efficiency and scaling factor to avoid introducing large uncertainties while the p_T value is chosen to be synchronous with the electron definition.

Moreover, the muon is required to pass the *Global Muon* requirement as described in [145]. For a good identification and a precise determination of the momentum, certain track quality requirements are applied. Therefore, the muon is required to have hits on more than five layers of the inner tracking system including at least one valid hit in the pixel detector. Furthermore, the muon is requested to have at least one valid hit in the muon system. The global muon track should be matched to the outer track segments of at least two stations of the muon system and the χ^2/ndof[b] for the global track fit

[b] The abbreviation ndof is commonly used for the *number of degrees of freedom*.

should be less than 10. To assure that the muon originates from the initial pp collision, the two-dimensional impact parameter d_B with respect to the beam spot is required to be smaller than 0.02 cm and the distance between the z position of the origin of the muon track and the z component of the primary vertex is required to be smaller than 0.5 cm.

In order to exclude leptons produced in semileptonic decays of hadrons within a jet, an isolation criteria is imposed. A relative lepton isolation I_{rel}^{PF} (ΔR^{max}, lep) is calculated. It is defined as the summed p_T of all neutral and charged PF hadrons and PF photons within an opening cone of ΔR^{max} around the selected lepton lep divided by the p_T of the lepton itself:

$$I_{rel}^{PF}(\Delta R^{max},\ \text{lep}) = \frac{\sum_{\Delta R<\Delta R^{max}} p_T^{\text{charged hadrons}} + p_T^{\text{neutral hadrons}} + p_T^{\gamma}}{p_T^{\text{lep}}}. \tag{6.1}$$

For muons, a cone of ΔR^{max}=0.4 is chosen. To remove the contribution of particles from PU interactions, only charged hadrons from the PV are considered (see CHS in Chapter 6.1). As neutral particles do not have a track and therefore cannot be assigned to a vertex, the summed p_T of the photons and neutral hadrons is corrected by subtracting half of the summed p_T of charged particles assigned to PU interactions. This is a good estimate because on average about $\frac{2}{3}$ of the particles produced in a particle interaction are charged hadrons.
Finally, a muon is considered to be isolated if I_{rel}^{PF} (ΔR^{max}=0.4, lep=μ)<0.12.

6.2.2 Additional Muon Veto

As only exactly one muon is expected in the $t\bar{t}$ semileptonic final state, events with additional muons are vetoed. For these PF muons, looser identification and isolation criteria are imposed.

All veto muons are required to be identified as *Global* or *Tracker* muons as described in [145] with a minimal transverse momentum of $p_T > 10$ GeV in the central region $|\eta| < 2.5$ of the detector. Furthermore, a minimum isolation of I_{rel}^{PF} (ΔR=0.4)<0.2 is required.

6.3 Electron Selection

The basic reconstruction of electrons in CMS is based on the hits in the tracking system and the energy deposits in the ECAL as detailed in [146]. Consequently, the identification of electrons is more difficult than for muons as the distinction from other particles is more complicated.
Furthermore, the lower mass of the electron leads to a sizeable amount of emitted photons from bremsstrahlung produced from interaction with the tracker material. This has to be considered within the reconstruction process to re-assign the radiated

photons and take into account potential kinks in the reconstructed electron trajectory. Technically, this is done using a *Gaussian Sum Filter* algorithm [147].
To distinguish electrons from other particles, an MVA-based pre-identification is applied within the PF procedure using shower shape, track quality and kinematic variables as well as a criteria for the momentum-energy matching between the tracker and the ECAL, which is dominated by bremsstrahlung effects.

6.3.1 Signal Electron Selection

For the $t\bar{t}$ electron+jets final state, the event is required to contain exactly one reconstructed electron with $E_T > 33\,\text{GeV}$[c] and $|\eta| < 2.1$. The cut on $|\eta|$ is chosen to be synchronous with the muon definition while the E_T cut is chosen to stay within the well described plateau region of the trigger efficiency.

Moreover, the electrons with an ECAL supercluster in the eta range of 1.4442 to 1.5660, corresponding to the transition region of the barrel and end-cap calorimeter, are not taken into account. Additionally, the electron track should point back to the interaction point and the impact parameter d_B is required to be smaller than 0.02 cm with respect to the beam spot. The distance between the longitudinal position of the track origin and the primary vertex is required to be smaller than 0.5 cm. Furthermore, a multivariate electron identification algorithm is used, which is e.g. constructed from shower shape variables and finally the value of the MVA electron ID is required to be above 0.5. In addition, electrons stemming from photon conversion are rejected. They are identified by missing hits in the layers of the inner tracking system and by a conversion veto [148].

Similar to the muon channel, an isolation requirement is imposed to veto electrons produced in non-prompt decays. Following the definition in Equation 6.1, a minimum relative isolation within a cone of ΔR^{max}=0.3 around the electron trajectory is derived. An electron is finally considered as isolated, if I_{rel}^{PF} (ΔR^{max}=0.3, lep=e)<0.1.

6.3.2 Additional Electron Veto

To suppress background processes with multi-lepton final states, events with additional loosely isolated electrons are vetoed.
Hence, events do not pass the event selection if they contain additional electrons with $E_T > 20\,\text{GeV}$, $|\eta| < 2.5$, an MVA electron ID value above 0 and and an isolation of I_{rel}^{PF} (ΔR^{max}=0.3, lep=e)< 0.15.

[c] At energies of at least a few GeV used within the analysis, momentum (p) and energy (E) are essentially identical while they differ in principle for energy regimes around the particle mass m. The general relation is: $E^2 = p^2 + m^2$.

6.4 Jets

From the semileptonic $t\bar{t}$ event topology, at least four quarks are expected in the final state. Due to QCD confinement (see Chapter 3.2.3), initially produced partons and gluons finally result in a collimated bunch of particles measured in the detector. Collecting these particles produced by a parton or gluon from the hard scattering process in the shower evolution and hadronisation process (see Chapter 5.3) in one jet is one of the most challenging tasks of the object reconstruction.

Within the PF algorithm, jets are clustered from the reconstructed PF candidates. For this analysis, loosely isolated leptons (as defined for the electron and muon vetoes) are excluded from the particle list for clustering.
The clustering itself is done using the infrared and collinear save sequential anti $-k_T$ clustering algorithm as detailed in [149]. A recombination parameter of $R = 0.5$ representing the size of the final jet cone is chosen.

6.4.1 Jet Calibration

The four-momentum of the initial parton or gluon is reflected by the sum of the four-momenta of all jet constituents. However, the momentum of the reconstructed jets does not correspond to the true particle-level energy obtained from gen jets clustered from stable particles on generator level. The main reasons for this difference are the non-uniform and non-linear response of the calorimeter, energy losses from particles outside the jet area or undetected neutrinos or additional clustered particles e.g. from PU interactions. Hence, a calibration of the measured raw jet momenta is needed [48, 150].

These *Jet Energy Corrections* (JECs) are centrally derived by the CMS *JetMET* Physics Analysis Group [151] using MC predictions as well as data from di-jet, Z+jet and γ+jet samples [152,153]. CMS follows a factorised approach involving the following

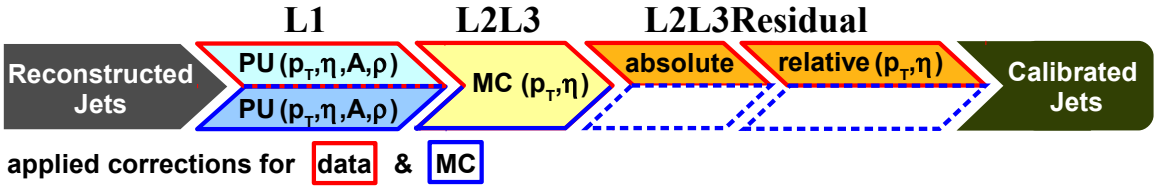

Figure 6.2: Illustration of the factorised approach of Jet Energy Calibration in CMS.

correction steps as illustrated in Figure 6.2:

- **L1 (data&MC):** p_T and η dependent correction for the contribution of underlying PU processes using the concept of jet areas [154, 155] with the median of the energy density (ρ) and the jet area (A)
- **L2L3 (data&MC):** p_T and η dependent corrections to compensate for the non-linear and non-uniform response of the calorimeters, derived from MC predictions,

corrected to generator jet level[d]

- **L2L3Residual (data only):** correct for remaining differences of data with respect to MC predictions, derived from the MC prediction to data ratio

For this analyis, jets have been calibrated up to the absolute *Jet Energy Scale* (JES) of *L3* in simulation and *L2L3Residual* in data using the official *Summer13_V4 CMS JECs* as detailed in [156]. Furthermore, the jet energy resolution[e] is found to be slightly larger in data than in simulation. Therefore, the energy of the reconstructed jets in simulation is smeared accordingly to match the data using the JER-SFs detailed in Table 6.1.

jet η range	JER-SF
$0.0 < \eta_{\text{jet}} < 0.5$	$1.052^{+0.063}_{-0.062}$
$0.5 < \eta_{\text{jet}} < 1.1$	$1.057^{+0.057}_{-0.056}$
$1.1 < \eta_{\text{jet}} < 1.7$	$1.096^{+0.065}_{-0.064}$
$1.7 < \eta_{\text{jet}} < 2.3$	$1.134^{+0.094}_{-0.092}$
$\eta_{\text{jet}} > 2.3$	$1.288^{+0.200}_{-0.199}$

Table 6.1: Scale factors for the jet energy resolution to be applied to the resolution obtained from simulation as provided by the CMS *JetMET* Physics Analysis Group.

6.4.2 Jet Selection

Events are requested to contain at least four high energetic jets in the central region of the detector by imposing $p_T^{\text{jet}} > 30\,\text{GeV}$ and $|\eta^{\text{jet}}| < 2.4$.
To reject fake jets from instrumental noise or other particle signatures, common PF jet identification requirements are applied to the selected jets. These requirements are minimalistic and based on the composition of the reconstructed jet:

- charged hadron energy fraction > 0
- neutral hadron energy fraction < 0.99
- electron energy fraction < 0.99
- photon energy fraction < 0.99
- multiplicity of charged particles > 0
- number of clustered PF candidates > 1

[d] Generator level jets are obtained from applying the same cluster algorithm as for reconstruction level jets on the stable particles of the event. In this context, stable particles are all generated particles of the event with a lifetime, which is long enough to transverse or leave a signal in the detector.

[e] In general, the resolution of a quantity x is defined as the width of the relative response distribution $\frac{x_{\text{reco}} - x_{\text{true}}}{x_{\text{true}}}$ of the reconstructed value x_{reco} over the true value x_{true}. More details on the definition and determination of the jet energy resolution can be found in [33].

6.5 Identification of b Jets

The majority of the events produced in pp collisions contain jets in the final state, while the presence of jets originating from b quarks is much less frequent. Therefore, identifying those jets initiated from b quarks is a powerful tool to separate events from $t\bar{t}$ production with at least two expected final-state b jets from other physics processes. The semileptonic $t\bar{t}$ final state involves also two non-b jets. For the full kinematic reconstruction of the event, the information of the b jet identification procedure is used to assign the reconstructed jets to the underlying top quarks initially produced. Therefore, events are required to contain at least two jets identified as b jets and two jets identified as non-b jets.

6.5.1 Principles

The procedure of b-jet identification (often called *b-tagging*) is based on the lifetime, mass and the decay of the B hadrons which are formed from the final-state b quark within the hadronisation process. With $\tau_B \approx \mathcal{O}(1\text{ps})$ lifetime, B hadrons can travel a short distance (typically $\approx \mathcal{O}(1\text{mm})$) before decaying. Therefore, the tracks of the decay products, which are finally measured in the detector, are extrapolated back to this point, forming a secondary vertex (displaced from the primary one). Hence, the identification of b jets as illustrated in Figure 6.3 is based on reconstructed secondary vertices and track related information such as the impact parameter with respect to the primary vertex. Also the relatively large mass of the B hadron in comparison to other particles in the shower and the possible decay into leptons can be used in the b-tagging procedure.

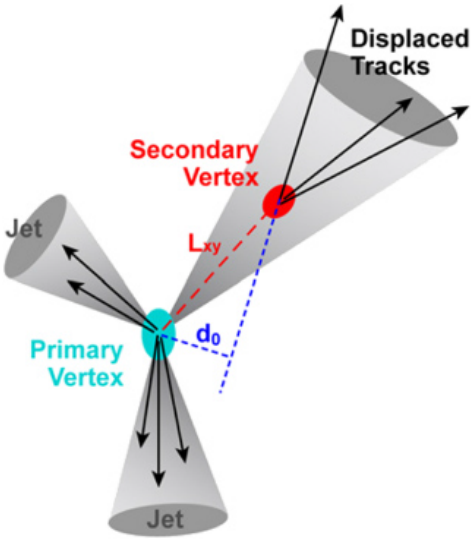

Figure 6.3: Illustration of the detector signal of a reconstructed b jet. L_{xy} denotes the flight distance of the B hadron in the transverse plane from the primary to the secondary vertex where it decays and creates tracks with a large impact parameter d_0 with respect to the primary vertex. Taken from [157].

Differences between data and simulation are corrected, using per jet SFs for the b-tag and mistag efficiencies provided by the CMS *B Tag & Vertexing Physics Object Group* (BTV POG) [158]. The conversion of these per jet SFs into an per event SF is explained in detail in Chapter 6.7.

6.5.2 B-Jet & Light-Jet Selection

Within the analysis the *Combined Secondary Vertex (CSV)* algorithm as detailed in [159] is used for b-jet identification. The CSV algorithm combines information about impact parameter significance of tracks, the reconstructed secondary vertex and jet kinematics in an MVA approach [160] into one discriminator value b_{disc}.

Jets with a b-jet discriminator value of $b_{\text{disc}} > 0.679$ are identified as b jets. This value refers to the medium working point of the algorithm (CSVM), which implies that only about 1.5% of the light jets are on average misidentified as b-jets while the average identification efficiency of b jets is about 65%. Following the $\mathrm{t\bar{t}}$ final-state topology, events are requested to contain at least two jets (as defined in Chapter 6.4.2) identified as b jets.

Furthermore, events are only considered for the analysis if they in addition contain at least two jets identified as non-b jets by inverting the discriminator criteria. This additional selection step was introduced to differentiate between events where the kinematic reconstruction is not successful because no light jets are found for the association with the hadronically decaying W boson and events where the kinematic reconstruction is not converging due to the incompatibility with the kinematic constraints.

6.6 Missing Transverse Energy

The undetected neutrino of the $\mathrm{t\bar{t}}$ semileptonic final state leads to a momentum imbalance of all measured particles in the transverse plane. The negative vectorial sum of the momenta of all PF candidates in the transverse plane is referred to as missing transverse momentum $\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}$ and its magnitude as missing transverse energy $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$:

$$\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{\substack{\text{all PF}\\ \text{candidates i}}} \vec{\mathrm{p}}_{\mathrm{T}}^{\,\mathrm{i}} \tag{6.2}$$

$$\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}} = |\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}|. \tag{6.3}$$

$\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}$ is a measure for the neutrino momentum in the transverse plane and therefore used in the kinematic event reconstruction (see Chapter 7) while no explicit event selection requirement is imposed. The reconstruction of $\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}$ is very sensitive to object mismeasurements, instrumental noise or detector malfunctions. Several corrections are applied during the $\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}$ computation as detailed in [161, 162], leading to resolutions of typically 20 GeV. For the analysis, "type-0" and "type-1" corrected $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ is used,

involving a PU correction and the propagation of the jet energy corrections (see Chapter 6.4.1).

6.7 Data-Driven Corrections

In order to use the maximal knowledge as input of the analysis, several correction factors are applied to the plain simulation:

- W→x *Branching Ratio* (BR) correction
- PU correction (official CMS procedure)
- trigger and lepton selection efficiencies (customised study on data)
- JEC (official CMS corrections)
- b-jet identification (official CMS scale factors)

Some of these *scale factors* (SFs) are directly provided from the corresponding *Physics Object Group* (POG) while others follow official prescriptions.

Branching Ratio

The BR for the decay of the W bosons which is used in the plain MadGraph+pythia $t\bar{t}$ prediction is of LO accuracy. The BR values for the W-boson decay as measured in data from the *Particle Data Group* (PDG) [32] are used to correct the different $t\bar{t}$ final states accordingly. The observed effect and an associated uncertainty are completely negligible.

Pile-Up

The generic PU distribution in the simulation is corrected according to the one observed in data. The method is detailed in Chapter 5.3.3 and finally, good agreement between data and simulation is found within uncertainties.

Trigger and Lepton Efficiency

Trigger and Lepton efficiencies have been determined individually for each final state using the common *Tag-and-Probe* (T&P) method. The strategy and technical implementation is taken from [23] and applied to the $\sqrt{s} = 8$ TeV dataset. The advantage of this customised determination of the efficiencies is that the applied lepton definition is identical to the one used for the event selection (see Chapter 6.2.1 and 6.3.1). Furthermore, the efficiencies can be determined differentially for any quantity necessary and in any binning. This is important as for the determination of normalised differential distributions a differential understanding of the efficiencies is needed.

The T&P method uses a sample of high purity leptons from Z→ll decays. From a sample with at least two oppositely charged lepton candidates in the considered kinematic phase space, one of them (*tag lepton*) is demanded to satisfy all the lepton selection criteria and to have fired the trigger. Furthermore, the second lepton candidate is only required to fulfill the basic PF object reconstruction. It is finally accepted as a *probe lepton* with a low misidentification probability if the combined invariant mass of the two leptons lies within 15 GeV around the Z boson mass of 91 GeV. From

this, relative efficiencies ϵ for a certain selection S with respect to previously applied selections S' are determined with the cut-and-count method:

$$\epsilon_{\mathrm{S[S']}} = \frac{\mathrm{N}^{\mathrm{probe}}_{\mathrm{S}\otimes\mathrm{S}'}}{\mathrm{N}^{\mathrm{probe}}_{\mathrm{S}'}}. \tag{6.4}$$

Here, $\mathrm{N}^{\mathrm{probe}}_{\mathrm{S}'}$ denotes the number of probe leptons passing the pre-selections S$'$ and $\mathrm{N}^{\mathrm{probe}}_{\mathrm{S}\otimes\mathrm{S}'}$ the number of leptons passing in addition selection S.

Moreover, the lepton efficiencies and scale factors are determined in a factorisation approach. Starting from probe leptons fulfilling the basic PF object reconstruction (labeled *reco*), the lepton selection (including identification and isolation cuts) and trigger requirements are applied subsequently:

$$\epsilon_{\mathrm{l[reco]}} = \epsilon_{\mathrm{sel[reco]}} \cdot \epsilon_{\mathrm{trig[sel\otimes reco]}} \tag{6.5}$$

$$\mathrm{SF_l} = \frac{\epsilon^{\mathrm{data}}_{\mathrm{l[reco]}}}{\epsilon^{\mathrm{MC}}_{\mathrm{l[reco]}}}. \tag{6.6}$$

Following the factorisation approach, the final scale factors $\mathrm{SF_l}(\mathrm{p}^{\mathrm{l}}_{\mathrm{T}}, \eta^{\mathrm{l}}) = \frac{\epsilon(\mathrm{data})}{\epsilon(\mathrm{MC})}$ for trigger and lepton selection efficiency are multiplied i.e. by taking all probe leptons before any trigger and lepton selection as denominator and taking the leptons that pass these requirements as the numerator.

In 2012, single-lepton triggers without prescale are available for the complete data-taking period. For the semileptonic muon (electron) final state, the trigger HLT_IsoMu24_eta2p1_v∗(HLT_Ele27_WP80_v∗) is applied to data and the MC prediction. The lepton selection criteria are defined in Chapter 6.2.1 for the muon and in Chapter 6.3.1 for the electron. The finally obtained T&P efficiencies and SFs for trigger and lepton selection are shown in Figure 6.4 for the muon and in Figure 6.5 for the electron. Both SFs are determined double differentially as a function of $\mathrm{p}^{\mathrm{l}}_{\mathrm{T}}$ and η^{l} of the lepton. Statistical uncertainties of the determined efficiencies refer to Clopper-Pearson confidence intervals [163].

In general, reasonably flat SFs in the order of 0.98 ± 0.03 are obtained. Only a moderate η dependence is observed for the muon and a moderate $\mathrm{p_T}$ dependence for the electron. Only the SFs for the kinematic region with electrons below $\mathrm{p_T} = 35\,\mathrm{GeV}$ shows larger discrepancies. For $30\,\mathrm{GeV} \geq \mathrm{p_T(electron)} < 33\,\mathrm{GeV}$ a large η(electron)-dependency for the SF of the trigger efficiency is observed. To avoid large uncertainties, a minimum requirement of $\mathrm{p_T} > 33\,\mathrm{GeV}$ is used for electrons within the analysis. Further details and one-dimensional projections of the two-dimensional efficiencies as function of $\mathrm{p}^{\mathrm{l}}_{\mathrm{T}}$ and as function of η^{l} can be found in Appendix A.

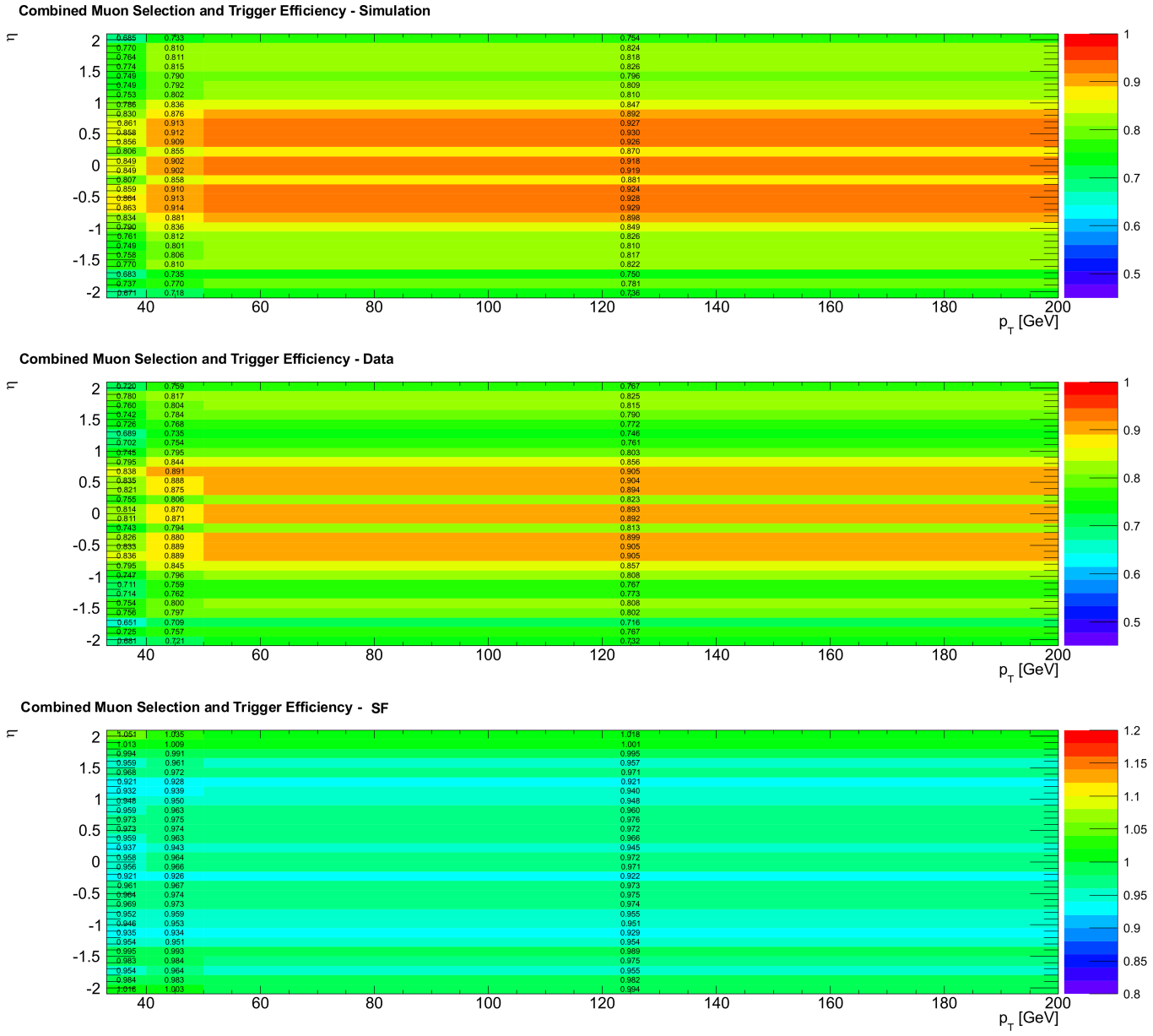

Figure 6.4: Combined trigger and lepton selection efficiency as obtained from a tag-and-probe study for the muon in the MC prediction (top) and the data (middle) as a function of p_T and η of the muon. The resulting data over MC scale factor is shown in the bottom. Taken from [164].

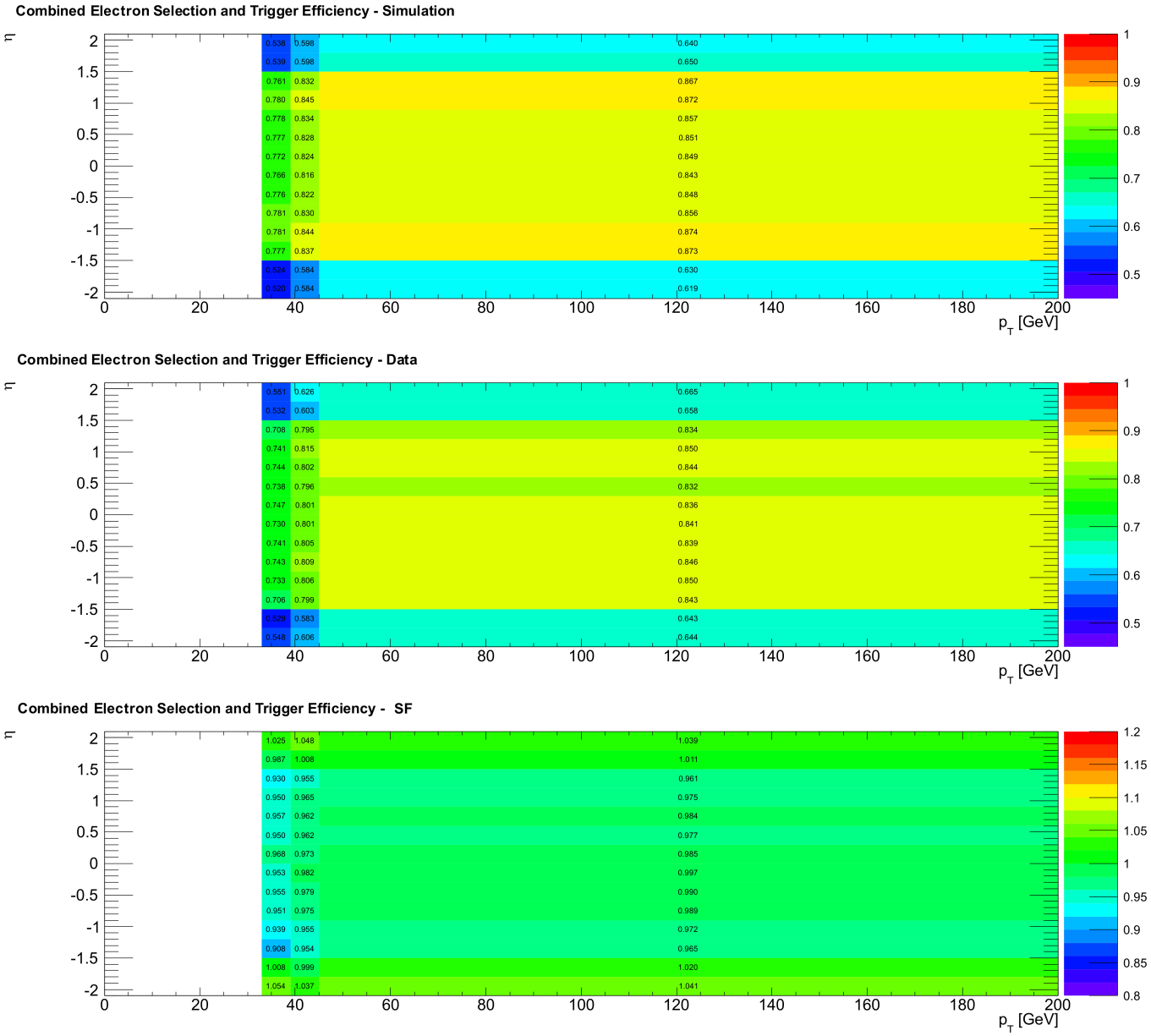

Figure 6.5: Combined trigger and lepton selection efficiency as obtained from a tag-and-probe study for the electron in the MC prediction (top) and the data (middle) as a function of p_T and η of the electron. The resulting data over MC scale factor is shown in the bottom. Taken from [164].

B-Jet identification

The efficiency for b-jet identification and light-jet[f] misidentification is found to be different in data and simulation by the CMS BTV POG [159, 165]. Per jet SFs parametrised as a function of p_T and η of the jet for the kinematic range $|\eta|$ <2.4 and $p_T > 20\,\text{GeV}$ are available. The SF for the b-jet identification efficiency per jet for the CSVM algorithm is measured to be between 0.94 and 0.97 with uncertainties of 2-4% as a function of the jet p_T in the range of $30 - 500\,\text{GeV}$.
The conversion of the per jet SFs into an per event SF follows a prescription which was also used for the analysis of differential $t\bar{t}$ cross sections with the CMS dataset recorded at $\sqrt{s} = 7$ TeV [23, 25].

In the first step, the efficiencies for b-jet identification are estimated from the MC prediction separately for each flavour (ϵ_b, ϵ_c and ϵ_l) as a function of the jet p_T and $|\eta|$ from the default $t\bar{t}$ signal sample (see Chapter 5.3.2) after the full lepton and jet selection without b-jet identification requirements. The number of simulated events in this sample is large enough to ignore the statistical uncertainty.

The corrected b-jet identification efficiencies ($\epsilon_{i,corr}$) can then be calculated by applying the provided SFs for the different jet flavors (SF_i):

$$\epsilon_{i,corr} = SF_i \cdot \epsilon_{i,sim} \quad , \quad i \in b, c, l. \tag{6.7}$$

Following the recommendation, the SF for b-jet identification is also used to correct the efficiency for c-jet misidentification with twice the quoted uncertainty. Furthermore, the SF for light-jet misidentification is measured to be between 0.95 and 1.45 with uncertainties of 8-15% as function of jet p_T in different jet η ranges.

Moreover, an event selection efficiency for at least two jets identified as b jets ($\epsilon_{\geq 2\text{ b jets}}$) is calculated for each MC event. As shown in Equation 6.8, this is done by subtracting from 1 the efficiency that no jet (1st row) or only one b-jet (2nd row), c-jet (3rd row) or light-jet (4th row) is identified as b-jet:

$$\begin{aligned}
\epsilon_{\geq 2\text{ b jets}} = 1 \; & - \prod_{i}^{N_b}(1-\epsilon_{b,i}) \prod_{j}^{N_c}(1-\epsilon_{c,j}) \prod_{k}^{N_l}(1-\epsilon_{l,k}) \\
& - \sum_{h}^{N_b} \epsilon_{b,h} \prod_{i\neq h}^{N_b}(1-\epsilon_{b,i}) \prod_{j}^{N_c}(1-\epsilon_{c,j}) \prod_{k}^{N_l}(1-\epsilon_{l,k}) \\
& - \sum_{h}^{N_c} \epsilon_{c,h} \prod_{i}^{N_b}(1-\epsilon_{b,i}) \prod_{j\neq h}^{N_c}(1-\epsilon_{c,j}) \prod_{k}^{N_l}(1-\epsilon_{l,k}) \\
& - \sum_{h}^{N_l} \epsilon_{l,h} \prod_{i}^{N_b}(1-\epsilon_{b,i}) \prod_{j}^{N_c}(1-\epsilon_{c,j}) \prod_{k\neq h}^{N_l}(1-\epsilon_{l,k}).
\end{aligned} \tag{6.8}$$

Moreover, $\epsilon_{\geq 2\text{ b jets}}$ depends on the specific event, like on the number of b, c and l jets

[f] Here, "light jets" indicate jets originating from gluons or quarks of the u, d and s flavor.

(N_b, N_c, N_l). Additionally, ϵ_b, ϵ_c and ϵ_l are assumed to depend on p_T and $|\eta|$, so that Equation 6.8 is individually evaluated for each event.

Finally, an event weight ($SF_{\geq 2 \text{ b jets}}$) is calculated from the ratio of selection efficiency with and without data to simulation SFs applied:

$$SF_{\geq 2 \text{ b jets}} = \frac{\epsilon_{\geq 2 \text{ b jets}}(SF_i \cdot \epsilon_{i,sim})}{\epsilon_{\geq 2 \text{ b jets}}(\epsilon_{i,sim})} \quad , \quad i \in b, c, l. \tag{6.9}$$

$SF_{\geq 2 \text{ b jets}}$ is separately calculated for each simulated event and after the b-jet selection step used as an event weight to the simulation.

As the misidentification rate (≈1% for the chosen CSVM working point) and the corresponding SFs are in the order of one, the two non-b jets are expected not to be misidentified as b-jets. Therefore, any correction related to the light jet identification step is negligible. The majority of events vetoed in this selection step, is expected to possess a light jet outside the kinematic acceptance[g].

Combination of Corrections

While the JECs are directly applied to the jets by adapting their four momenta, data to simulation scale factors are derived for all other data-driven corrections. Following a factorisation approach, these SFs are multiplied for each event:

$$SF_{event} = SF_{BR} \cdot SF_{PU} \cdot SF_l \cdot SF_{\geq 2 \text{ b jets}}. \tag{6.10}$$

The ultimately obtained result (SF_{event}) is applied as event weight to the simulation.

[g] The phase space defined by the p_T and η requirements of one or several objects is called kinematic acceptance.

6.8 Cutflow and Expected Event Composition

After applying all relevant corrections to the MC predictions, the event selection as defined in Chapter 6.2.1 to 6.5.2 is applied step by step and the selected number of events in data is compared to the prediction from simulation. The obtained "cut-flow" for the individual muon and electron semileptonic final states is shown in Table 6.2 for the lepton and jet selection, b-jet and light-jet selection. Furthermore, the number of events with a successful kinematic event reconstruction and the number of well reconstructed events fulfilling an additional 2% requirement on the minimal χ^2-probability of the event reconstruction (as explained in Chapter 7) are listed as well.

The prediction from simulation is obtained by normalising the MC prediction to the recorded luminosity using the theoretical cross sections discussed in Chapter 5.3.2. For each selection step, the corresponding correction factors for the MC prediction as detailed in Chapter 6.7 are taken into account.

Final State	μ+jets		e+jets		ℓ+jets combined		
Selection Step	N_{Data}	N_{MC}	N_{Data}	N_{MC}	N_{Data}	N_{MC}	$\frac{N_{Data}}{N_{MC}}$
$\equiv$ 1 lepton + $\geq$ 4 jets	290,557	273,396	316,567	283,416	607,124	556,811	1.090
+ $\geq$ 2 b jets	68,818	64,187	65,686	61,471	134,504	125,658	1.070
+ $\geq$ 2 light jets	64,899	61,022	62,060	60,300	126,959	121,323	1.046
+ kinematic reco	56,492	52,964	53,586	50,493	110,078	103,456	1.064
+ χ^2-prob.$\geq$2%	26,843	25,741	24,927	24,369	51,770	50,110	1.033

Table 6.2: Number of events obtained after lepton and jet selection, b-jet and light-jet selection, after a successful kinematic reconstruction of the full event (kinematic reco) and a quality requirement on the minimal χ^2-probability of the event reconstruction. The prediction from simulation (N_{MC}) includes pile-up correction, branching ratio and jet energy corrections as well as trigger and lepton selection efficiency corrections and, after b-jet selection, also b-tag efficiency scale factors and is compared to the result in data (N_{Data}).

After the b-jet and light-jet selection step, 64899 (62060) events are selected in data compared to 64187 (61471) expected events from the simulation in the muon (electron) final state. A successful kinematic reconstruction assuming a semileptonic $t\bar{t}$ event topology as needed to access the top-quark kinematics and discussed in Chapter 7 is possible for 85% (87%) of those events in simulation (data) for both final states. After requiring a minimum χ^2-probability of 2% for the obtained kinematic solution to enhance well reconstructed $t\bar{t}$ events, a total number of 51770 data events is used as input for the analysis. In this final selection step, the agreement between the number of selected events in data and simulation is at the level of about 3%. This small discrepancy is fully covered by systematic uncertainties. A more detailed discussion about the agreement of the total rate between data and simulation is given in Chapter 11.1.

The predicted event composition is derived from the expected events in simulation and is listed for different selection steps in Table 6.3. After the requirement of two

Final State	μ+jets				e+jets			
Physics Process	pre-tagged	tagged	kin-reco	prob	pre-tagged	tagged	kin-reco	prob
$t\bar{t}$ signal (e/μ prompt)	51	81	82	88	47	81	82	88
$t\bar{t}$ other	7	12	11	7	7	12	11	7
Single Top Q. (s-,t-&tW-channel)	3	4	4	2	3	4	4	2
$t\bar{t}$ +V	1	1	1	1	<1	1	1	1
W+Jets	31	2	2	1	27	2	2	1
Z+Jets	4	<1	<1	<1	5	1	<1	<1
QCD	1	<1	<1	<1	9	<1	<1	<1
Diboson	1	<1	<1	<1	<1	<1	<1	<1

Table 6.3: Expected relative event composition [%] as obtained from the simulation after the lepton and jet selection (pre-tagged), after b-jet requirement (tagged), after successful kinematic event reconstruction (kin-reco) and after a quality requirement on the minimal χ^2-probability of the event reconstruction (prob). The prediction from simulation includes pile-up, branching ratio and jet energy corrections as well as trigger and lepton selection efficiency corrections and, after b-jet selection, also corrections for the b-jet identification efficiencies.

identified b jets, the expected relative event composition is similar for both decay channels. For the final selection step, the expected composition of events used for the analysis comprises about 95% events originating from $t\bar{t}$ pair production. About 88% of the selected events are expected to originate from $t\bar{t}$ events with a prompt semileptonic decay into muon or electron which are treated as signal events ("$t\bar{t}$ signal"). About 7% of the selected events originate from other $t\bar{t}$ decay modes ("$t\bar{t}$ other").
Additionally, the composition per decay mode of the $t\bar{t}$ other events is shown in Figure 6.6. The majority of these events involve τ leptons in the decay chain. Therefore, additional neutrinos are involved, resulting in a more complex final state and a difference in the reconstructed distributions. Hence, these events are treated as background. The same argumentation holds for the production of $t\bar{t}$ pairs which are additionally accompanied by a vector boson.

Furthermore, the largest remaining background contribution from non-$t\bar{t}$ events stems from the production of single top quarks (≈2%) while the expected contribution of all other processes is expected to be around 3%. The remaining single-top-quark events are mainly from the tW-channel production mode (≈80%) while only few events are produced in the t-channel (≈17%) or the s-channel (≈2%).

Due to the combination of a large cross section and a similar final state, QCD multijet events and events with W bosons have been the most important background processes

for semileptonic $t\bar{t}$ analyses. Here, those backgrounds are efficiently suppressed by the applied selection criteria: QCD multijet events already by requiring an isolated, high-energetic lepton and events with W bosons by the requirement of at least two identified b jets.

As illustrated in Table 6.3, the kinematic event reconstruction has almost no influence on the expected event composition. In contrast to this, the requirement of a minimal χ^2-probability suppresses all events not originating from the $t\bar{t}$ signal.

The finally expected purity of 95% events from $t\bar{t}$ pair production is a demonstration of the possibilities offered by the large LHC dataset. The dataset recorded at $\sqrt{s} = 8$ TeV allows for a tight event selection resulting in an expected purity of $t\bar{t}$ events which is even larger than for the dileptonic final state [25], which has intrinsically much lower background contributions.

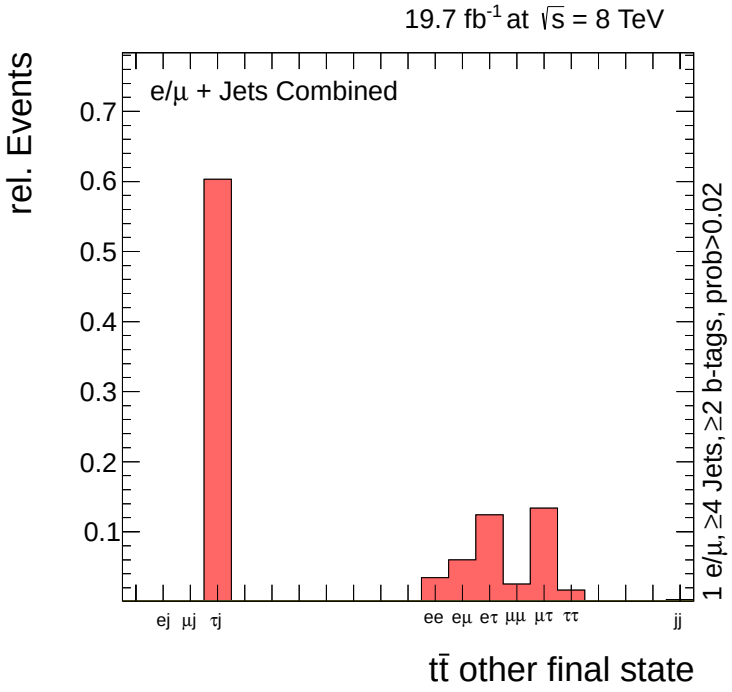

Figure 6.6: Expected event composition of the selected non-signal $t\bar{t}$ events from final states other than semileptonic with prompt muon or electron as obtained from the MadGraph+pythia MC prediction for the final selection step including a quality requirement on the minimal χ^2-probability of the event reconstruction.

6.9 Control Plots

To verify the understanding of the applied selection criteria, the shapes of the MadGraph+pythia MC prediction are compared to data for several basic kinematic quantities for the "combined" ℓ+jets final state (see Figure 6.7 to 6.12).

The combination of the semileptonic electron and muon final states is done by adding the selected event yields respectively. All further steps of the analysis use exactly this combined ℓ+jets final state as input. Before adding the event yields, all corrections as discussed in Chapter 6.7 are applied for the corresponding selection step.

Furthermore, the $t\bar{t}$ contribution is scaled to the in-situ measured total inclusive cross section individually for each selection step to ensure the same number of events in data and simulation.

Within the hatched uncertainty band for the simulation, the main uncertainties are taken into account:

- $t\bar{t}$ modelling uncertainties
  - ME-PS matching scale uncertainty
  - hard scattering Q^2 scale uncertainty
  - hadronisation uncertainty[h]

- experimental uncertainties
  - JES uncertainty
  - JER uncertainty
  - b-tag SF uncertainty (also shape effects)

A detailed discussion about the uncertainties and their precise definition can be found in Chapter 10. For all evaluated uncertainty sources, only the effect on the in-situ scaled shape of the $t\bar{t}$ component is considered for the control plots. Moreover, the provided data over MC ratios and the corresponding uncertainties are obtained from linearly rescaling the number of data events by the number of events from the MC prediction in each bin.

Furthermore, all plots shown are after the selection step of requiring at least two identified b jets. Quantities related to the lepton are shown in Figure 6.7, quantities related to jets in Figure 6.8 to 6.10, quantities related to b jets in Figure 6.11 and the missing transverse energy in Figure 6.12.

In general, good agreement between the shapes obtained from data and predicted by the simulation is found at the $\leq 10\%$ level and mostly within the uncertainties. However, all momentum-related distributions (e.g. p_T (jet), p_T (lepton) or E_T^{miss}) are observed to be slightly softer in data than in simulation. This observation will be discussed in detail in Chapter 12 and is assumed to be an effect of the modelling of the top quark transverse momentum which propagates through the decay to the final-state leptons, neutrinos and jets. Furthermore, a small difference for the pseudorapidity of

[h]The hadronisation uncertainty is defined as the difference between the predictions of the powheg+pythia and the mc@nlo+herwig predictions as detailed in Chapter 10.

the jets and leptons is observed. In data, the reconstructed jets and leptons appear to be slightly less central. This might be a hint for higher order contributions which are not modelled in the MadGraph+pythia prediction but e.g. in the powheg+pythia prediction which shows an improved consistency with the data. The visible dips in the barrel region around $\eta^{l} = \pm 0.25$ are due to the gaps between the wheels of the yoke, which result in a lower reconstruction efficiency for muons [166].

Moreover, an equally good agreement for the separate electron and muon final states is observed. As no difference between the final states is expected or observed and they both contribute approximately the same number of events to the combined control plots, no separate plots are shown.

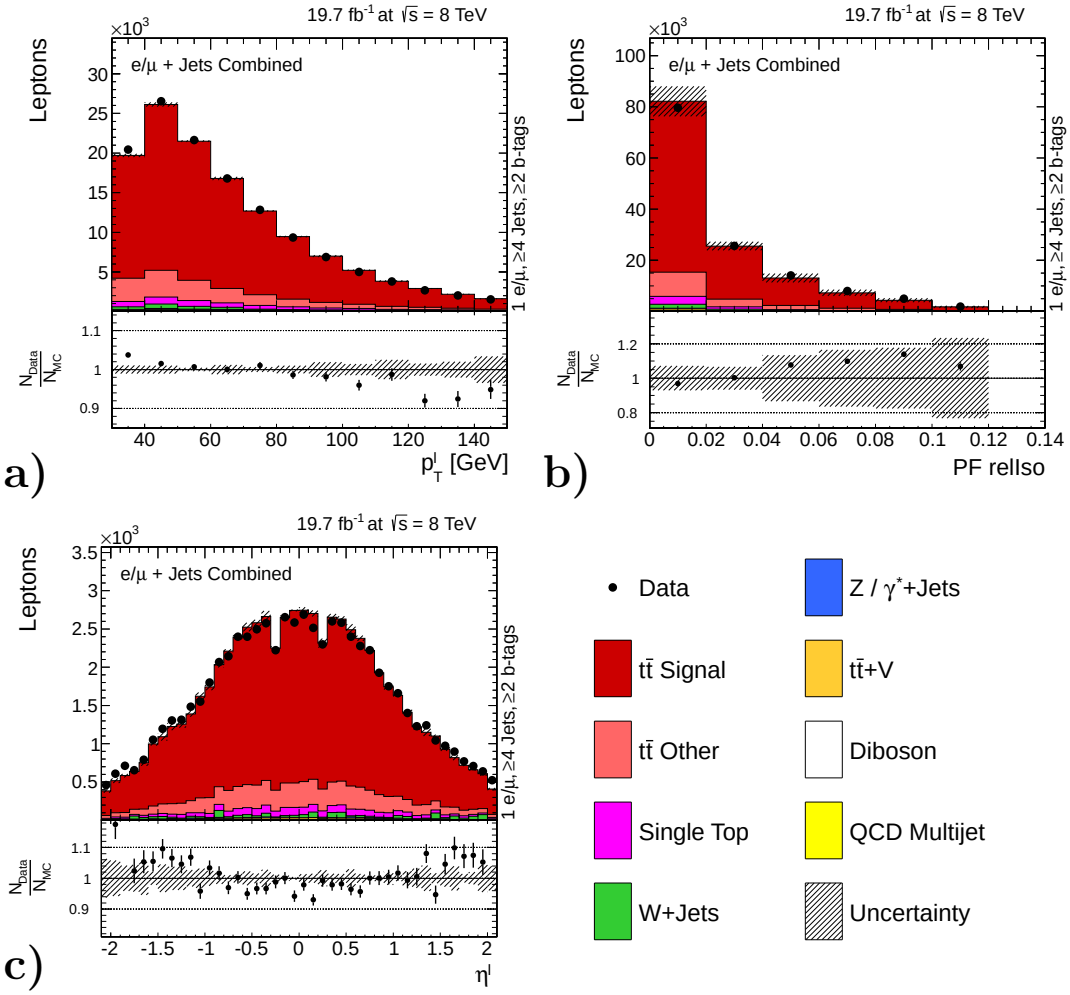

Figure 6.7: Transverse momentum (a), relative isolation (b) and pseudorapidity (c) for the selected lepton (e/μ) after lepton selection, jet selection and b-jet identification in the combined ℓ+jets channel. All correction factors are applied to the simulation.

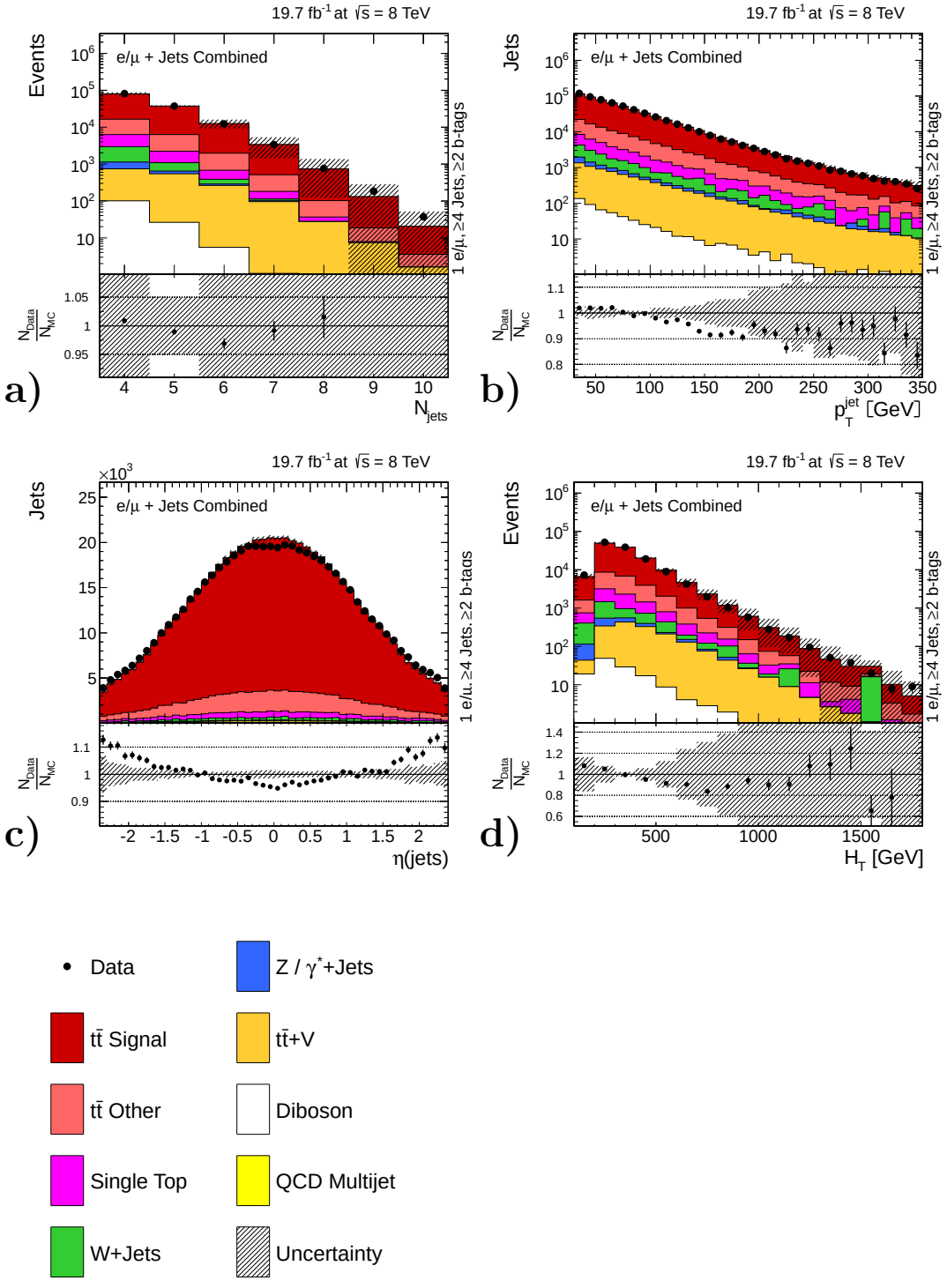

Figure 6.8: Multiplicity (a), transverse momentum (b), relative isolation (b), pseudorapidity (c), and the scalar sum of all jet p_T (d) for all selected jets after lepton selection, jet selection and b-jet identification in the combined ℓ+jets channel. All correction factors are applied to the simulation.

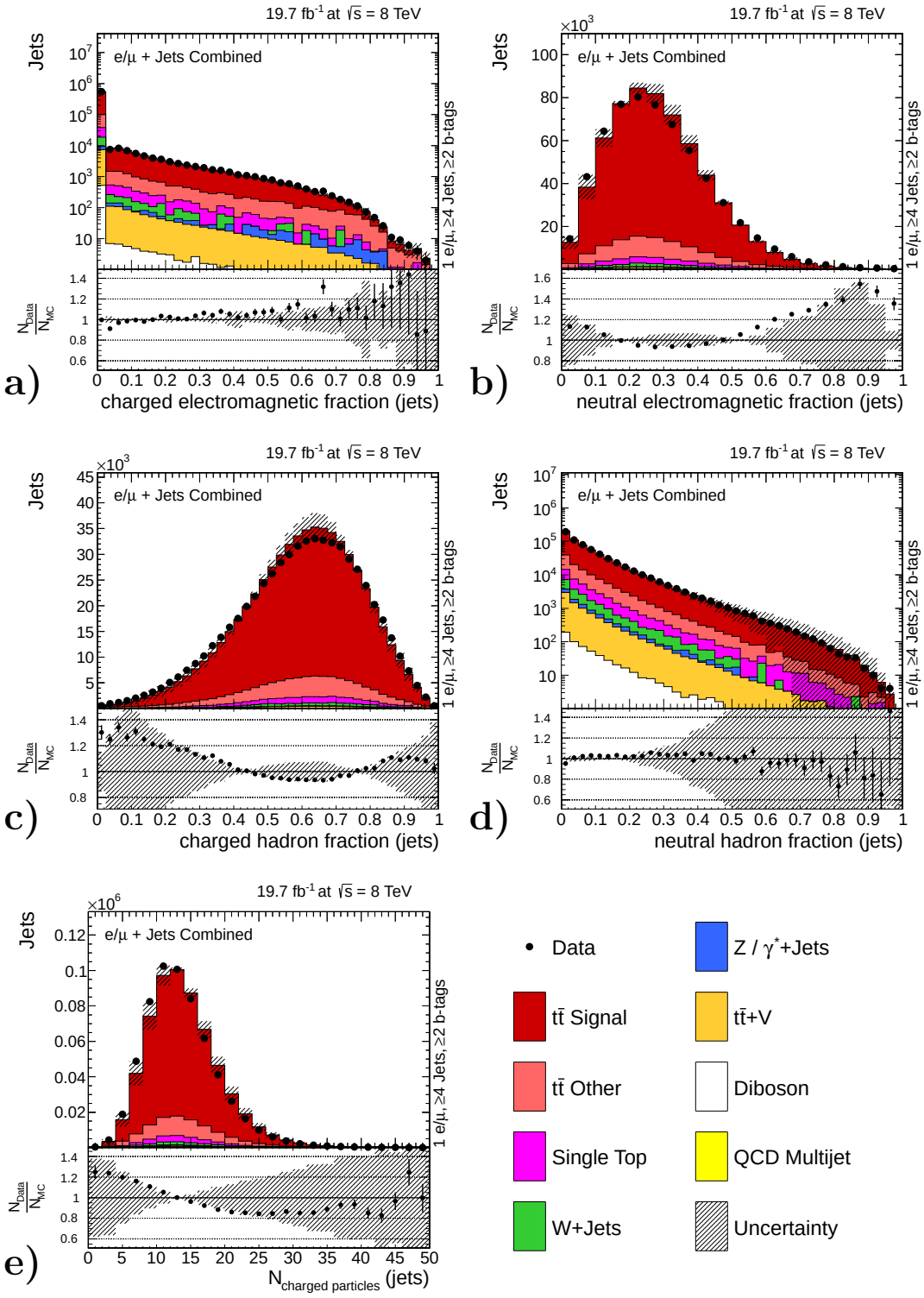

Figure 6.9: Variables for jet identification: charged electromagnetic fraction (a), neutral electromagnetic fraction (b), charged hadronic fraction (c), neutral hadronic fraction (d) and number of charged particles for all selected jets after lepton selection, jet selection and b-jet identification in the combined ℓ+jets channel. All correction factors are applied to the simulation.

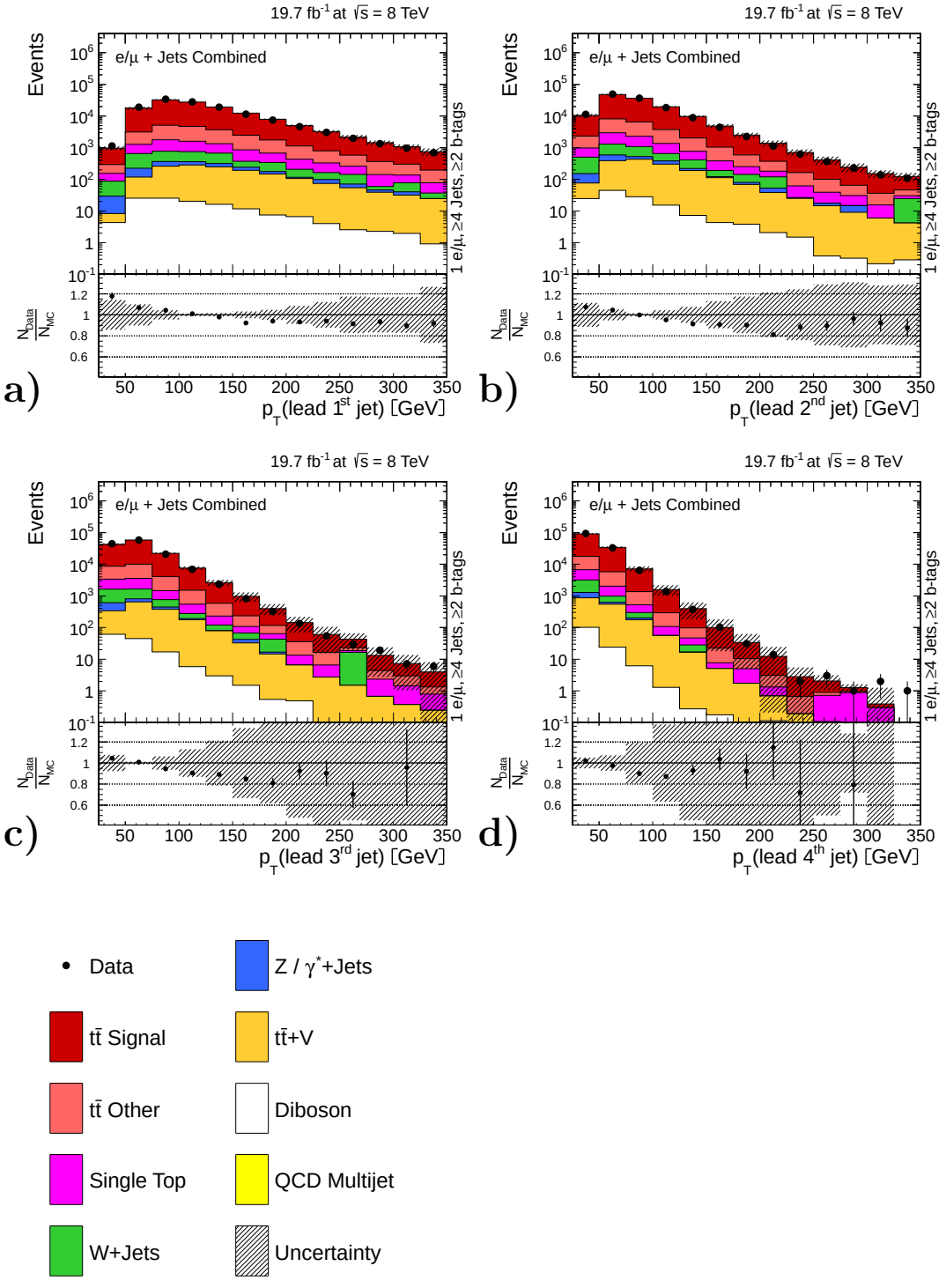

Figure 6.10: Transverse momentum (p_T) for the four jets with highest (a), second (b), third (c) and fourth (d) highest p_T for all selected jets after lepton selection, jet selection and b-jet identification in the combined ℓ+jets channel. All correction factors are applied to the simulation.

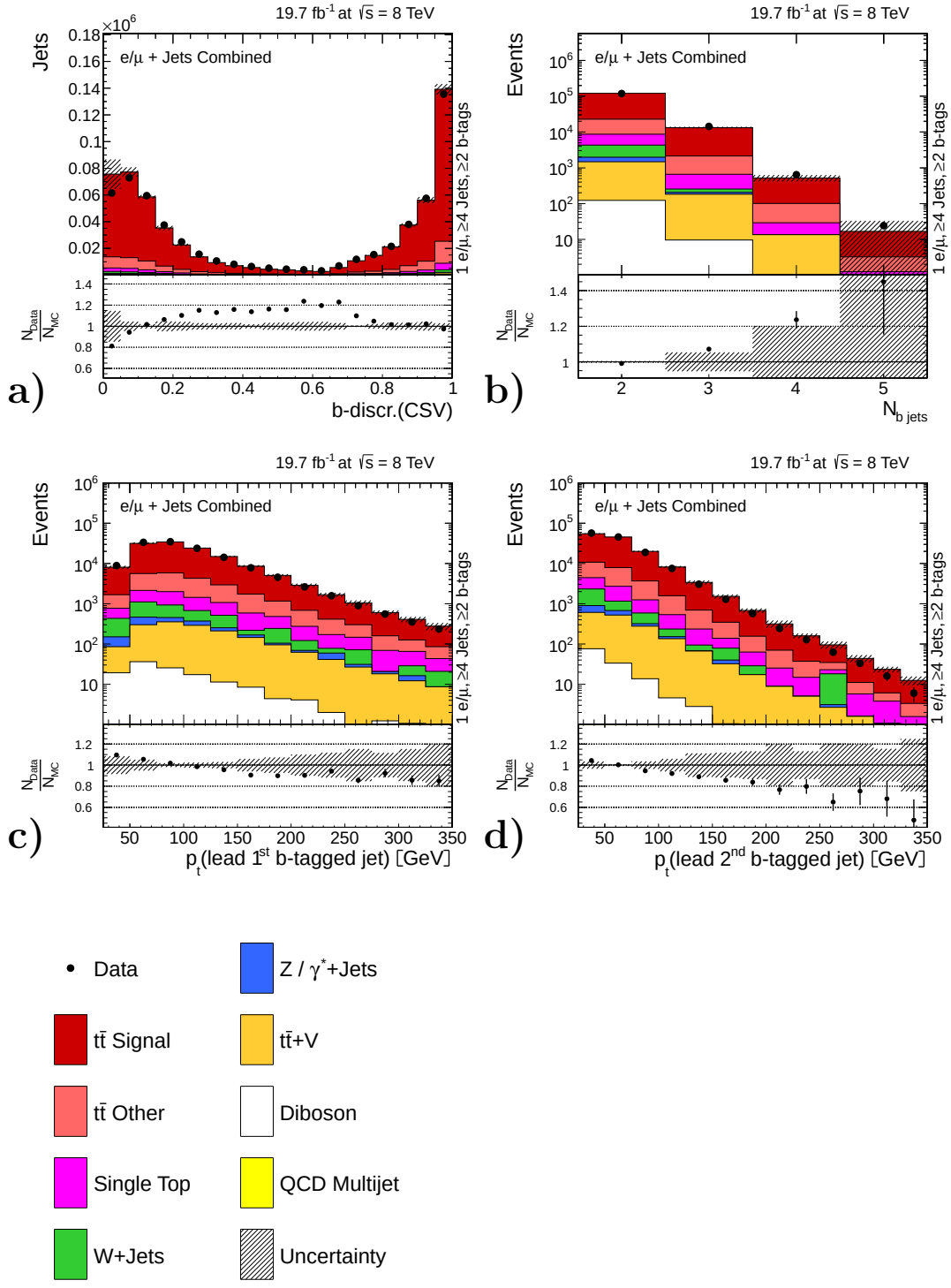

Figure 6.11: CSV b-tag discriminator for all selected jets (a) and resulting jet multiplicity (b) and transverse momentum of the identified b jet with highest (c) and second highest (d) p_T after lepton selection, jet selection and b-jet identification in the combined ℓ+jets channel. All correction factors are applied to the simulation.

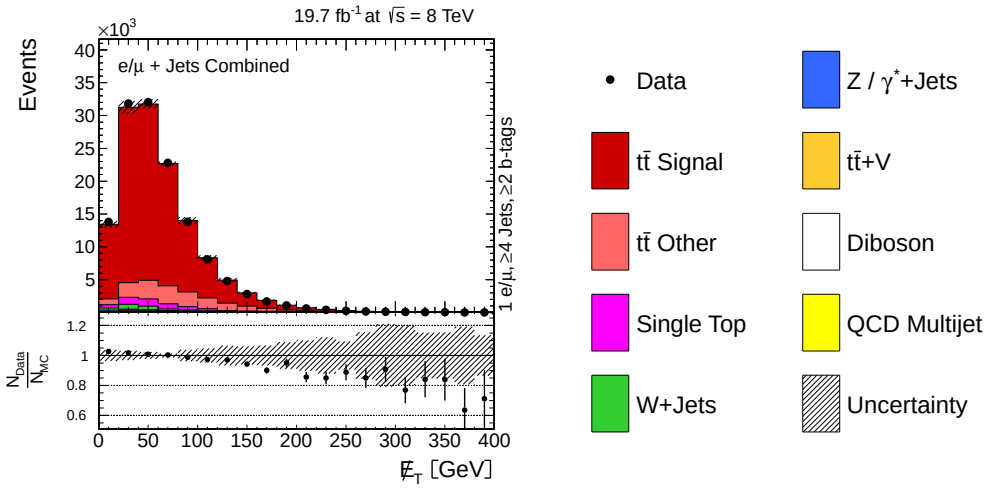

Figure 6.12: Missing transverse energy after lepton selection, jet selection and b-jet identification in the combined ℓ+jets channel. All correction factors are applied to the simulation.

Chapter 7

Kinematic Reconstruction of the $t\bar{t}$ Event Topology

To access the kinematics of the top quark itself, a full interpretation of the event topology is needed. This kinematic reconstruction involves assigning the reconstructed lepton and jets to the initial particles from the $t\bar{t}$ decay and finally obtain the top-quark four momentum as sum of the four momenta of its measured final state objects.

In the semileptonic decay channel (without additional jets), the final state comprises four jets, two of them initiated from b quarks of the top-quark decays and two from the hadronically decaying W boson. Additionally, in approximately 50% of all events further jets (e.g. from ISR/FSR) are passing the jet selection as described in Chapter 6.4.2 which increases the number of possible jet assignments. Another challenge is the a priori unmeasured neutrino of the leptonic W-boson decay. Assuming momentum conservation in the transverse plane, its x and y component is associated to the measured momentum imbalance ($\vec{E}_T^{miss}$). Hence, only the z component of the neutrino momentum remains unmeasured.

A commonly used method at Hadron colliders for the event reconstruction and interpretation of $t\bar{t}$ events is a kinematic fit. Its principles are explained in Chapter 7.1. The performance of the event reconstruction is essential for every high precision differential cross section measurement as function of the top-quark and top-quark pair quantities. Therefore, an improved strategy in comparison to the CMS analysis of differential top cross sections at $\sqrt{s} = 7$ TeV [23] is developed in Chapter 7.2 and its performance in data is tested in various ways in Chapter 7.3. Finally, the obtained distributions for all quantities are used as input for the calculation of differential cross sections are presented in Chapter 7.4.

7.1 Kinematic Fitting

The technical tool, used to perform the event interpretation is a kinematic fit as explained in [167] and implemented in the *KinFit* package [168].

All measured four momenta of the selected lepton and the selected five jets with the highest transverse momentum are given as input to the kinematic fit. Furthermore, the missing transverse momentum is associated with the transverse momentum of the neutrino while the z component of the neutrino momentum is set to 0 as initial value. To reduce the number of possible jet combinations only jets identified as b jets are considered as b quarks and only jets identified as non-b jets are considered as light quarks in the association of the reconstructed objects to the decay products of the semileptonic top-quark decay. The permutation of the two light jets is irrelevant because they are both assigned to the same W boson. Thus, in case of exactly four selected jets the only ambiguity left is the permutation of the two b jets.

Within the fitting procedure, the kinematic properties of the input objects (lepton, neutrino and jets) are changed to fulfill certain topological constraints. At least one constraint is necessary to solve the kinematic equations due to the unmeasured neutrino p_z. Further constraints help to find the correct jet assignment and to improve the resolution of the measured object kinematics within the fitting procedure. All constraints applied involve the reconstructed top-quark and W-boson masses:

- W-boson mass constraints
  - $m_W^{reco} = m(jj)$
    $= m(l\nu) \equiv 80.4\,\text{GeV}$
- top-quark mass constraints
  - either: $m_{top}^{reco} = m(l\nu b) \equiv m(jjb) = \text{free}$
  - or: $m_{top}^{reco} = m(l\nu b)$
    $= m(jjb) \equiv 172.5\,\text{GeV}$

The invariant mass of the light jets as well as the invariant mass of the lepton and the neutrino are required to form exactly the W-boson mass of 80.4 GeV. Furthermore, mass constraints for the reconstructed top-quark masses are used. This is done either by requiring equal reconstructed masses for the top and antitop quark– leaving the actual value of the mass free or by forcing the reconstructed top-quark mass to a fixed value. Additionally, it was studied that including a width in the mass constraints has an negligible effect on the result of the fit. All constraints are illustrated in Figure 7.1.

The technical implementation of the kinematic fit minimises a likelihood function ($\mathscr{L}$):

$$\mathscr{L} = \underbrace{(\Delta\vec{y})^t \cdot \mathbf{COV}_y^{-1} \cdot (\Delta\vec{y})}_{\mathbf{A}} + \underbrace{2 \cdot \sum_{k=1}^{\#\text{constraints}} \lambda_k \cdot f_k}_{\mathbf{B}} \tag{7.1}$$

$$\Delta y_i = y_i^{fit} - y_i^{input}. \tag{7.2}$$

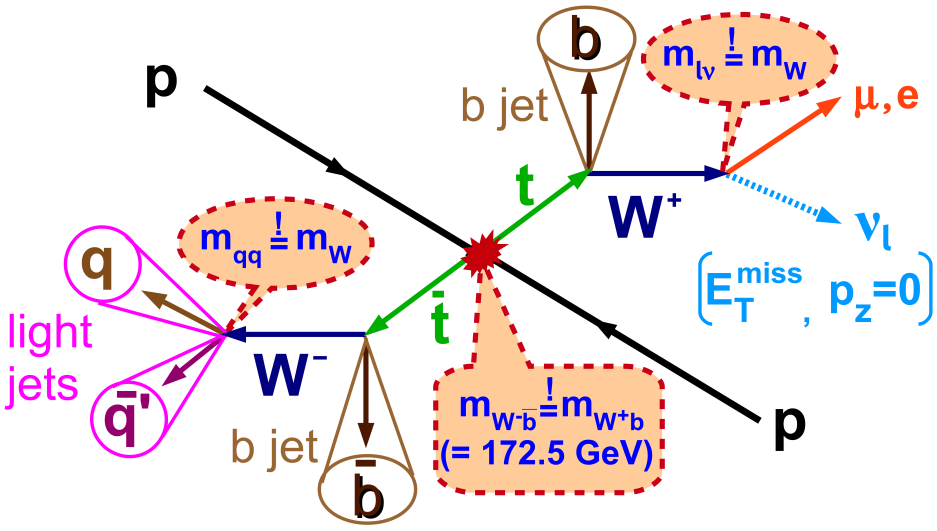

Figure 7.1: Illustration of the topological constraints for the reconstructed W-boson and top-quark masses used for the event interpretation of the semileptonic $t\bar{t}$ final states with one prompt muon or electron.

which is constructed from the initial (superscript "input") and fitted (superscript "fit") kinematic properties (i=p_T, η, ϕ) of the input objects (y=lepton, neutrino, two b jets and two non-b jets), their covariance matrix ($\mathbf{COV}_y$) and the applied constraints (f_k) which are implemented using Lagrangian Multipliers (λ_k). Moreover, the constraints are constructed such that f_k=0 means they are perfectly fulfilled, e.g. f_1=m_W^{reco}-80.4 GeV. Therefore, a small numerical value for expression **B** in Equation 7.1 means that the applied constraints are well fulfilled.

As covariance matrix, a diagonal matrix with the object resolutions (res_y) on the diagonals is used:

$$\left(\mathbf{COV}_y\right)_{ij} = \delta_{ij} res_{y_i}. \tag{7.3}$$

Therefore, expression **A** represents the amount of change to the kinematics of the input objects relative to their resolutions. Consequently, input objects with poor resolutions (in this case the neutrino and the jets) are on average changed by larger values than input objects which are measured more accurately (like the lepton).
The resolutions of all reconstructed final-state objects are obtained from a $t\bar{t}$ simulation relative to their parton level truth values. Details are given in Chapter 7.3.1. A small numerical value for **A** indicates solutions with minimal changes to the input objects.

Consequently, minimising the Likelihood function $\mathscr{L}$ in Equation 7.1 leads to a solution where the constraints are fulfilled best and the kinematics of the input objects are only minimally changed according to their resolutions.
In order to find the minimum of the likelihood $\mathscr{L}$, a linearisation approach and an iterative solving is performed (see [167] for details) so that finally the constraints are nearly exactly fulfilled. Finally, the kinematic fitting procedure is considered as

successful if a solution is found within 500 steps of the iterative solving routine while the convergence criteria are a maximum absolute change of 0.00005 for **A** between the last two iterations and a maximum absolute value of 0.0001 for **B**.

Depending on the number of selected jets, several jet assignments are possible and therefore multiple solutions per event can be obtained. To choose the most likely correct one, an estimator for the goodness of the fit is used. For this purpose, the compatibility of the obtained solution with the assumed $t\bar{t}$ topology is quantified. Here, the assumed $t\bar{t}$ topology is reflected by the constraints. Therefore, the difference between the initial and fitted object kinematics relative to the resolution is evaluated. Top like events are expected to require only a minimal change if all the relevant final state objects are reconstructed and the correct jet assignment is chosen as they naturally form the W-boson and top-quark mass. To quantify this compatibility, expression **A** in Equation 7.1 is interpreted as a χ^2 expression and simplifies to:

$$\mathbf{A} \overset{7.1}{=} (\Delta\vec{y})^t \cdot \mathbf{COV}_{\mathrm{y}}^{-1} \cdot (\Delta\vec{y}) \overset{7.3}{=} \sum_{\mathrm{i}} \left(\frac{y_{\mathrm{i}}^{\mathrm{fit}} - y_{\mathrm{i}}^{\mathrm{input}}}{\mathrm{res}_{\mathrm{y}}} \right)^2 = \chi^2. \tag{7.4}$$

Moreover, the χ^2 is converted into a χ^2-probability (prob) with values between 0 (bad solution) and 1 (good solution):

$$\mathrm{prob}(\chi^2, \mathrm{n}) = \frac{\int_{\chi^2}^{\mathrm{inf}} \mathrm{t}^{\frac{\mathrm{n}}{2}-1} \cdot \exp(-\frac{\mathrm{t}}{2})\mathrm{dt}}{2^{\frac{\mathrm{n}}{2}} \cdot \int_{0}^{\mathrm{inf}} \mathrm{t}^{\frac{\mathrm{n}}{2}-1} \cdot \exp(-\mathrm{t})\mathrm{dt}}. \tag{7.5}$$

Here, n is the number of degrees of freedom that depends on the number of constraints and the number of unmeasured parameters. For the constraints with equal (fixed) top-quark mass, n is 2 (3).

Finally, the fitted object kinematics of the lowest χ^2 (=highest χ^2-probability) jet assignment is used as input for all differential cross section measurements. Thereby, the correction of the input parameter kinematics within the fitting procedure helps limiting systematic uncertainties, e.g. the momentum of the light jets is automatically calibrated by the W-boson mass constraint. This is an advantage of the semileptonic over the dileptonic final state which is kinematically underconstrained due to the two neutrinos.

7.2 Optimisation of the Kinematic Reconstruction

The event reconstruction is one of the key points of this analysis. If the jets are wrongly associated to the top quarks, an arbitrary wrong solution is obtained. Therefore, the choice of different constraints and configurations for the kinematic fit is studied and an improved strategy with respect to previous analyses is developed.

7.2.1 Expected Jet Assignment as Quality Criteria

In order to quantify the performance of the chosen setup of the kinematic fit, the assignment of the reconstructed jets from the solution with the minimum χ^2 is compared to the correct assignment.

Determining the correct jet assignment comprises the difficulty of associating the reconstructed and selected b and light jets to the original b and light quarks of the underlying top-quark decay. This association is done using a jet-parton matching algorithm where every jet is associated to every parton within ΔR(jet,parton)<0.3. Only events with an unambiguous result for all partons are considered as successfully matched. This assures that the matching result is correct for most of the events but the algorithm does not find a solution in up to 50% of all cases.

These *unmatched* events originate for example from cases where one of the jets from the $t\bar{t}$ decay is not selected (e.g. because it is outside the kinematic acceptance or not passing the jet identification criteria), cases where two quarks from the $t\bar{t}$ decay are merged in one single jet because of a boosted event topology or cases where one parton from the $t\bar{t}$ decay results in two or more reconstructed jets due to FSR or misreconstruction. For all these scenarios, a correct reconstruction of the top quarks is not possible using the kinematic fitting procedure.
On the other hand, the matching can fail also for correct solutions if the result is ambiguous because jets are close-by or if ΔR(parton,jet)>0.3.

Possible permutations when comparing the jet assignment from the kinematic fit to the result of the jet-parton matching algorithm are:

- all jets correctly assigned (ok)
- b jets interchanged (bb)
- for more than four selected jets:
  wrong four out of the five jets with the highest transverse momentum chosen (wrongj)
- at least one of the jets from the $t\bar{t}$ decay is not among the five jets with the highest transverse momentum (jmis)
- cyclic permutation of two or three jets (blepq, bhadq, bbqlep, bbqhad)
- all jets are wrongly assigned (bbqq)

The ordering of the listed particles in the abbreviations for the cyclic jet permutations specifies which jets are interchanged:

- blepq : leptonic b jet and one light jet interchanged

- bhadq : hadronic b jet and one light jet interchanged
- bbqlep: light jet assigned as leptonic b jet, leptonic b jet as hadronic b jet and hadronic b jet as light jet
- bbqhad: light jet assigned as hadronic b jet, hadronic b jet as leptonic b jet and leptonic b jet as light jet

It was found that using the five instead of only the four jets with the highest transverse momentum in the fitting procedure results in about 20% more correct jet assignments and reduces the cases where one jet from the $t\bar{t}$ decay is missing by a factor of almost four. Using six or more jets does not improve the result further but enhances cases where the wrong jets are chosen.
Furthermore, it was found that using the information of the b-jet identification for the jet-quark assignment reduces all permutations involving a b jet and a non-b jet by up to 70%. For the typical setup of the kinematic fit with five jets as input and the b jet information as it was already used in the CMS analysis of differential top cross sections at $\sqrt{s} = 7$ TeV [23, 25], the major contributions to events without correct jet assignment are events without matching and events with the two b jets interchanged. In the following, an improved strategy is developed to reduce both cases.

7.2.2 The Double Kinematic Fit

Assigning the b jets correctly turns out to be the main challenge of the jet assignment if all relevant jets from the $t\bar{t}$ decay are reconstructed and selected. One solution is to constrain the reconstructed value of the top-quark mass, e.g. using $m_{top}^{reco} \equiv 172.5\,\text{GeV}$ as constraint reduces solutions with interchanged b jets by approximately 50%.

On the other hand, the choice of the mass value can bias those differential distributions which are very sensitive to the top-quark mass, leading therefore to an additional uncertainty.

In contrast to the CMS analysis of differential top cross sections at $\sqrt{s} = 7$ TeV [23,25], the value of the top-quark mass is used as a constraint in an improved event interpretation. To use the knowledge of the true top-quark mass without depending on its exact value, an approach with two sequential fitting steps in the following called "Double Kinematic Fit" is performed.

The first kinematic fit is applied in order to choose the best jet assignment. Permutations are tested using the selected five jets with the highest transverse momentum. Besides the W-boson mass constraint, a fixed top-quark mass constraint of $m_{top}^{reco} = 172.5\,\text{GeV}$ is used to suppress permutations with interchanged b jets. Finally, only the jet assignment with the minimum χ^2 is chosen as input for the second fit.

All kinematic properties for the differential cross section calculation are determined in the second kinematic fit. As input, the jet assignment from the first fit is used while all four momenta of the input objects (including the ones of the jets) are obtained from the reconstruction step before the first fit. Besides the W-boson mass constraint, only equal reconstructed top- and antitop-quark masses are required, leaving the actual

value of m_{top}^{reco} free. Therefore, the obtained kinematic distributions will depend only weakly on the choice of the value for the m_{top} constraint of the first fit.

Finally, the obtained χ^2 and kinematics of the second kinematic fit are used in the further analysis. This is consistently done for the top quarks as well as for the input objects. The Double Kinematic Fit is needed for the top-quark reconstruction but also improves the resolution for the final state objects (especially the jets) due to the constraints.

7.2.3 Minimum χ^2-Probability Requirement

Signal events (e/μ+jets $t\bar{t}$ final states) with correct jet assignment naturally fulfill the applied W-boson and top-quark mass constraints. Hence, the obtained average value for the χ^2-probability of the kinematic reconstruction (as defined in Equation 7.5) is expected to be higher than for background events or signal events with a wrong jet assignment. Figure 7.2 shows the expected shape of the χ^2-probability distribution for $t\bar{t}$ signal events with correct jet assignment, all other $t\bar{t}$ signal events and $t\bar{t}$ non-signal events. A clear separation of correctly assigned signal events can be observed.

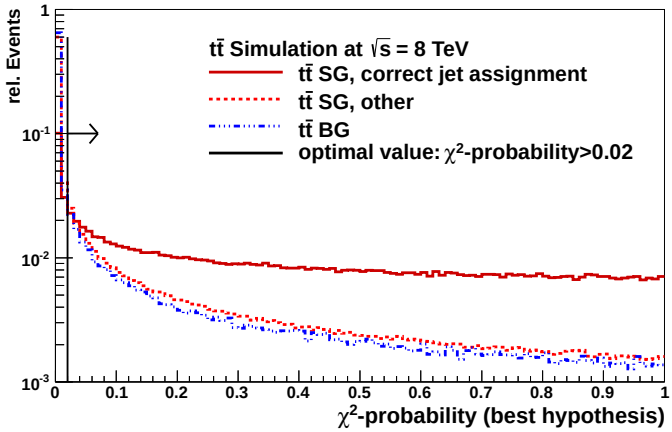

Figure 7.2: Relative shape of the χ^2-probability distribution for the minimum χ^2 solution as obtained from the double kinematic fit event reconstruction procedure for simulated $t\bar{t}$ events. Three different type of $t\bar{t}$ events are studied: signal events (ℓ+jets final state) with correctly assigned jets ("$t\bar{t}$ SG, correct jet assignment"), other signal events ("$t\bar{t}$ SG, other") and non-signal $t\bar{t}$ events ("$t\bar{t}$ BG"). The black line indicates the probability value of 2% used as selection criteria to enhance correctly reconstructed events.

Consequently, a minimum χ^2-probability requirement is introduced to enhance the fraction of good reconstructed events. Finally, this results in limited migration effects for the cross section quantities (see Chapter 8.4).
On the other hand, choosing larger values of the minimal χ^2-probability will reduce the statistics in data too much. This results in larger statistical uncertainties, especially in the tails of steeply falling distributions like $m^{t\bar{t}}$.

In order to find the best minimal χ^2-probability requirement as a compromise of a high efficiency and a high fraction of correctly reconstructed signal events, the significance for correct reconstructed signal events relative to all reconstructed $t\bar{t}$ events is optimised:

$$\begin{aligned} \text{significance} &= \frac{\text{sig}}{\sqrt{\text{sig}+\text{bkg}}} \\ \text{sig} &= \mathrm{N}^{t\bar{t}\ \mathrm{SG}}_{\text{correct jet assignment}} \\ \text{bkg} &= \mathrm{N}^{t\bar{t}\ \mathrm{SG}}_{\text{wrong jet assignment}} + \mathrm{N}^{t\bar{t}\ \mathrm{SG}}_{\text{no jet}-\text{parton matching}} + \mathrm{N}^{t\bar{t}\ \mathrm{BG}}. \end{aligned}$$

As the fraction of non-$t\bar{t}$ background events is already fairly small after the b-jet identification (see Chapter 6.8), only $t\bar{t}$ events are involved to optimise the χ^2-probability requirement. The expected significance and selection efficiency as determined from the $t\bar{t}$ simulation for different choices of the minimal χ^2-probability requirement is shown in Figure 7.3 for the Double Kinematic Fit.

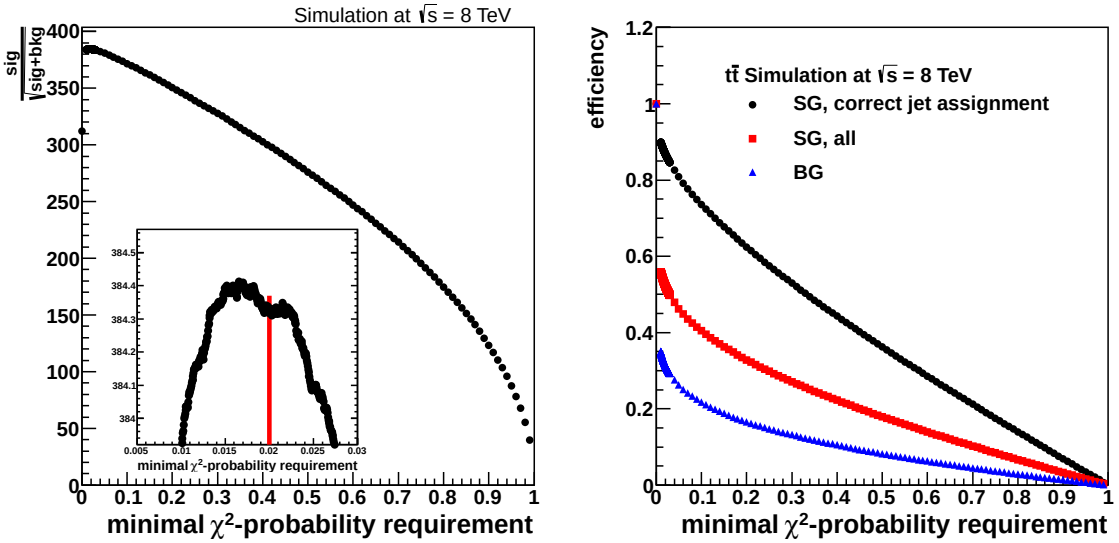

Figure 7.3: Left: Significance for correctly reconstructed $t\bar{t}$ signal events relative to all reconstructed $t\bar{t}$ events. Right: Efficiency for $t\bar{t}$ signal events with correct jet assignment (black), other $t\bar{t}$ signal events (red) and $t\bar{t}$ non-signal events (blue) as expected from simulation for different requirements of the minimal χ^2-probability. The finally chosen value of 2% is marked with a red line in the left figure.

As optimal value, a minimal χ^2-probability requirement of 2% is found which corresponds to a maximum χ^2 requirement of 7.82. The value of 2% turns out to be independent of the choice of the top-quark mass constraint for all studied scenarios (equal mass, fixed mass and Double Kinematic Fit), while all numbers quoted in the following refer to the Double Kinematic Fit.

Only few correctly reconstructed events are vetoed while events with wrong jet assignment as well as non-signal events are efficiently suppressed. The efficiency of the χ^2-probability>2% requirement is about 50% for all $t\bar{t}$ events, 87% for signal events with correct jet assignment, 36% for all other $t\bar{t}$ signal events and 31% for all non-signal

$t\bar{t}$ events. The separate efficiencies for all jet assignments of the $t\bar{t}$ signal events are shown in Figure 7.4.
Events without jet-parton matching are effectively suppressed and the expected fraction of signal events with correctly assigned jets is raised to over 50% while the expected ratio of $t\bar{t}$ signal events over $t\bar{t}$ non-signal events is increased from 7 to 12.

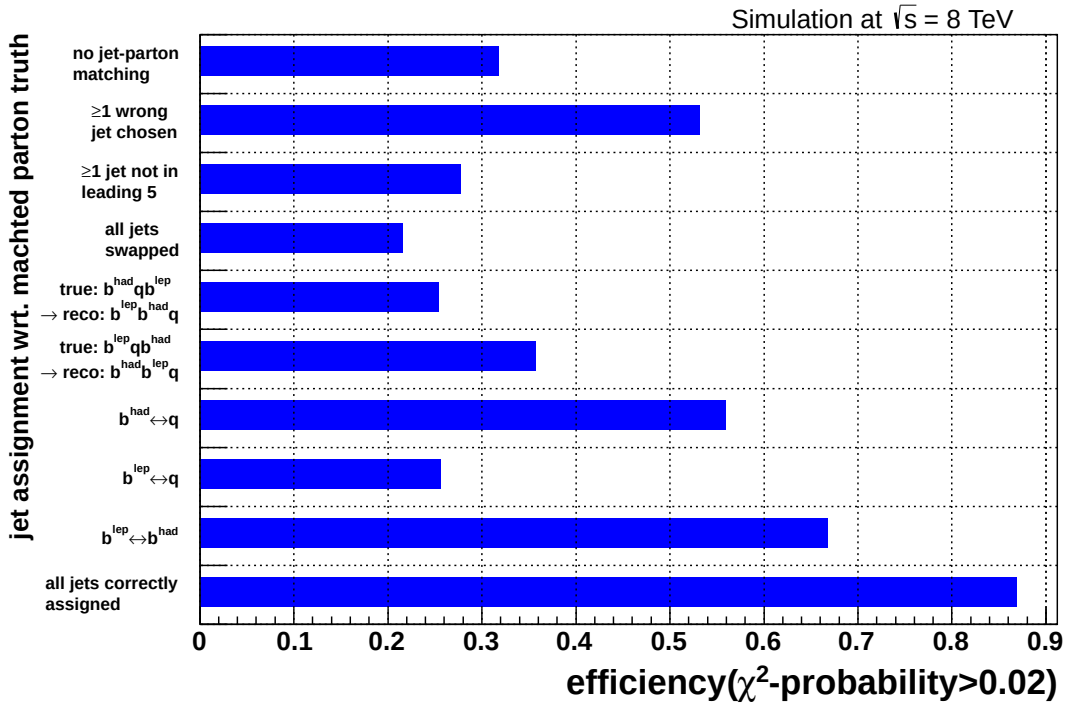

Figure 7.4: Expected efficiency for all jet assignments of $t\bar{t}\rightarrow e/\mu$+jets events for the minimal χ^2-probability requirement of 2%.

Furthermore, the relative composition of $t\bar{t}$ background events before and after the minimal χ^2-probability requirement is shown in Figure 7.5. Dileptonic $t\bar{t}$ events are suppressed more efficiently and therefore, the most important decay channel among the non-signal $t\bar{t}$ events is the τ+jets channel.

In summary, the requirement of a minimal χ^2-probability is an efficient method to convert the larger statistics of the $\sqrt{s} = 8$ TeV dataset into an enhancement of well-reconstructed events. This results in limited migration effects, which is needed in order to improve a systematically limited differential measurement.

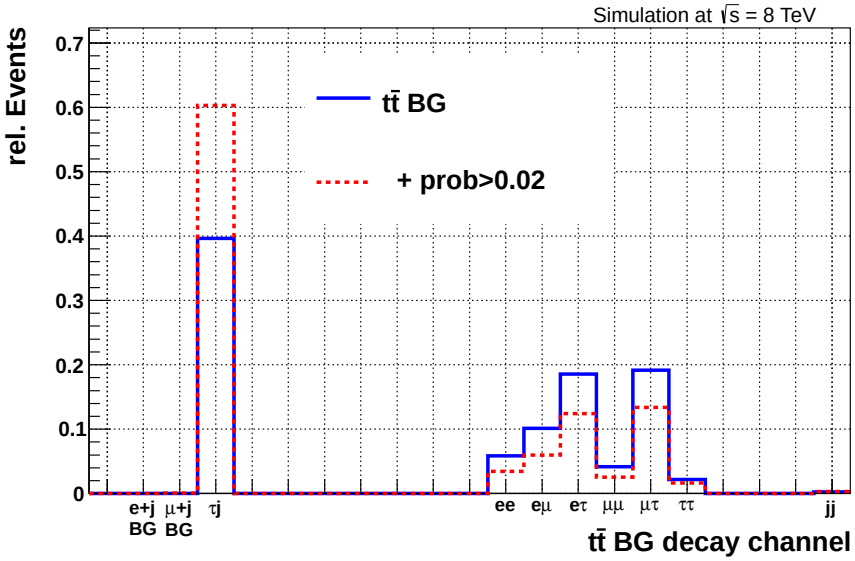

Figure 7.5: Expected relative composition of the non-signal $t\bar{t}$ events before (blue) and after (red) a minimal χ^2-probability requirement of 2% for the combined e/μ+jets final state. The "e+j BG" and "μ+j BG" contributions are in the respective other final state.

7.2.4 Comparison of Different Top-Quark Mass Constraints

The expected jet assignment for $t\bar{t}$ signal events has been studied for three different top-quark mass constraints with and without a minimal χ^2-probability requirement as introduced in Chapter 7.2.3:

- equal top-quark mass constraint with free mass value
- fixed top-quark mass constraint of 172.5 GeV
- Double Kinematic Fit approach

The expected number of events for the data luminosity for all three scenarios is separately shown for all possible jet assignments in Figure 7.6. All abbreviations for the different jet permutations are introduced in Chapter 7.2.1. The best performance in terms of a high fraction of solutions with correct jet assignment is obtained with the fixed top-quark mass approach. Especially the most important permutation where the two b jets are interchanged, is significantly reduced in comparison to the equal top-quark mass constraint. However, the performance of the double kinematic fit approach is only slightly worse compared to the fixed top-quark mass constraint and can reduce events with interchanged b jets by roughly the same amount.

Additionally, the minimum χ^2-probability requirement reduces the number of events without solution for the jet-parton matching efficiently. Those events are expected to originate mostly from events without correct reconstruction (see Chapter 7.2.1).

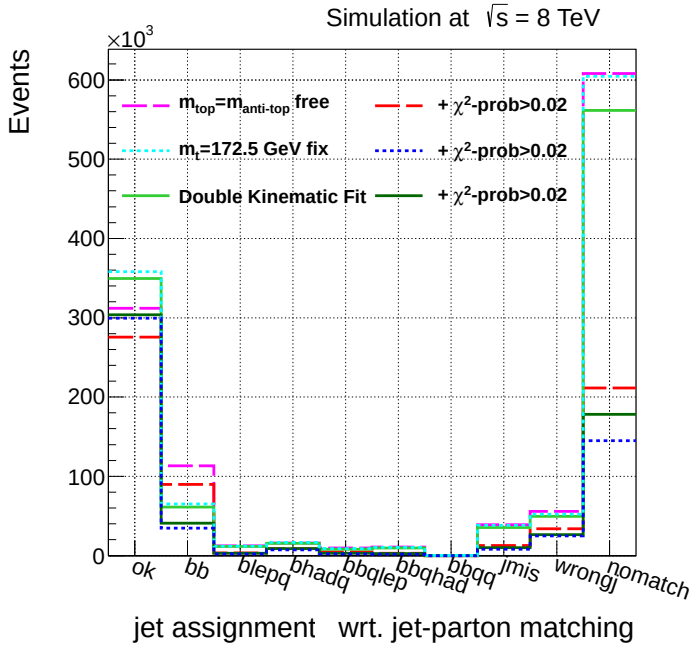

Figure 7.6: All jet assignments of reconstructed $t\bar{t}$ signal events for different choices of the top-quark mass constraint in the kinematic fit with and without the requirement of a minimal χ^2-probability. The number of events from the simulation is normalised according the data luminosity using the total inclusive cross section $\sigma_{t\bar{t}}^{\mathrm{NNLO+NNLL}}$.

Thus, the Double Kinematic Fit with the requirement of the minimal χ^2-probability, which is finally chosen as event reconstruction for this thesis, shows an improved performance in comparison to the setup with an equal top-quark mass constraint which was used for my previous CMS analyses of differential top cross sections [25, 26]. The absolute number of events with interchanged b jets is reduced by a factor of approximately three and the absolute number of events without jet-parton matching by a factor of approximately 3.5 while the absolute number of correctly reconstructed events is approximately the same. As illustrated in Figure 7.7, the relative number of events with correctly assigned jets is raised from about 25% to over 50%.

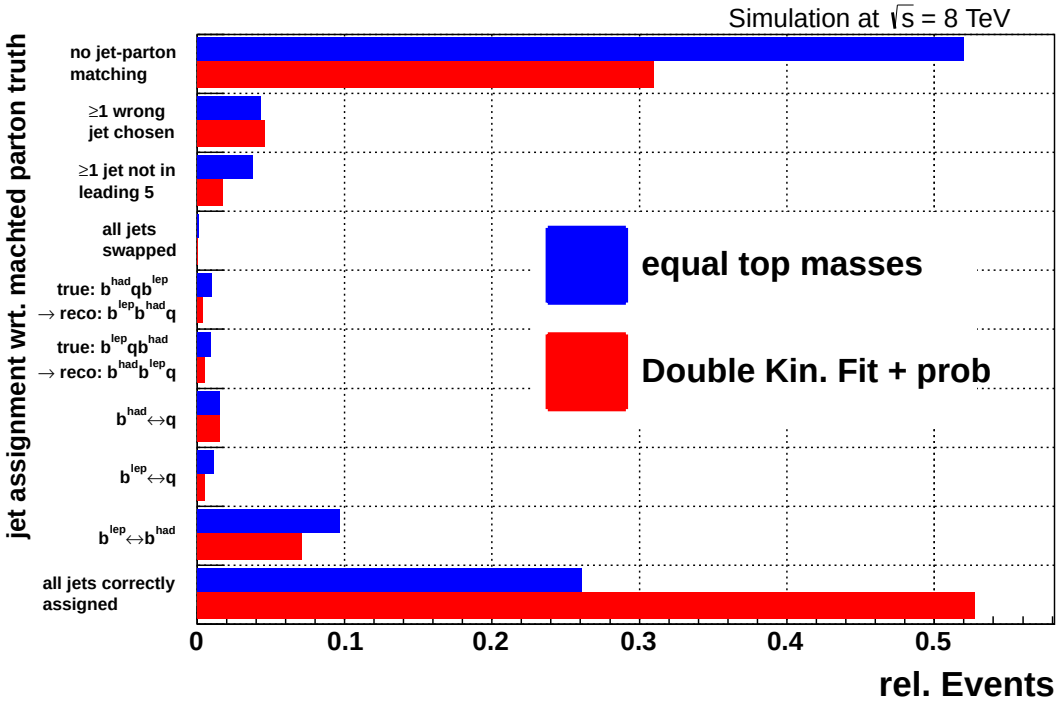

Figure 7.7: Expected relative composition for the performance of the jet assignment with the equal top-quark mass constraint as used for previous CMS analyses of differential top cross sections (blue) and with the Double Kinematic Fit and a minimal χ^2-probability requirement of 2% as used for this thesis (red).

7.2.5 Top-Quark Mass-Constraint Value in the Double Kinematic Fit

The to date experimental precision of top-quark mass measurements is in the order of one GeV or even below [169–175] and the obtained result is between 172 and 175 GeV. This is very close to the value of 172.5 GeV which is used for m_{top}^{gen} in the simulated $t\bar{t}$ MC samples. Therefore, the mass constraint for the first step of the Double Kinematic Fit is also chosen to be $m_{top}^{reco} \equiv 172.5\,\text{GeV}$.
Potential differences between the true top-quark mass in data and the assumed top-quark mass for the simulation are addressed by a systematic uncertainty assigned to the choice of the top-quark mass for the simulation in the procedure of the differential cross section calculation.

The key idea of the Double Kinematic Fit is based on the assumption that the result as obtained from the event interpretation and reconstruction does not depend on the exact choice of the top-quark mass constraint because the value of 172.5 GeV is only used for the jet assignment in the first step. To test this assumption, the expected prediction for all reconstructed differential distributions used as input for the cross section measurements as derived from the default $t\bar{t}$ simulation with $m_{top}^{gen} = 172.5\,\text{GeV}$ is compared for different values of the top-quark mass constraint.
It was found that for a kinematic reconstruction with simple fixed top-quark mass constraint, $m^{b\bar{b}}$ and $p_T^{b\bar{b}}$ are the most sensitive distributions. They reveal changes of up to 5% for a variation of one GeV ($\equiv 0.6\%$) in the value of the m_{top}^{reco} constraint. In contrast, when using the Double Kinematic Fit a negligible dependency on the value of typically less than 2% is observed. The results for the Double Kinematic Fit are shown in Figure 7.8 to 7.10. The dashed error bars denote the statistical uncertainty of the simulation for the default value of the top-quark mass constraint. Furthermore, the top-quark and $t\bar{t}$ system quantities used for differential cross section measurements are found to be in general more robust regarding the choice of the top-quark mass constraint value.

Moreover, it was checked that the reconstructed shape of the ρ_S distribution (defined in Equation 3.16) does not depend strongly on the choice of the exact value for the top-quark mass constraint. This is of essential importance for a potential extraction of m_{top} as studied in Chapter 13. A small dependency (as observed for the last bin) is not a problem for the concept of the measurement. The sensitivity of the reconstructed distribution on the true top-quark mass might be reduced but is still present. Consequently, the extracted result will be less precise but the measurement itself is still valid. Additionally, the observed effect is small in comparison to the total uncertainty. This is discussed in more detail in Chapter 13.

Overall, a negligible dependence for the reconstructed distributions compared to the uncertainties of the measurement is found for reasonable variations in the order of a few GeV. Only for mass constraint values far beyond the known precision on m_{top}, a sizeable effect on the reconstructed distributions is observed because then the correct jet assignment is effectively suppressed by the wrong constraint value.

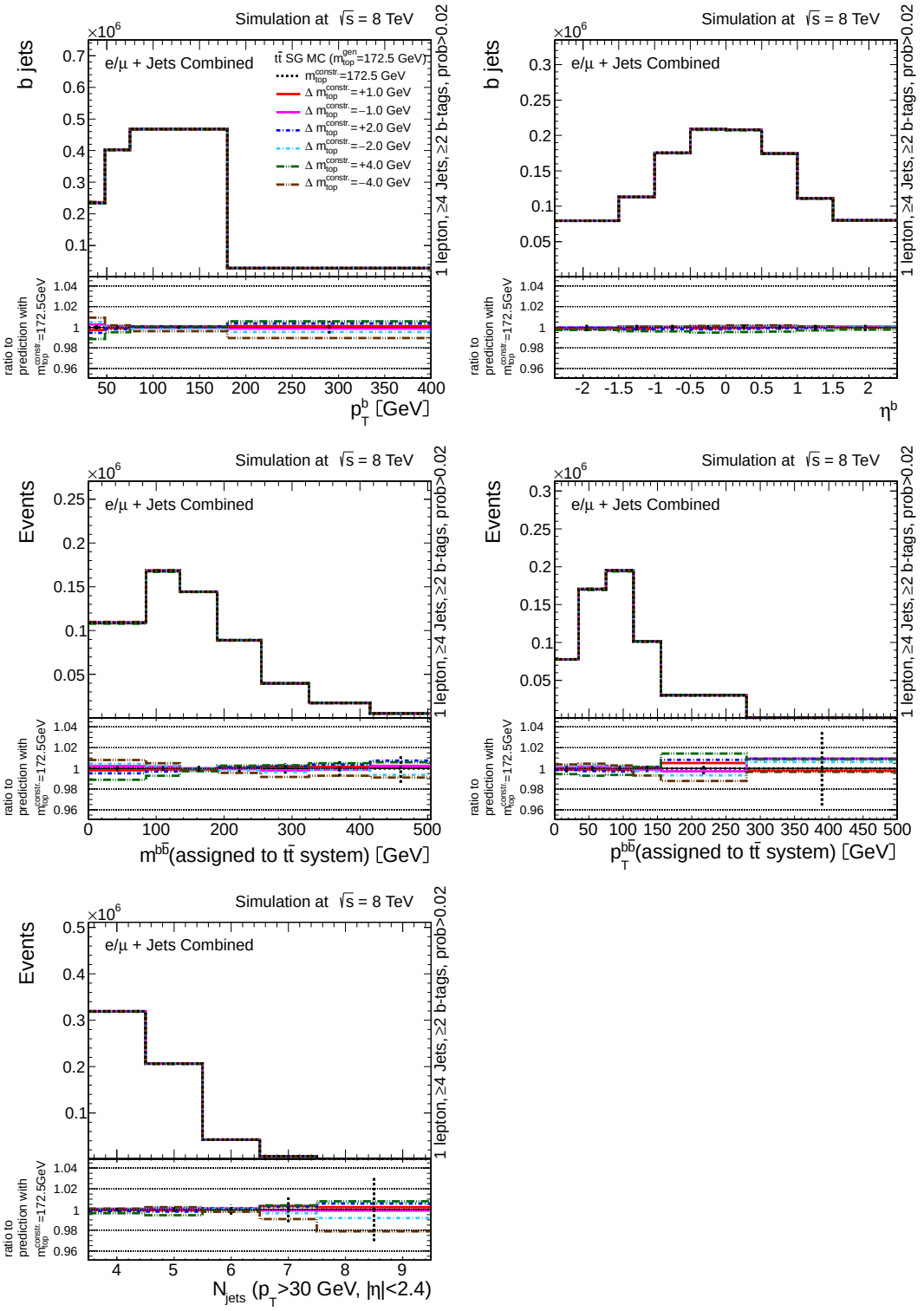

Figure 7.8: Reconstructed distributions for the momentum (p_T^b) and pseudorapidity of the b jets (top), the invariant mass ($m^{b\bar{b}}$) and transverse momentum ($p_T^{b\bar{b}}$) of the b-jet pair associated to the $t\bar{t}$ system (middle) and the jet multiplicity (N_{jets}) as predicted for different choices of the m_{top} constraint (m_{top}^{constr}) by ±1, ±2 and ±4 GeV around the default of 172.5 GeV in the Double Kinematic Fit for $t\bar{t}$ signal events. In addition, the ratio to the default value of m_{top}^{constr} is shown. The dashed error bars denote the statistical uncertainty of the simulation for the default value of m_{top}^{constr}.

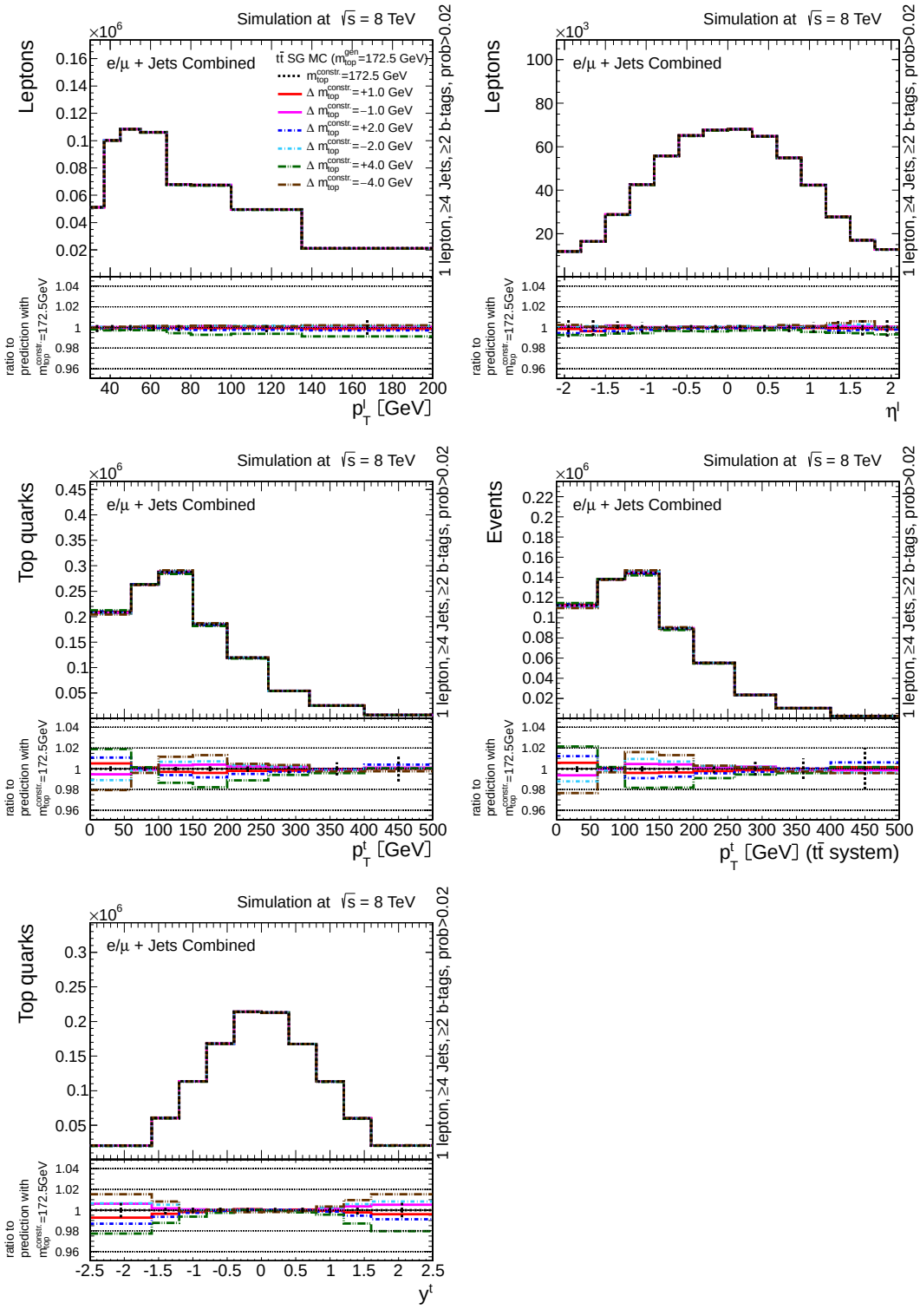

Figure 7.9: Reconstructed distributions for the transverse momentum (p_T^l) and pseudo-rapidity (η^l) of the lepton (top row) and the transverse momentum (p_T^t) and rapidity (y^t) of the top quark as well as the transverse momentum of the top quark in the $t\bar{t}$ rest frame (p_T^t ($t\bar{t}$ com)) (middle and bottom row) as predicted for different choices of the m_{top} constraint (m_{top}^{constr}) by ±1, ±2 and ±4 GeV around the default of 172.5 GeV in the Double Kinematic Fit for $t\bar{t}$ signal events. In addition, the ratio to the default value of m_{top}^{constr} is shown. The dashed error bars denote the statistical uncertainty of the simulation for the default value of m_{top}^{constr}.

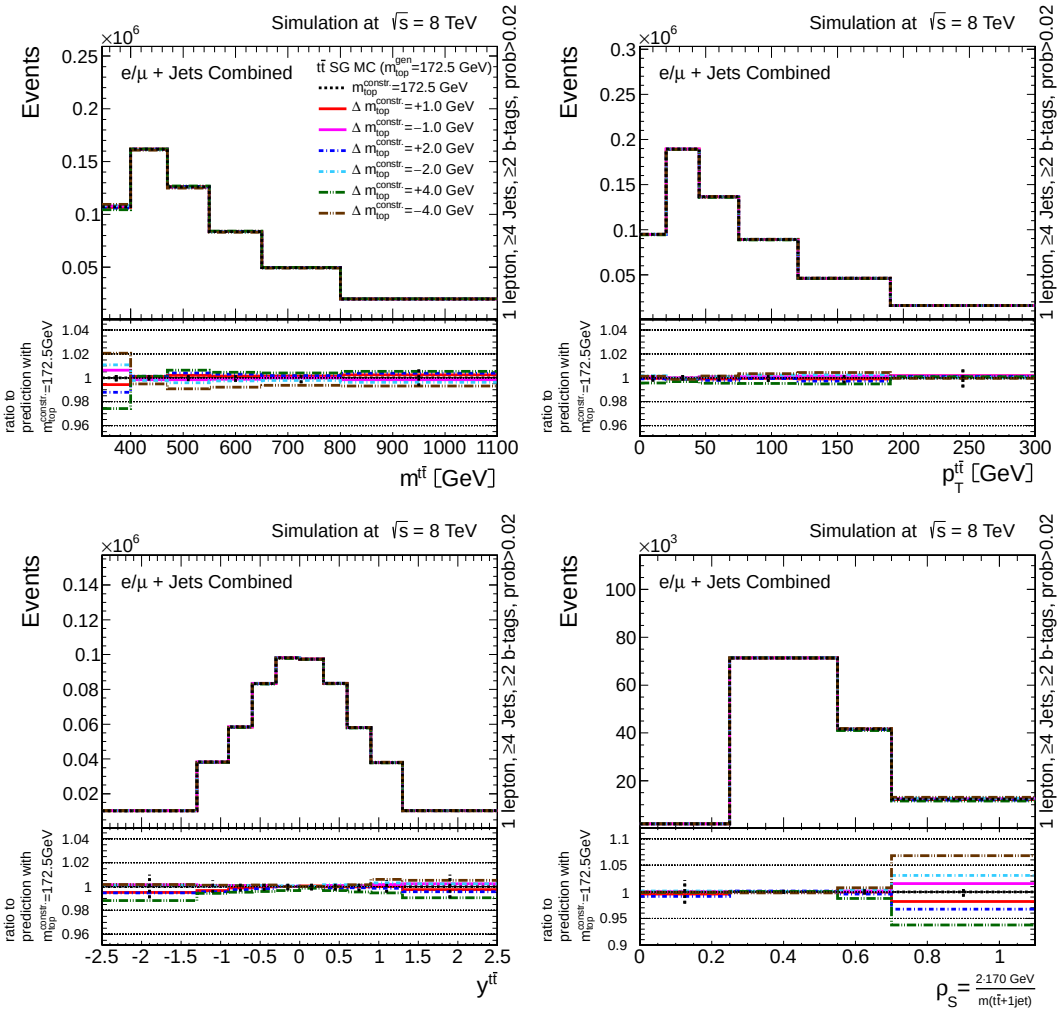

Figure 7.10: Reconstructed distributions for the invariant mass ($m^{t\bar{t}}$), transverse momentum ($p_T^{t\bar{t}}$) and rapidity ($y^{t\bar{t}}$) of the $t\bar{t}$ system as well as ρ_S as predicted for different choices of the m_{top} constraint (m_{top}^{constr}) by ±1, ±2 and ±4 GeV around the default of 172.5 GeV in the Double Kinematic Fit for $t\bar{t}$ signal events. In addition, the ratio to the default value of m_{top}^{constr} is shown. The dashed error bars denote the statistical uncertainty of the simulation for the default value of m_{top}^{constr}.

7.3 Applying the Double Kinematic Fit in Data

Ensuring a detailed understanding of the performance of the Double Kinematic Fit event interpretation and reconstruction procedure in data, various control distributions are compared between data and simulation.

If not explicitly specified, the expected contribution from events of non-$t\bar{t}$ processes for all control plots is normalised to the expected luminosity using theory cross sections (see Chapter 5.3.2). In contrast, the $t\bar{t}$ contribution is scaled to the in-situ measured total inclusive cross section to ensure the same number of events in data and simulation. Within the uncertainty bands for the simulation, the effect of the main uncertainties on the shape of the $t\bar{t}$ distribution is taken into account:

- $t\bar{t}$ modelling uncertainties
  - ME-PS matching scale uncertainty
  - hard scattering Q^2 scale uncertainty
  - hadronisation uncertainty

- experimental uncertainties
  - JES uncertainty
  - JER uncertainty
  - b-tag SF uncertainty (also shape effects)

The details for these uncertainties are explained in Chapter 10. All Figures are shown for the combined e/μ+jets final state.

7.3.1 Resolution and Kinematic Shift of the Input Quantities

The resolutions of the input objects determine their individual contributions to the χ^2-expression (Equation 7.4) and are therefore an important ingredient for the obtained solution.

The resolution of each quantity represents the average difference between the measured reconstruction level object (lepton, jets, missing transverse energy) and the initial parton (lepton, quarks, neutrino). Different object resolutions are used for b jets and light jets while an effectively infinite large resolution is chosen for the unmeasured neutrino component.

The quantities p_T, η and ϕ are chosen to parametrise the resolution. All resolutions are derived double differentially as function of p_T in intervals of η at reconstruction level. A $t\bar{t}$ simulation is used for the determination of the resolutions. The results can be found in [176]. As they were already used in the previous analysis at $\sqrt{s} = 7$ TeV, they are discussed in more detail in [23]. No update is needed as no relevant changes of the resolutions at $\sqrt{s} = 8$ TeV are expected due to the very similar object definitions. Moreover, the exact choice for the resolution parameters does not matter to first approximation. As data and simulation are treated identical, differences in the fitting procedure should cancel in case of a good simulation.

More important than the exact choice of the values used for the covariance matrix of

the event reconstruction is the intrinsic modelling of the input object resolutions in the simulation. A difference in the jet energy resolution between the selected jets in data and simulation would e.g. lead to different results and consequently wrong differential cross sections as the assumed reconstruction efficiency and the minimum χ^2-probability selection efficiency are different. To verify that this is not the case, the absolute change (x^{fit}-x^{reco}) in p_T, η and ϕ due to the fit is compared between data and simulation for all input objects. These distributions are directly sensitive to the modelling of the resolutions and their size reflects also the relative value for the resolution, e.g. lepton quantities are expected to be changed only minimally as their resolution is best among all input objects.

Typical magnitudes for the kinematic shift of all input objects are extracted from the average absolute change and summarised in Table 7.1. As expected, the momentum of the jets (due to their relatively large resolution) and the quantities of the neutrino are shifted most.

quantity x	Average value for $\langle \Delta x^{fit} \rangle = \langle |x^{fit}\text{-}x^{reco}| \rangle$
p_T^{lepton}	10^{-2} GeV
p_T^{ν}	15 GeV
$p_T^{b\ jet}$	10 GeV
$p_T^{light\ jet}$	15 GeV
η^{lepton}	10^{-7}
η^{ν}	1
$\eta^{b\ jet}$	10^{-3}
$\eta^{light\ jet}$	10^{-2}
ϕ^{lepton}	10^{-8}
ϕ^{ν}	0.1
$\phi^{b\ jet}$	10^{-3}
$\phi^{light\ jet}$	10^{-2}

Table 7.1: Average absolute difference between initial x^{reco} and final x^{fit} value of all kinematic properties x of the input objects for the Double Kinematic Fit as obtained from data.

The distribution for the absolute shift Δx^{fit} for the kinematic properties (p_T, η, ϕ) of all measured final-state objects (lepton, neutrino, two light jets, two b jets) for for data and simulation is shown in Figure 7.11 and 7.14. Good agreement between data and simulation within uncertainties in the order of 5-10% is observed for all quantities. Non-Gaussian tails in the resolutions and wrong jet assignments in the Double Kinematic Fit can result in non-symmetric distributions for several quantities.

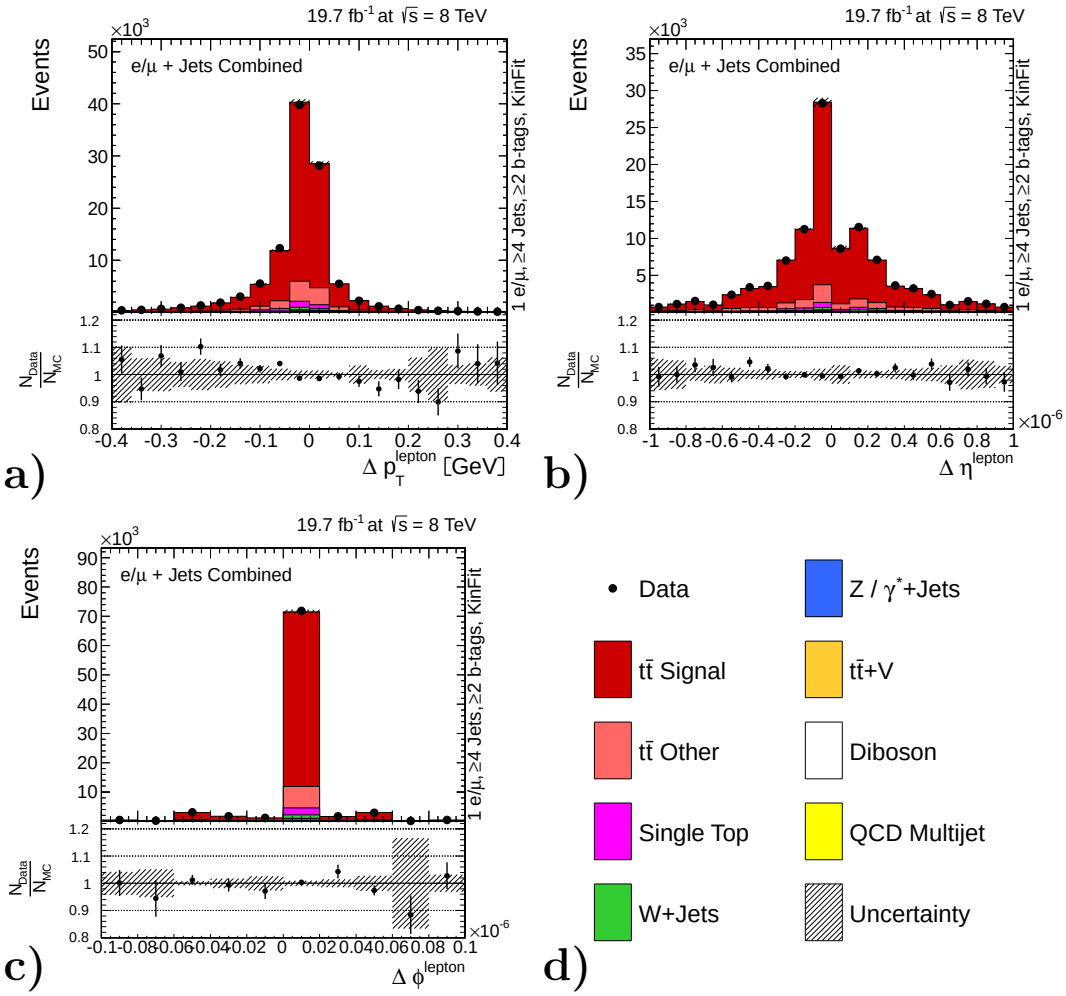

Figure 7.11: Absolute shift $\Delta x = x^{fit} - x^{reco}$ of the kinematic quantities $x = p_T$, η, ϕ for the lepton within the Double Kinematic Fit event reconstruction procedure. The hatched band corresponds to the main experimental and $t\bar{t}$ model uncertainties of the simulation.

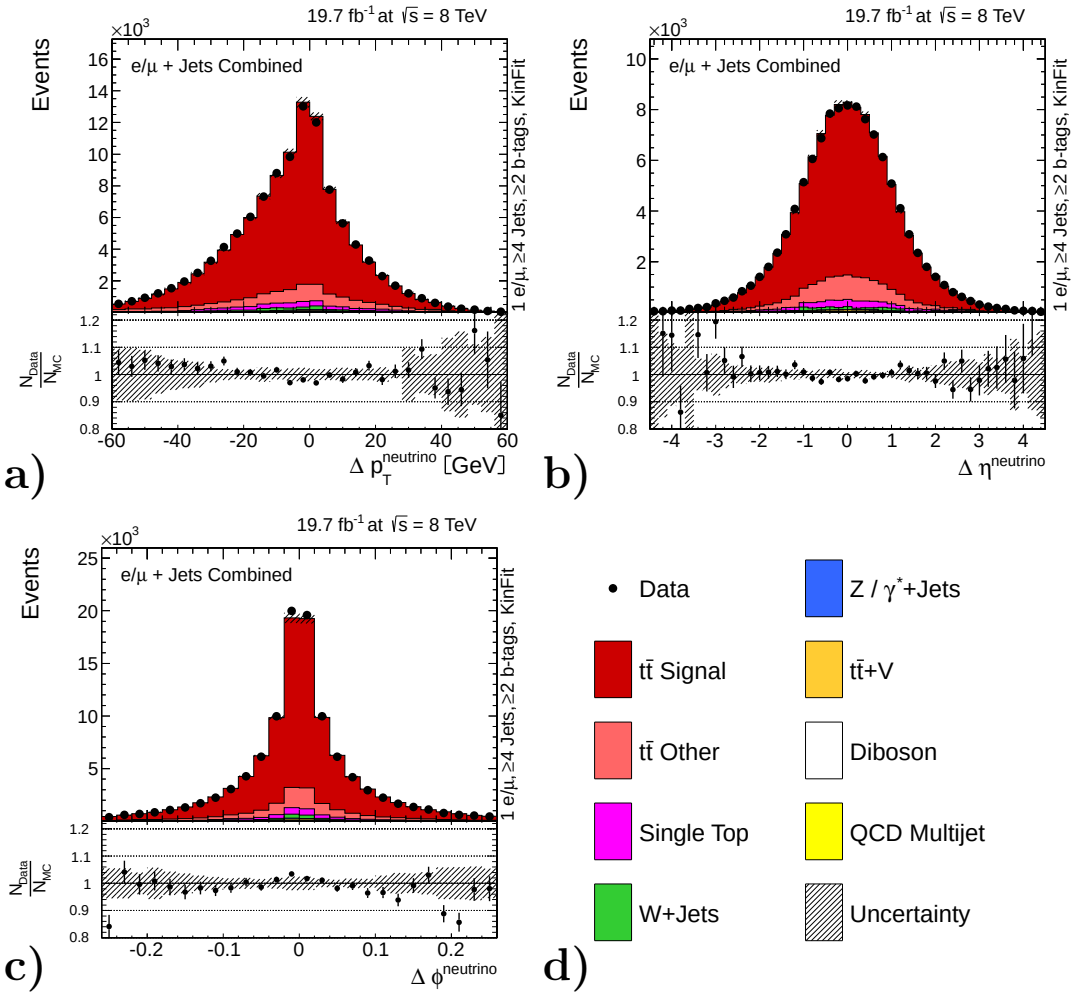

Figure 7.12: Absolute shift $\Delta x = x^{fit} - x^{reco}$ of the kinematic quantities $x = p_T$, η, ϕ for the neutrino within the Double Kinematic Fit event reconstruction procedure. The hatched band corresponds to the main experimental and $t\bar{t}$ model uncertainties of the simulation.

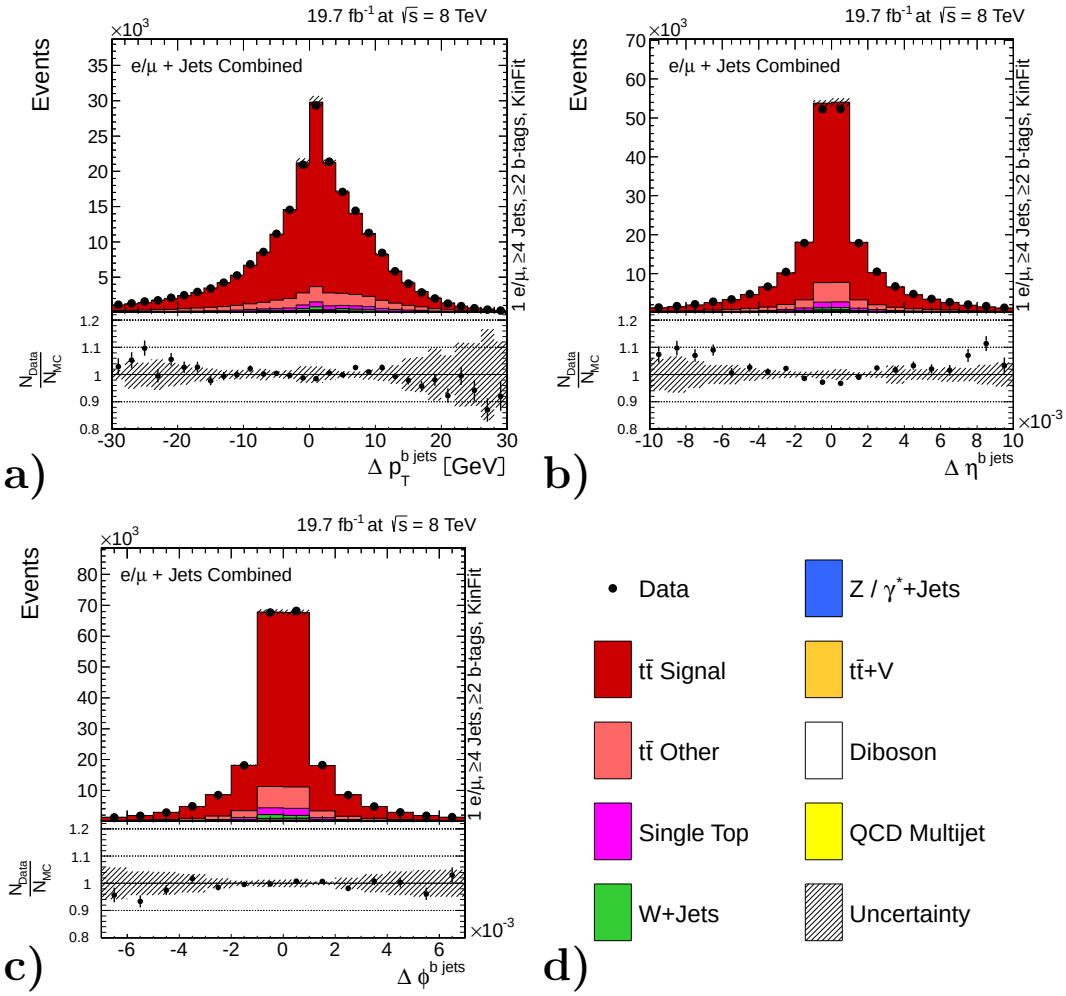

Figure 7.13: Absolute shift $\Delta x = x^{fit} - x^{reco}$ of the kinematic quantities $x = p_T$, η, ϕ for the b jets within the Double Kinematic Fit event reconstruction procedure. The hatched band corresponds to the main experimental and $t\bar{t}$ model uncertainties of the simulation.

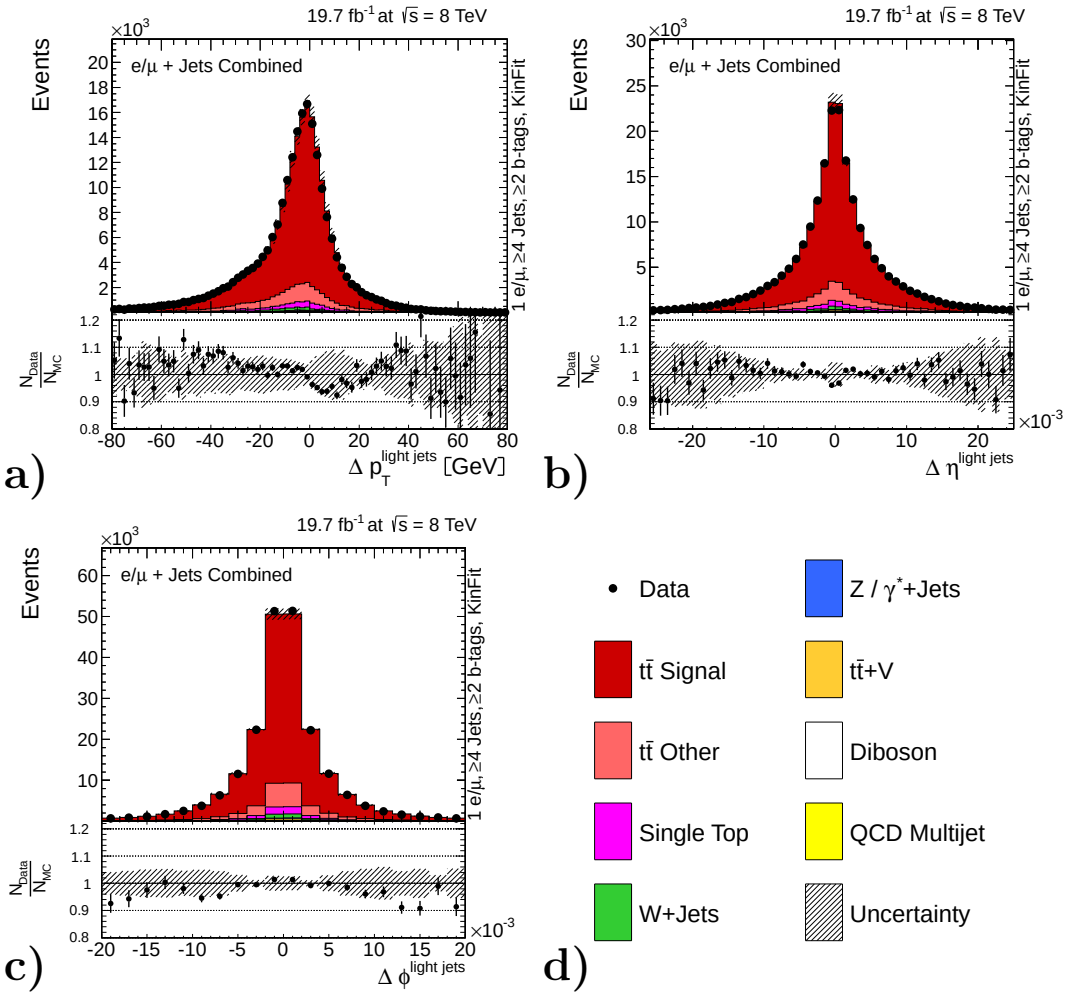

Figure 7.14: Absolute shift $\Delta x = x^{fit} - x^{reco}$ of the kinematic quantities $x = p_T$, η, ϕ for the light jets within the Double Kinematic Fit event reconstruction procedure. The hatched band corresponds to the main experimental and $t\bar{t}$ model uncertainties of the simulation.

7.3.2 Reconstructed W-Boson and Top-Quark Mass

The reconstructed masses of the top quark and the W boson are important control distributions to assure a good understanding of the Double Kinematic Fit and the validity of the input object kinematics in data.

The value for the reconstructed mass of the W boson, which is obtained from the two associated light jets, is used as a constraint in the Double Kinematic Fit. Therefore, the obtained value after the fit is fixed. Nevertheless, it is valuable to study the reconstructed W mass before the fit, e.g. by using the jet association from the fit but the jet kinematics as obtained beforehand. This is also done for other top-quark analyses like CMS top-quark mass measurements [169, 174] that use this distribution to calibrate the JES in-situ.

Good agreement between data and simulation is observed for the reconstructed mass of the W-boson within one sigma of the JES and JER uncertainty as shown in Figure 7.15. This assures that the JES as derived centrally by CMS is also valid for the phase space of this analysis.

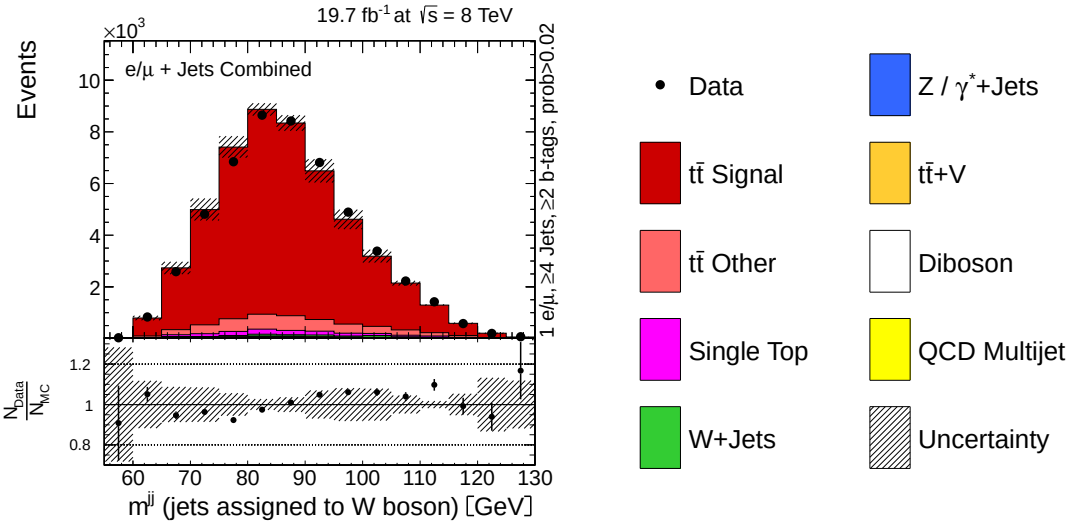

Figure 7.15: Reconstructed W-boson mass for events with a successful event interpretation using the Double Kinematic Fit and a minimal χ^2-probability requirement of 2%. The association of the reconstructed jets to the W boson is obtained from the event interpretation while the jet kinematics is taken from the reconstruction, e.g. before the Double Kinematic Fit. The hatched uncertainty band for the simulation contains JES and JER uncertainties.

Furthermore, the obtained top-quark mass m_{top} as obtained from the second fit of the Double Kinematic Fit approach and after a minimal χ^2-probability requirement of 2% is compared between data and simulation in Figure 7.16. Good agreement within the systematic uncertainties in the order of 5-10% is observed, assuring a good

understanding of the event reconstruction and input objects.

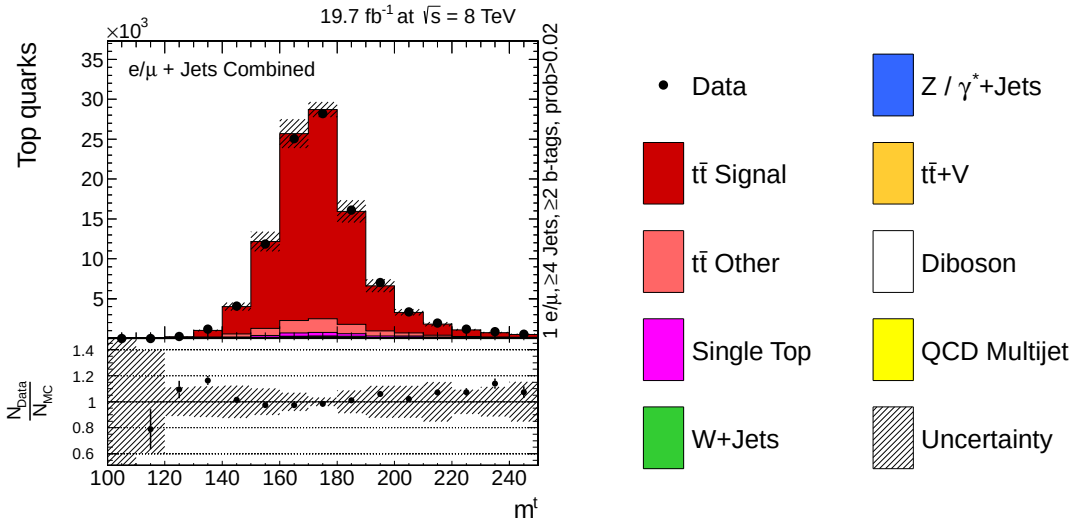

Figure 7.16: Reconstructed top-quark mass for the Double Kinematic Fit and a minimal χ^2-probability requirement of 2%. The hatched band corresponds to the main experimental and $t\bar{t}$ model uncertainties of the simulation.

7.3.3 χ^2 and χ^2-Probability

The most important control distributions for the kinematic event reconstruction are the χ^2 distribution for the final kinematic solution and the corresponding χ^2-probability distribution.

Figure 7.17 shows both distributions for data and simulation as obtained from the Double Kinematic Fit for the best solution. Agreement between data and simulation is found within one sigma of the main uncertainties which are shown as hatched band for the simulation. The absolute agreement is in the order of 5-10%.

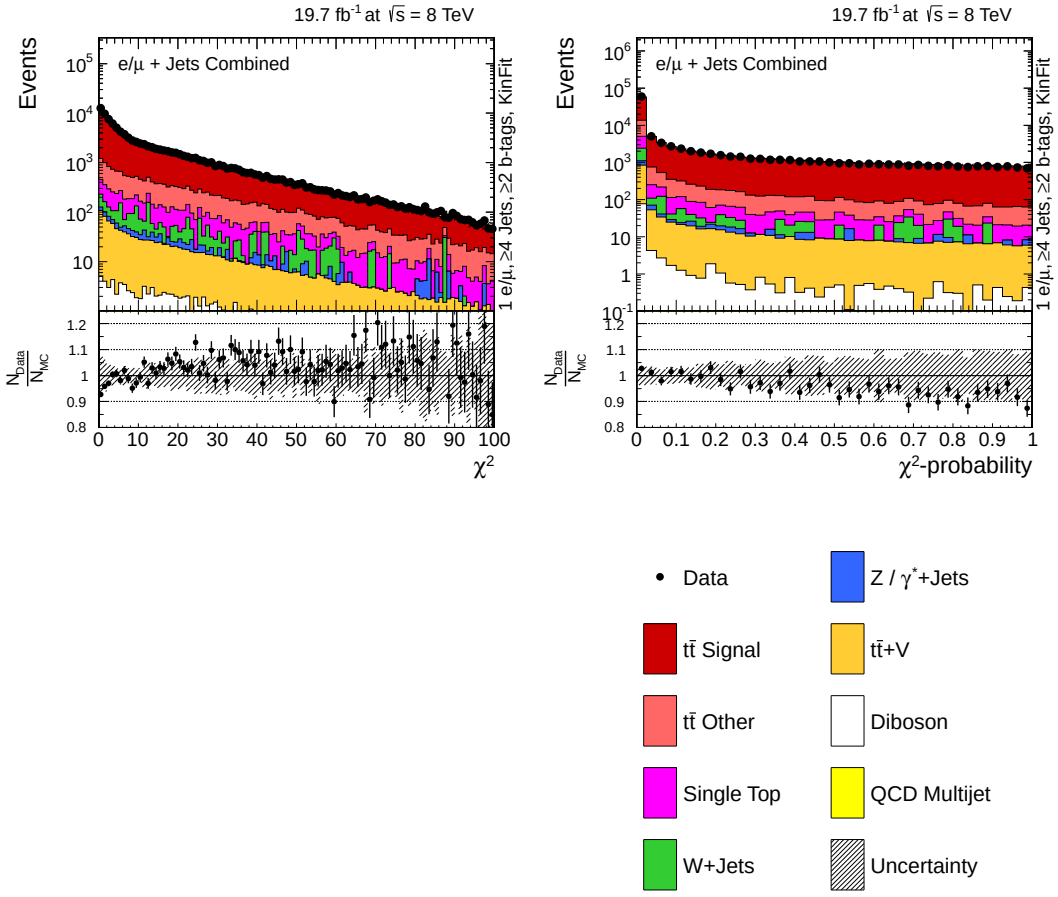

Figure 7.17: χ^2 (left) and χ^2-probability (right) as obtained from the Double Kinematic Fit event interpretation and reconstruction for the best solution.

In the ideal case, the χ^2-probability distribution would be flat for signal events. However, due to background events, signal events with wrong jet association (see Chapter 7.2.1), non-Gaussian tails in the resolutions and experimental cuts, a rise towards lower χ^2-probabilities is observed. The flatness of the χ^2-probability distribution for high probabilities ($\gtrsim 0.3$) and the good agreement between data and simulation justify the modelling and the choice of resolutions.

Furthermore, a minimum χ^2-probability of 2% ($\hat{=}$ maximum χ^2 of 7.824) is required to enhance well-reconstructed events as discussed in Chapter 7.2.3. To assure a good understanding, the efficiency of this selection step is monitored as function of the cross section quantities in Appendix B. Good agreement between simulation and data is found within uncertainties. Therefore, no corrections are required. The same holds for the reconstruction efficiency of the Double Kinematic Fit itself.

As shown in Figure 7.18, the agreement of 3% between the total number of events observed in data and predicted by the simulation (using the $t\bar{t}$ production cross section at NNLO+NNLL accuracy) improves when requiring a higher value for the minimum χ^2-probability. For P(χ^2)>20% or higher, the agreement is at the level of 1% and for all

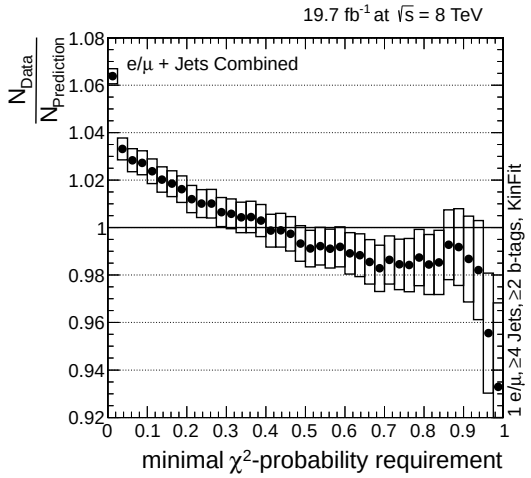

Figure 7.18: Observed ratio of the total number of events observed in data and predicted by the simulation as function of the minimal χ^2-probability requirement. The simulation is normalised to the integrated luminosity using theoretical cross sections. For the expected by far dominant $t\bar{t}$ component, the perturbative QCD calculation at NNLO+NNLL accuracy [44] is used. The uncertainties are statistical uncertainties of the selected number of events in data only.

requirements of P(χ^2)>45% or higher, the data over prediction ratio is approximately constant (0.99).

As already studied in Chapter 7.2.3, the majority of the events that are vetoed with a higher minimum χ^2-probability requirement are expected to originate from non-signal $t\bar{t}$ events (mainly involving τ leptons) and signal events where at least one jet from the $t\bar{t}$ decay is not correctly reconstructed and selected. Therefore, one possible explanation for the improved agreement with data for higher values of the minimum χ^2-probability requirement is the difficulty in modelling those poorly reconstructed events. For example, events passing the kinematic reconstruction although one jet from the $t\bar{t}$ decay is not selected involve always an additional jet. Therefore, the modelling of these events depends much more on the chosen theory parameters of the simulation such as the amount of additional radiation. As suggested by the high fraction of up to 50%

of successfully reconstructed events without jet-parton matching (see Chapter 7.2.1), this is a quite common scenario. Supporting this hypothesis, the jet multiplicity is found to be described best for events with large probability values.
Furthermore, this would also explain the small difference for the absolute efficiency of 2% for the applied minimum χ^2-probability requirement in data and simulation, which results in a 3% improved ratio between the observed number of events in data and the predicted number of events by the simulation.

In summary, good agreement between data and simulation is observed for all tests performed for the Double Kinematic Fit event interpretation and reconstruction and remaining differences between data and simulation are covered by the main uncertainties in all cases. As those uncertainties are propagated through the cross section calculation, they will also be reflected in the final result. Consequently, an additional contribution to the uncertainties due to the minimal χ^2-probability requirement is expected. This contribution is found to be smaller than the reduction of the uncertainties due to the limited migration effects. Therefore, the total uncertainty is finally reduced due to the improved kinematic reconstruction with the Double Kinematic Fit the minimum χ^2-probability requirement.

7.4 Reconstructed Cross Section Quantities

Finally, the kinematic quantities as extracted from the Double Kinematic Fit after a minimum χ^2-probability requirement are used as input for the determination of differential cross sections.

All obtained cross section quantities are shown in Chapter 7.4.1 for the top quarks and $t\bar{t}$ system and in Chapter 7.4.2 for the jets, b jets and lepton. Those distributions as obtained after the full event selection and kinematic reconstruction before any further step of the analysis is in the following referred to as *event yield*.

The normalisation of the single physics processes from the simulation is done as described in Chapter 7.3 using theory cross sections and also the hatched uncertainty bands of the simulation comprises the main uncertainties as described previously. Therefore, the event yields illustrate the compatibility of the shape for the kinematic distributions between data and simulation within the expected major uncertainties. Consequently, the event yields are already a preview for the agreement between data and the prediction of MadGraph+pythia for the final differential cross section results.

7.4.1 Top-Quark and Top-Quark-Pair-System Quantities

In Figure 7.19, the event yields for the transverse momentum (p_T^t) of both top quarks is shown (a) in the detector rest frame, (b) in the $t\bar{t}$ rest frame ($p_T^{t\ (t\bar{t}\ com)}$) and separated for the top quark with highest p_T ($p_T^{lead\ t}$, c) and the top quark with lowest p_T ($p_T^{sublead\ t}$, d) of each event in the detector rest frame.
Furthermore, the event yields for the rapidity (y^t) of both top quarks (left) and the difference in the azimuthal angle ($\Delta\phi\,(t,\bar{t}) = \phi^t - \phi^{\bar{t}}$) between them (right) are shown in Figure 7.20.
Additionally, Figure 7.21 shows the event yield for (a) the transverse momentum ($p_T^{t\bar{t}}$), (b) the rapidity ($y^{t\bar{t}}$) and (c) the invariant mass ($m^{t\bar{t}}$) of the top-quark-pair system as well as (d) the quantity ρ_S which is derived from the invariant mass ($m_{t\bar{t}j}$) of the $t\bar{t}$ system and an additional jet with transverse momentum above 50 GeV ($\rho_S = \frac{2 \cdot 170\,\text{GeV}}{m_{t\bar{t}j}}$).

In general, reasonable agreement between data and simulation is observed. The uncertainties of the simulation are of the order $\mathcal{O}(10\%)$. A trend towards lower values in data is consistently observed for all transverse momentum distributions related to individual top quarks. Additionally, the rapidity of the individual top quarks tends to be slightly less central for data than predicted by MadGraph+pythia. The agreement between data and MC prediction is discussed in detail for the final differential cross sections with the full uncertainties.

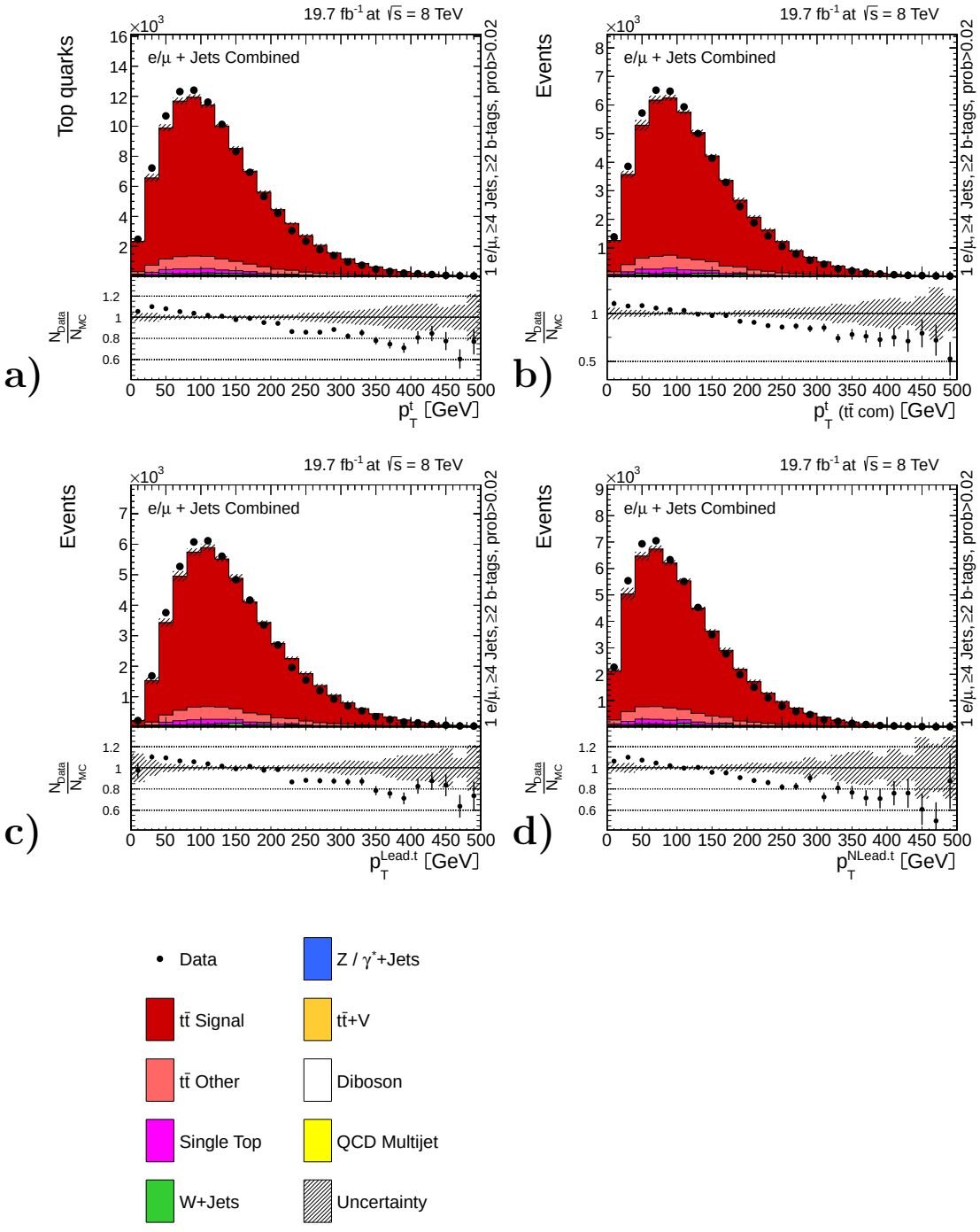

Figure 7.19: Reconstructed event yields for the combined semileptonic muon and electron final states for the top-quark transverse momentum for both top quarks (a), in the $t\bar{t}$ rest frame (b) and separated by their p_T in the top quark with lower momentum (c) and higher momentum (d). All quantities are obtained from the Double Kinematic Fit in data and simulation after imposing a minimal χ^2-probability requirement of 2% and used as input for the calculation of differential cross sections.

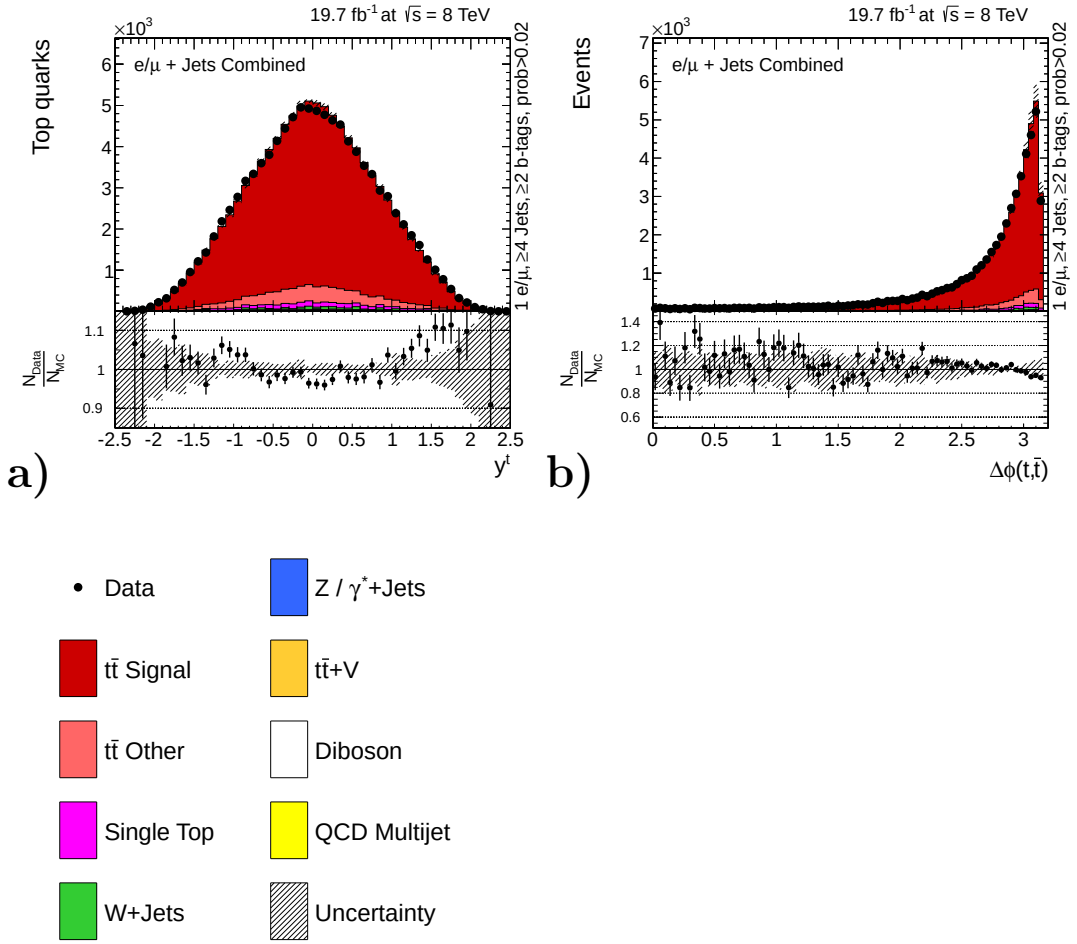

Figure 7.20: Reconstructed event yields for the combined semileptonic muon and electron final states for the top-quark rapidity (left) and the difference in the azimuthal angle between the two top quarks (right). All quantities are obtained from the Double Kinematic Fit in data and simulation after imposing a minimal χ^2-probability requirement of 2% and used as input for the calculation of differential cross sections.

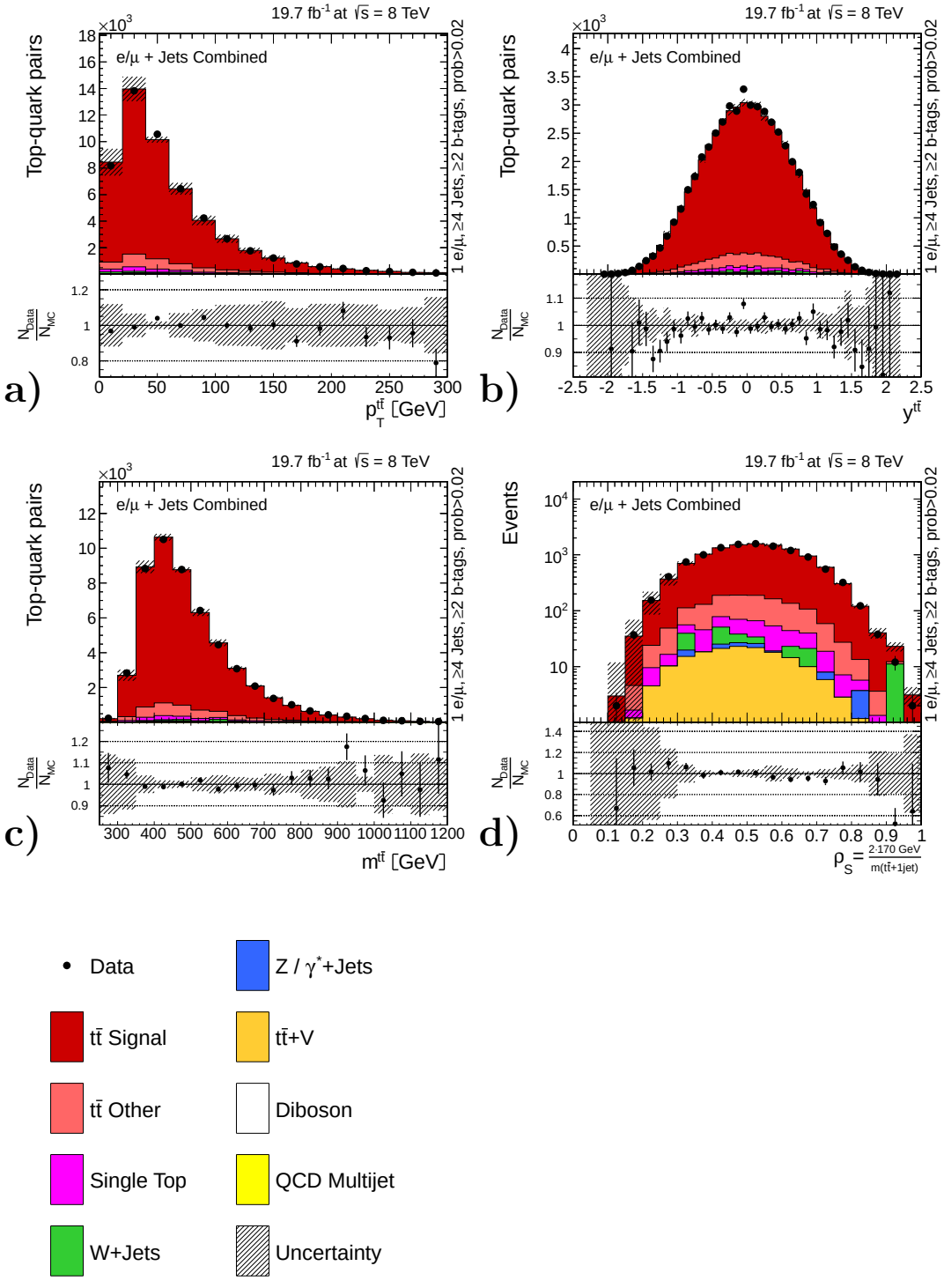

Figure 7.21: Reconstructed event yields for the combined semileptonic muon and electron final states for the $t\bar{t}$ system quantities. All quantities are obtained from the Double Kinematic Fit in data and simulation after imposing a minimal χ^2-probability requirement of 2% and used as input for the calculation of differential cross sections.

7.4.2 Quantities Related to the Lepton, Jets and B Jets

Figure 7.22 shows the event yields for (a) the transverse momentum (p_T^l) and (b) the pseudorapidity (η^l) of the final-state lepton as well as (c) the transverse momentum (p_T^b) and (d) the pseudorapidity (η^b) of the final-state b jets associated to the top-quark decay.
Moreover, the event yields for the jet multiplicity (N_{jets}, a) and the invariant mass of the lepton and b jet associated to the leptonic top-quark decay (m^{lb}, b) as well as the transverse momentum ($p_T^{b\bar{b}}$, c) and the invariant mass ($m^{b\bar{b}}$, d) of the b-jet-pair associated to the top-quark decay are shown in Figure 7.23.

Reasonable agreement between data and simulation is observed. The uncertainties of the simulation are of the order $\mathcal{O}(10\%)$. A trend towards smaller values of the transverse momentum in data is observed for the lepton and the single b jets. This observation is consistent with the trend observed for the transverse momentum of the individual top quarks. Furthermore, this trend was already visible after the b-jet identification selection step (see Figure 6.7a and Figure 6.11c,d in Chapter 6.9) and is therefore not artificially introduced by the kinematic reconstruction and event interpretation.
Moreover, the pseudo rapidities of the lepton and the b jets are observed to be slightly less central in data than predicted by MadGraph+pythia. The agreement between data and prediction is discussed in detail for the final differential cross sections with full uncertainties.

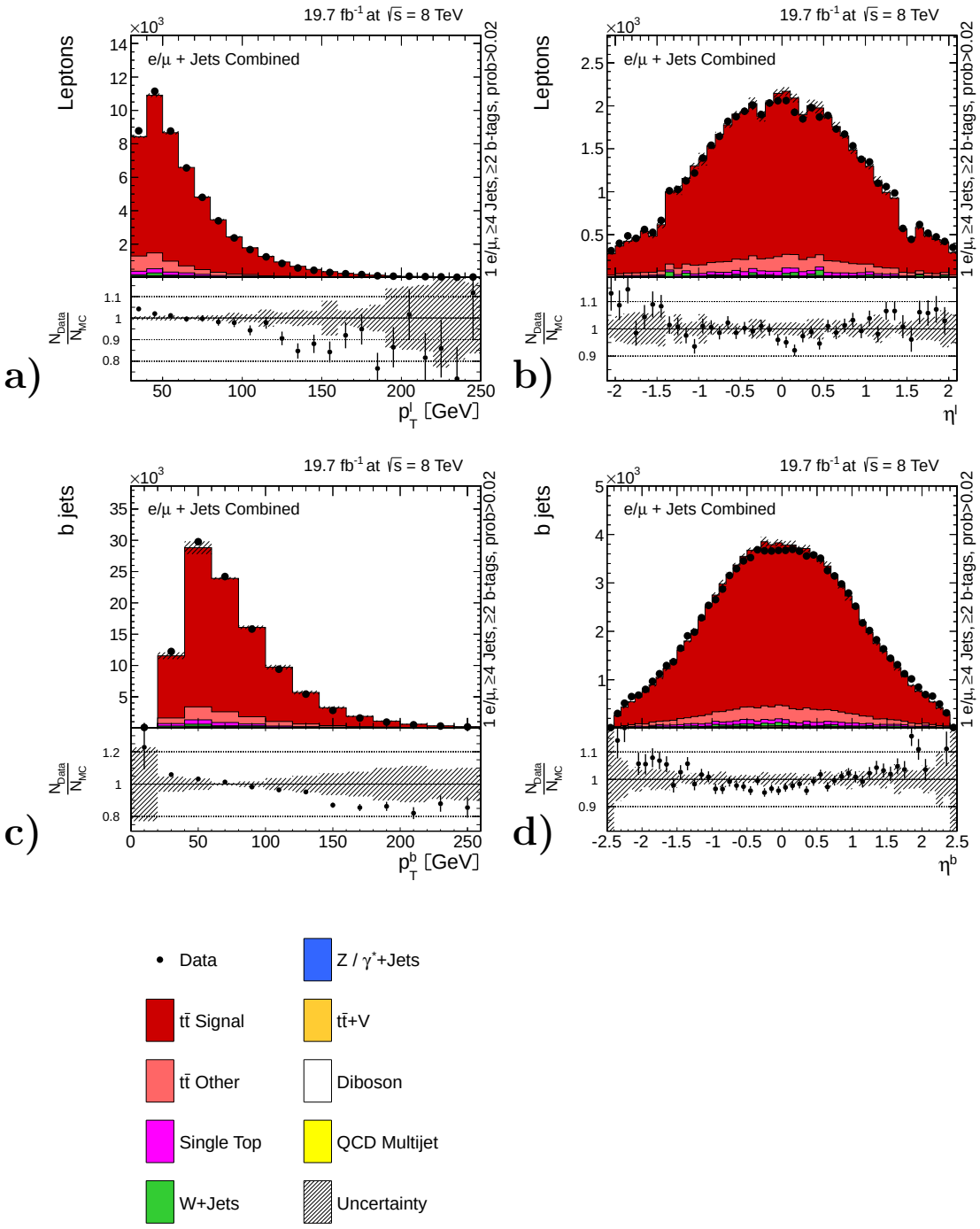

Figure 7.22: Reconstructed event yields for the combined semileptonic muon and electron final states for the lepton (a,b) and b jet (c,d) quantities. All quantities are obtained from the Double Kinematic Fit in data and simulation after imposing a minimal χ^2-probability requirement of 2% and used as input for the calculation of differential cross sections.

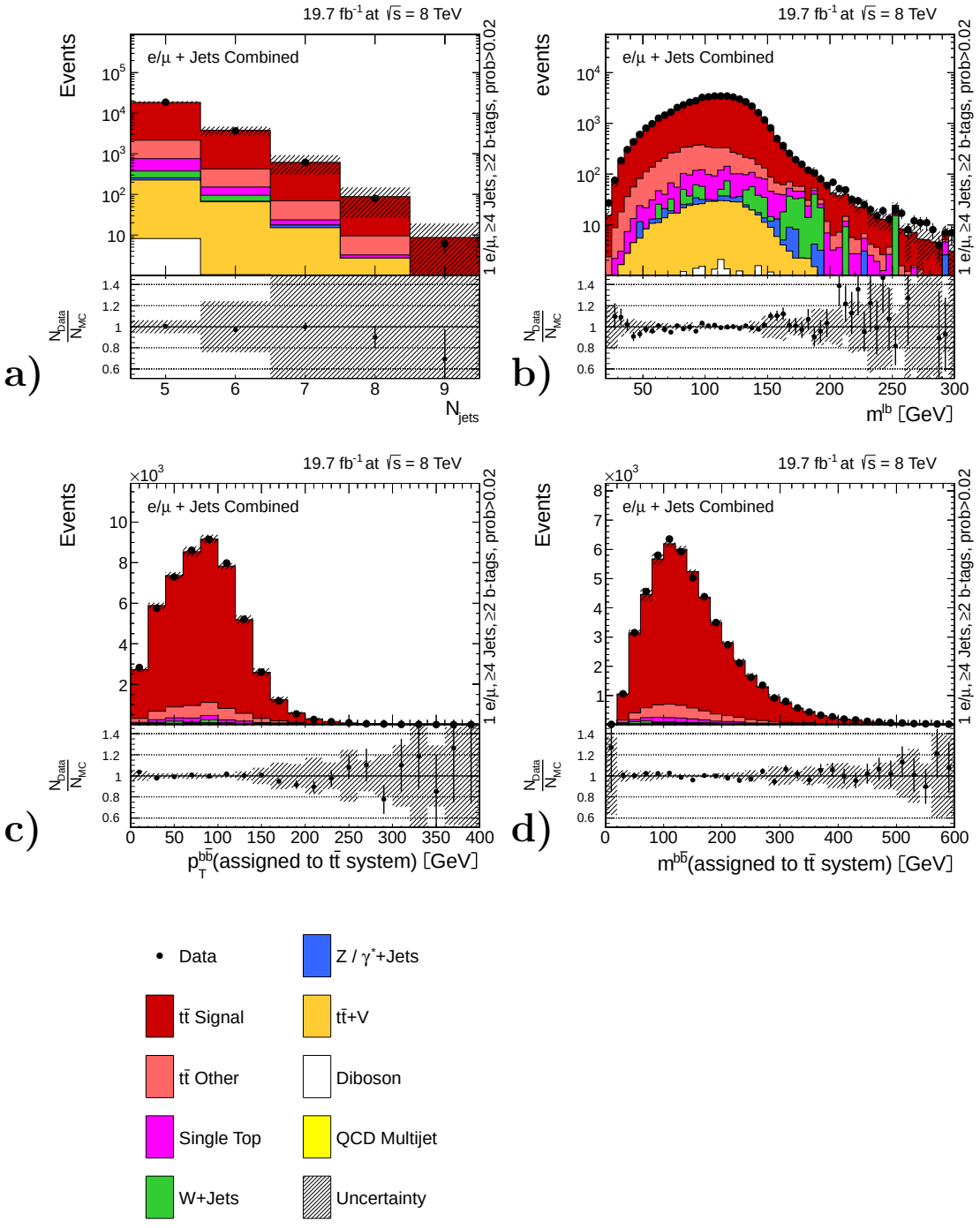

Figure 7.23: Reconstructed event yields for the combined semileptonic muon and electron final states for the jet multiplicity (a), the invariant mass of the lepton and the leptonic b jet associated to the top-quark decay (b) and the quantities of the $b\bar{b}$ system associated to the top-quark decay (c,d). All quantities are obtained from the Double Kinematic Fit in data and simulation after imposing a minimal χ^2-probability requirement of 2% and used as input for the calculation of differential cross sections.

Chapter 8

Cross Section Definitions

Calculating differential cross sections from the event yields of the reconstructed kinematic quantities after the Double Kinematic Fit involves several analysis steps.

Every cross section measurement is done for a well defined kinematic region (*phase space*) which represents the underlying physics process and the corrections that are applied to the measured detector level objects. The phase space for each cross section measurement is defined by the multiplicity and the kinematic properties of the true physics objects of the $t\bar{t}$ production (top quarks) or $t\bar{t}$ decay (leptons, quarks, hadrons or gen jets). In this context, the abbreviations *true* and *gen* refer to objects at the level before detector reconstruction. Reconstructed objects in the detector are in the following labelled *rec*. An unambiguous definition for each true physics object and the phase spaces for each cross section measurement is given in Chapter 8.1.
Furthermore, the procedure of calculating cross sections in a specific phase space from the reconstructed event yields is explained in Chapter 8.2 and 8.3. Another crucial part for the calculation of differential cross sections is the unfolding, which will be detailed in Chapter 9.

8.1 Phase Space Definition

The major aspects characterising each phase space definition are the objects which are involved and the specific kinematic range requested. An unambiguous definition for each object involved in the definition of cross section quantities is needed. Here, two different classes of objects are used: *parton level* objects as produced in the partonic interaction (e.g. quarks) and *particle level* objects defined similar to the corresponding reconstruction level objects (e.g. gen jets).
These objects can be used to define a *visible phase space* which is defined by a minimal requirement of the overall event activity, e.g. how many particles with specific kinematic properties need to be present in the event to consider it as "inside" the phase space. In contrast, an *extrapolated phase space* definition contains all events without specific requirements on the kinematic properties of the objects.

For this thesis, two different phase space definitions are used, involving different correction steps of the measured reconstructed event yields. As detailed in Chapter 8.1.1, the top-quark and $t\bar{t}$ system quantities are corrected for detector and hadronisation effects and extrapolated to the full phase space, resulting in parton level cross section distributions, which are directly comparable to perturbative QCD calculations. In contrast to this, all cross section measurements involving final-state objects (leptons or jets), are only corrected for detector effects and measured in a visible particle level phase space. As detailed in Chapter 8.1.2, this minimises the applied corrections.

In the following, *generator truth* and *true spectrum* refers to the corrected distributions of the corresponding phase space definition of each cross section quantity while *reconstruction level event yield* refers to the reconstructed distributions from the detector information.

8.1.1 Extrapolated Parton Level Phase Space

This phase space definition is used for the differential cross section measurement of all top quark and $t\bar{t}$ quantities. From the physical point of view, it is based on top quarks after radiation directly before their decay, following the factorisation theorem (see Chapter 3.1.1). Performing a cross section measurement in this phase space involves correcting detector and hadronisation effects of the reconstructed top quarks. Furthermore, an extrapolation from the limited kinematic acceptance of the measurement to the full acceptance is performed.

Therefore, a measurement in this phase space involves many corrections derived from the simulation. These corrections are based on information like the inter-bin migration of signal events, which is contained in the covariance matrix used within the unfolding step (see Chapter 9).
The advantage of this phase space definition is its well defined status which enables an easy and direct comparison with available perturbative QCD theory predictions, e.g. to probe the effect of including higher orders in the calculation. Additionally, it allows a comparison between measurements of different $t\bar{t}$ final states or experiments as different acceptance definitions are corrected. Even a combination of different measurements is feasible although this holds additional complications like for example the aspect of different binnings or the need of a precise knowledge of the correlations.

Top-Quark Definition

A precise and consistent definition of the top quark is essential to compare the obtained results with any other measurement or with a prediction. It has to be ensured in the first step that the comparison is done on equal footings before any conclusion can be drawn. This discussion is similar to the question of the compatibility between the measured top-quark mass and the predicted top-quark mass of a fixed order QCD calculation [177].

Therefore, the physics meaning of the technical implementation of the top quarks is

investigated for all used CMS MC predictions. Furthermore, the findings are compared to the MC predictions of the ATLAS experiment within the *Top-quark LHC working group* (TopLHCWG) [178].

From the physics point of view, the top-quark definition characterises the two top quarks produced from the pp initial state after all radiation directly before their decay.

From the technical point of view, the definition of the top quark is based on the top quarks listed in the decay chain event record of the simulated MC event samples. As the simulation is done in several steps, several top quarks are listed, labelled by a specific status code. These status codes differ for different simulations and depend on the choice of the MC generator for the parton showering.

For the default MadGraph +pythia MC prediction which is used to perform the differential cross section measurement, Pythia6 is used for the parton showering. Here, the top quark after radiation directly before its decay is labeled as "status 3" (see Table 8.1, top). Gluons stemming from additional radiation are produced in the first step of this status code chain. Additional gluons produced in the decay step of the top quark ($t \rightarrow Wb$) are expected to originate from the radiation of the b quark produced by the decaying top quark.

For several comparisons and the evaluation of the systematic uncertainties, also simulations involving Herwig6 for the parton shower modelling are used. For these simulations, "status 155" labels the top quarks after radiation and directly before their decay [179]. In comparison to Pythia6, additional intermediate steps are listed in the status code chain for the top quarks. Gluons from additional radiation are produced within the step 144/143→3 in the status code chain of the Herwig6 status codes. For all MC predictions which use Herwig6, it was explicitly checked that the kinematic properties of the Herwig status code 3 top quarks are equal to the ones of the status 155 top quarks. Similar to Pythia, additional gluons produced in the decay step of the top quark are expected to originate from the radiation of the b quark produced in the top-quark decay.
In order to make the treatment of simulations using pythia and herwig for the parton showering more uniform, a tool is used within CMS for Herwig6 to emulate the Pythia6 status codes. Therefore, all official CMS simulations have only the emulated status codes stored. Therefore, status 3 top quarks are consistently used for all MC prediction using pythia and herwig with status code emulation as they label top quarks after radiation but before decay.
Table 8.1 shows the detailed status code chain of the top-quark evolution with (bottom) and without (middle) Pythia status code emulation. It was carefully checked for all MC predictions used in the analysis, that the status codes follow the scheme listed in Table 8.1. The choice of the MC generator for the ME calculation step has no influence on the status code labelling.
Therefore, this top-quark definition based on status code 3/155 seems to be a distinct definition for all $t\bar{t}$ simulations and a comparison between different measurements as well as between different experiments is feasible. To prove this, it was checked that the obtained kinematic top-quark quantities for the same simulation agree between

Status code chain for the production and decay of $t\bar{t}$ pairs
– **Pythia6:**
$gg \to t/\bar{t}\ (3) \to W\ b/\bar{b}$
– **Herwig6:**
$gg \to \bar{t}\ (124) \to \bar{t}\ (144) \to \bar{t}\ (3) \to \bar{t}\ (155) \to W\ \bar{b}$
$gg \to t\ (123) \to t\ (143) \to t\ (3) \to t\ (155) \to W\ b$
– **Herwig6 with status code emulation:**
$gg \to \bar{t}\ (2) \to \bar{t}\ (2) \to \bar{t}\ (3) \to \bar{t}\ (2) \to W\ \bar{b}$
$gg \to t\ (2) \to t\ (2) \to t\ (3) \to t\ (2) \to W\ b$

Table 8.1: Status code decay chain from the production to the decay of the top quarks as listed in the event record of $t\bar{t}$ simulations using Pythia6 (top), Herwig6 (middle) or Herwig6 with Pythia status code emulation (bottom) for the parton showering. Status 3 top quarks are consistently used for this thesis.

ATLAS and CMS [178]. This was a priori not obvious and is one of the improvements in understanding to which the work documented in this thesis contributed. For other simulations involving for example different programs for the parton showering (like Herwig++ or Pythia8), the status code chain and therefore the technical definition of the top quark is different.

Besides the comparability between different simulations, the main question is whether this top-quark definition is comparable to the top quarks from a higher order perturbative QCD calculation. The comparison stays valid as long as the defined top quark is consistently the one directly before its decay after all radiation.

A follow-up study of the additionally radiated gluons in the decay step of the top quark might contribute further to a more precise understanding. Due to colour-connection effects, there is a non-trivial interference between the final state radiation of the top quark and the bottom quark. As all MC predictions investigated within this thesis do not include additional radiation within the decay, this interference is not taken into account. It is not clear how this would change the kinematics of the top quark but it is consistently not incorporated in the status 3 top-quark definition and the perturbative QCD predictions used for further comparison. Furthermore, this is exactly where the definition of the top-quark "truth" becomes ambiguous as effects of the production and decay of the top quarks are mixed.

Event Definition

In principle, it is also possible to define a visible parton level phase space like it is e.g. done in [23] by defining a kinematical acceptance for the parton level objects. For this thesis, only the extrapolated parton level phase space definition is used, which involves an extrapolation to the full kinematic phase space. Therefore, this phase space includes all $t\bar{t}$ signal events because each event contains per definition two top quarks as characterised above.

8.1.2 Visible Particle Level Phase Space

This phase space definition is used for all differential cross section measurements as a function of final state objects (leptons and jets). Those are directly measured within a limited but unambiguously defined kinematic acceptance (i.e. the p_T and η selection criteria). This phase space definition is supposed to be as close to the reconstruction level as possible. Only detector effects are corrected and gen jets are used for the definition of the particle level phase space as generator truth equivalent of the jets at reconstruction level. Furthermore, this phase space includes only events containing the same number of final state objects within the identical kinematic acceptance (p_T, η) as used for the reconstruction level selection (see Chapter 6.2 to 6.5.2). Defining a "visible" (i.e. in the detector accessible) acceptance region is naturally possible for the lepton and jets while the top quark itself is only defined at parton level[a].

Restricting the phase space to the visible region and using object definitions close to the ones at reconstruction level minimises the corrections derived from simulation which have to be applied to the reconstructed event yields as much as possible. Furthermore, an extrapolation into unmeasured regions of the total kinematic acceptance is avoided. On the other hand, events may migrate from outside the visible phase space into the reconstructed event sample. This type of migration is not present for the extrapolated parton level phase space. Another disadvantage of this visible particle level phase space definition is that the obtained results cannot be directly compared to QCD theory calculations. Moreover, also the comparison to other measurements in different $t\bar{t}$ final states is not possible and the comparison to other measurements in the same final state is only possible if identical object definitions are used. Therefore, the comparison to other experiments will be difficult as e.g. the choice of kinematic acceptances is often constrained by the available triggers.

Consequently, the obtained results in the visible particle level phase space are mainly used to compare different MC predictions and are therefore a helpful validation for the modelling of the kinematic properties of the $t\bar{t}$ final state objects. This can be important for all searches where $t\bar{t}$ production is considered as background process, e.g. Supersymmetry analyses or $t\bar{t}$ +Higgs analyses.

Visible Kinematic Acceptance

The definition of the kinematic acceptance follows for all objects the reconstruction level definition (see Chapter 6):

- gen lepton acceptance: $p_T > 33\,\text{GeV}$, $|\eta| < 2.1$
- gen jet acceptance: $p_T > 30\,\text{GeV}$, $|\eta| < 2.4$

[a] One alternative approach to define top quarks within a visible phase space is the so-called "pseudo top concept" discussed in [180] where the top quarks are basically defined as sum of the four momenta of their visible final state decay products.

Lepton Definition

As visible gen lepton, the electron or muon from the list of generated particles is chosen that originates directly from the decay of the W boson produced by the top-quark decay before any additional (photon) radiation. The advantage of a lepton definition before radiation is the equality of muon and electron. Otherwise, the electron momenta would be on average softer due to the higher probability to radiate a photon. Using the leptons before this radiation step facilitates the combination of the two final states.

Technically, the lepton from the W-boson decay before additional radiation is labelled as "status 3 lepton" in the event record. To distinguish leptons produces in other processes (e.g. the decay of tauon), only leptons with a status code 3 W boson as ancestor are considered. Similar to the top-quark status code definition, status code 3 holds for simulations that use Pythia6 for the parton shower or Herwig6 with the Pythia status code emulation while the initial Herwig6 status code is different (123). All status codes for the leptons can be found in Table 8.2.

Status code chain for lepton production from top-quark decay
– **Pythia6:** top-quark decay → W (3) → W-boson decay W-boson decay → lepton (3) → lepton (1)
– **Herwig6:** top-quark decay → W (123) → W (3) → W (155) → W-boson decay W-boson decay → lepton (123) → lepton (1)
– **Herwig6 with Pythia status code emulation:** top-quark decay → W (2) → W (3) → W (2) → W-boson decay W-boson decay → lepton (3) → lepton (1)

Table 8.2: Status code decay chain for the production of leptons from the decay of top quarks as listed in the event record of $t\bar{t}$ simulations using Pythia6 (top), Herwig6 (middle) or Herwig6 with Pythia status code emulation (bottom) for the parton showering. The first lepton in the table is before, the last after additional radiation. Additional decay products like b quarks from the top-quark decay and neutrinos from the W-boson decay are not listed.

Visible Gen Jet Definition

Gen jets are obtained by applying the same anti $-k_T$ ($R = 0.5$) cluster algorithm as for reconstructed jets (see Chapter 6.4) to all generated, stable final-state particles. Neutrinos, electrons, muons originating from a resonance (top quark, W boson, Z boson or photon) are excluded from the clustering step. Moreover, a requirement of $\Delta\mathrm{R}(\text{gen lepton}, \text{gen jet}) > 0.4$ is applied to remove gen jets originating from photons radiated from the leptons. This is in particular important in the e+jets channel. All gen jets fulfilling the jet acceptance requirement as defined above are called visible gen jets and considered further.

Visible Gen B Jet Definition

From all visible gen jets, the one is identified as visible (anti)b jet that contains an (anti)B hadron originating from the (anti)bottom quark of the top-quark decay. If the same gen jet contains both B hadrons or both B hadrons cannot be successfully associated to visible gen jets as defined above, the event is not considered for this phase space. Without acceptance requirement for the gen jets, this b jet identification procedure is found to give a result for more than 99% of all events. This number includes events where the same gen jet is identified as b and antib gen jet that are finally excluded from the phase space definition.

Visible Gen Light Jet Definition

From all visible gen jets that are not identified as visible b jets, the pair of gen jets is identified as visible light jets from the hadronically decaying W boson for which the invariant mass is closest to the W-boson mass of 80.4 GeV. To suppress picking up a jet not originating from the W-boson decay (for example in case one of the real jets from the W-boson decay is outside the kinematic acceptance), a requirement of $65\,\text{GeV} < m_{jj} < 105\,\text{GeV}$ is introduced. The values for the size of the W-boson mass window was concluded from the invariant mass distribution of the reconstructed light jets associated to the W boson by the event reconstruction for simulated $t\bar{t}$ signal events with correct jet association. If no pair of visible gen jets is fulfilling this requirement, the event is not considered for this phase space.

Additional Gen Jet Definition

An additional jet is only needed to derive the cross section quantity ρ_S. From the visible gen jets, which are neither identified as b gen jets nor as light gen jets, the gen jet with the largest p_T value is chosen as additional jet. Following [49], it is analogously only considered for the calculation of ρ_S if its transverse momentum is at least 50 GeV.

Event Definition

A $t\bar{t}$ signal event (i.e. ℓ+jets final state) is considered for the visible particle level phase space, if it contains the same number of visible generator level objects like used for the reconstruction level definition:

- exactly one visible gen lepton (e or μ)
- at least four visible gen jets
- two identified visible gen b jets
- two identified visible gen light jets

This common global event definition for all quantities involves all final state objects that are also used within the kinematic reconstruction procedure for the measurement. Consequently, the same events are consistently used for all cross section definitions at

generator truth level.

The visible particle level phase space definition contains about 15% percent of all signal events in the semileptonic final state involving muons or electrons. Furthermore, only ≈1% of all signal events fulfil the phase space requirement and contain an additional gen jet with $p_T > 50\,\text{GeV}$. Detailed numbers for the acceptance can be found in Table 8.3.

	$\mathbf{N_{gen}^{events}}$		**Acceptance**
	(e+jets)	(μ+jets)	(ℓ+jets)
total, no selection	9,206,900	9,208,500	100%
≡1 visible gen lepton	5,471,500	5,475,200	59%
+ ≥4 visible gen jets	2,615,300	2,643,100	29%
+ ≡2 visible gen b jets	2,163,500	2,195,900	24%
+ ≡2 visible gen light jets	1,411,600	1,434,800	15%
+ ≥1 additional gen jet			
− with $p_T > 30\,\text{GeV}$	685,400	699,200	8%
− with $p_T > 50\,\text{GeV}$	124,100	127,300	1%

Table 8.3: Number of events (left and middle) for all steps of the visible particle level phase space definition (top) and for the presence of an additional gen jet (bottom) in semileptonic $t\bar{t}$ final states with electron (left) or muon (middle) obtained from the MadGraph+pythia MC prediction. The acceptance of the combined ℓ+jets channel is shown in the last column.

The fraction of reconstructed events (N^{rec}) which fulfil the visible particle level phase space definition is:

$$\frac{N^{rec}_{\in \text{visible particle level phase space}}}{N^{rec}} = 75\%. \tag{8.1}$$

This number is a measure of the amount of migration from outside the PS definition into the sample of reconstructed events.

8.2 Inclusive Cross Section

The calculation of the total inclusive cross section follows the idea of a counting experiment. For an observed number of $t\bar{t}$ events ($N_{t\bar{t}}$), the total inclusive cross section ($\sigma_{t\bar{t},\mathrm{tot}}$) is obtained by dividing by the integrated luminosity ($\mathscr{L}_{\mathrm{int}}$) of the analysed data sample:

$$\sigma_{t\bar{t},\mathrm{tot}} = \frac{N_{t\bar{t}}}{\mathscr{L}_{\mathrm{int}}} \tag{8.2}$$

Due to the imperfectness of the detector, $N_{t\bar{t}}$ cannot be obtained directly. In this thesis, only the $t\bar{t}$ events decaying into semileptonic final states with muon or electron ($N_{t\bar{t}\to e/\mu+\mathrm{jets}}$) are selected. To compensate for this, the BR for the $t\bar{t}$ decay into these final states ($BR_{t\bar{t}\to e/\mu}$) is used:

$$N_{t\bar{t}\to e/\mu+\mathrm{jets}} = N_{t\bar{t}} \cdot BR_{t\bar{t}\to e/\mu} \tag{8.3}$$

Furthermore, due to inefficiencies of the reconstruction (ϵ_{reco}) and the limited kinematic acceptance (Acc) of the detector, only a fraction of events ($N^{t\bar{t}\to e/\mu+\mathrm{jets}}_{\mathrm{reco,selected}}$) is reconstructed and selected:

$$N^{t\bar{t}\to e/\mu+\mathrm{jets}}_{\mathrm{reco,selected}} = N_{t\bar{t}\to e/\mu+\mathrm{jets}} \cdot \epsilon_{\mathrm{reco}} \cdot \mathrm{Acc} \tag{8.4}$$

The exact separation of efficiency and acceptance is a matter of definition and for the calculation of the total inclusive cross section, only the product of the two matters.

For this thesis, efficiency and acceptance are calculated from the simulation, using the phase space definitions given in Chapter 8.1. For this purpose, all data-driven correction factors as detailed in Chapter 6.7 are applied to the simulation. The calculation involves the number of reconstructed events ($N^{\mathrm{events,sim}}_{\mathrm{reco,selected}}$), the number of events contained in the visible particle level phase space ($N^{\mathrm{events,sim}}_{\mathrm{gen\ vis.}}$) and the number of events in the extrapolated parton level phase space ($N^{\mathrm{events,sim}}_{\mathrm{gen\ all}}$), which is equal to the total number of generated $t\bar{t}$ signal events:

$$\mathrm{Acc} = \frac{N^{\mathrm{events,sim}}_{\mathrm{gen\ vis.}}}{N^{\mathrm{events,sim}}_{\mathrm{gen\ all}}} \tag{8.5}$$

$$\epsilon_{\mathrm{reco}} = \frac{N^{\mathrm{events,sim}}_{\mathrm{reco,selected}}}{N^{\mathrm{events,sim}}_{\mathrm{gen\ vis.}}} \tag{8.6}$$

$$\longrightarrow \mathrm{Acc} \cdot \epsilon_{\mathrm{reco}} = \frac{N^{\mathrm{events,sim}}_{\mathrm{reco,selected}}}{N^{\mathrm{events,sim}}_{\mathrm{gen\ all}}} \tag{8.7}$$

Moreover, the selected number of events in data ($N^{\mathrm{events,data}}_{\mathrm{reco,selected}}$) contains not only $t\bar{t}$ signal events but also background events. Therefore, a correction for the expected number of background events is needed. Within this thesis, two different types of

background processes are distinguished. The expected contribution from non-$t\bar{t}$ events ($N^{\text{non}-t\bar{t}\text{ events,sim}}_{\text{reco,selected}}$) is small (approximately 5% of all events for all background processes, see Chapter 6.8). Therefore, this contribution is estimated from the simulation, using theory cross sections (see Chapter 5.3.2). The obtained expectation is directly subtracted:

$$N^{t\bar{t}\text{ events,data}}_{\text{reco,selected}} = N^{\text{events,data}}_{\text{reco,selected}} - N^{\text{non}-t\bar{t}\text{ events,sim}}_{\text{reco,selected}} \tag{8.8}$$

In contrast to this, the expected contribution from $t\bar{t}$ events originating from a different final state than e/μ+jets (approximately 7% of all events) is corrected for in a multiplicative approach:

$$N^{t\bar{t}\to e/\mu\text{ events,data}}_{\text{reco,selected}} = N^{t\bar{t}\text{ events,data}}_{\text{reco,selected}} \cdot \underbrace{\frac{N^{t\bar{t}\to e/\mu\text{ events,sim}}_{\text{reco,selected}}}{N^{t\bar{t}\to e/\mu\text{ events,sim}}_{\text{reco,selected}} + N^{t\bar{t}\text{ other events,sim}}_{\text{reco,selected}}}}_{f_{t\bar{t}\text{ SG}}} \tag{8.9}$$

The $t\bar{t}$ signal fraction $f_{t\bar{t}\text{ SG}}$ characterises the fraction of $t\bar{t}$ signal events ($N^{t\bar{t}\to e/\mu\text{ events,sim}}_{\text{reco,selected}}$) among all $t\bar{t}$ events ($N^{t\bar{t}\to e/\mu\text{ events,sim}}_{\text{reco,selected}} + N^{t\bar{t}\text{ other events,sim}}_{\text{reco,selected}}$) as predicted by the MC event generators at reconstruction level after all selection steps. The main advantage of a multiplicative correction is the independence of the total normalisation, e.g. the value for $\sigma_{t\bar{t},\text{tot}}$ is not needed. Furthermore, for all uncertainties propagated through $f_{t\bar{t}\text{ SG}}$ to the final result, correlated uncertainties between $t\bar{t}$ signal and $t\bar{t}$ other events cancel, resulting in a reduced systematic uncertainty.

In summary, the total inclusive $t\bar{t}$ production cross section can be obtained as:

$$\sigma_{t\bar{t},\text{tot}} = \frac{f_{t\bar{t}\text{ SG}} \cdot \left(N^{\text{events,data}}_{\text{reco,selected}} - N^{\text{non}-t\bar{t}\text{ events,sim}}_{\text{reco,selected}} \right)}{\epsilon_{\text{reco}} \cdot \text{Acc} \cdot \mathscr{L}_{\text{int}} \cdot \text{BR}_{t\bar{t}\to e/\mu}} \tag{8.10}$$

The visible inclusive cross section ($\sigma_{t\bar{t}\to X,\text{vis}}$) is defined by dropping the acceptance correction and the BR correction to restrict the measurement to the kinematic accessible region of the detector in the corresponding final state X:

$$\begin{aligned} \sigma_{t\bar{t}\to X,\text{vis}} &= \frac{f_{t\bar{t}\text{ SG}} \cdot \left(N^{\text{events,data}}_{\text{reco,selected}} - N^{\text{non}-t\bar{t}\text{ events,sim}}_{\text{reco,selected}} \right)}{\epsilon_{\text{reco}} \cdot \mathscr{L}_{\text{int}}} \\ X &\in (\text{e+jets, } \mu\text{+jets, e/}\mu\text{+jets}) \end{aligned} \tag{8.11}$$

Visible and total inclusive cross section are measured in this thesis.

8.3 Normalised Differential Cross Sections

The calculation of differential cross sections for all kinematic quantities x follows Equation 8.10 for the extrapolated parton level phase space and Equation 8.11 for the visible particle level phase space:

$$\frac{\mathrm{d}\sigma}{\mathrm{dx}} = \frac{\mathrm{N}^{\mathrm{t\bar{t}}\to \mathrm{e}/\mu \text{ events,data}}_{\text{unfolded, bin i}}}{\mathscr{L}_{\mathrm{int}} \cdot \Delta \mathrm{x_i} \; (\cdot \mathrm{BR}_{\mathrm{t\bar{t}}\to \mathrm{e}/\mu})} \tag{8.12}$$

Here, $\Delta \mathrm{x_i}$ is the width of the considered bin i of the kinematic quantity x. Moreover, $\mathrm{N}^{\mathrm{t\bar{t}}\to \mathrm{e}/\mu \text{ events,data}}_{\text{unfolded, bin i}}$ is the background-, efficiency-, acceptance- and migration-corrected number of data events in bin i, analogue to $\mathrm{N}_{\mathrm{t\bar{t}}\to \mathrm{e}/\mu+\mathrm{jets}}$ in the definition of the inclusive cross section. Obtaining $\mathrm{N}^{\mathrm{t\bar{t}}\to \mathrm{e}/\mu \text{ events,data}}_{\text{unfolded, bin i}}$ for each bin from the reconstructed event yields is the essential part of the calculation and is done within the unfolding procedure described in Chapter 9.

Finally, the obtained differential cross sections are normalised by their integral to obtain high precision shape measurements, exploiting the cancellation of correlated uncertainties:

$$\frac{1}{\sigma}\frac{\mathrm{d}\sigma}{\mathrm{dx}} \quad , \quad \sigma = \sum_{\substack{\text{measured} \\ \text{bins i}}} \left(\frac{\mathrm{d}\sigma}{\mathrm{dx}}\right)_{\mathrm{i}} \cdot \Delta \mathrm{x_i} \tag{8.13}$$

8.4 Migration and Choice of Binning

The largest difference between a differential and an inclusive cross section measurement is the presence of migration effects. For a kinematic quantity x, it occurs that due to misreconstruction and the limited resolution of the detector the reconstructed value $\mathrm{x}^{\mathrm{reco}}$ is not identical to the true value $\mathrm{x}^{\mathrm{gen}}$. As all differential measurements are done in a finite binning, this difference can lead to a difference in the observed distribution $\mathrm{f}(\mathrm{x}^{\mathrm{reco}})$ with respect to the true distribution $\mathrm{f}(\mathrm{x}^{\mathrm{gen}})$ if $\mathrm{x}^{\mathrm{reco}}$ is reconstructed in a different bin than $\mathrm{x}^{\mathrm{gen}}$ for some events. Therefore, migration effects introduce a correlation between the reconstructed number of events in different bins which has to be resolved to obtain an unbiased estimate for the true distribution. Moreover, migration effects mix with the different efficiencies and acceptances for different bins.

Correcting migration, efficiency and acceptance effects is done in the unfolding step (Chapter 9). Especially the correction of migration effects depends on a good understanding of the underlying process. Consequently, it is beneficial to reduce migration effects as much as possible to obtain a high precision result.
As implied beforehand, the choice of the binning has an impact on the final migration effects. Therefore, it is necessary to quantify migration effects and choose the binning for the final differential cross section measurements according to a criterion to limit migrations.

The two quantities purity (p_i) and stability (s_i) are used to quantify migration effects. Their definition depends on the number of events which are generated in bin i and reconstructed anywhere[b] ($N_{\text{reco}}^{\text{gen in bin i}}$), on the number of events which are reconstructed in bin i and generated anywhere ($N_{\text{reco in bin i}}^{\text{gen}}$) and on the number of events that are generated and reconstructed in the same bin ($N_{\text{reco in bin i}}^{\text{gen in bin i, }\in\text{phase space}}$):

$$p_i = \frac{N_{\text{reco in bin i}}^{\text{gen in bin i, }\in\text{phase space}}}{N_{\text{reco in bin i}}^{\text{gen}}} \tag{8.14}$$

$$s_i = \frac{N_{\text{reco in bin i}}^{\text{gen in bin i, }\in\text{phase space}}}{N_{\text{reco}}^{\text{gen in bin i}}} \tag{8.15}$$

$$p_i, s_i \in [0.0, 1.0] \tag{8.16}$$

Without migration effects, purity and stability would be equal to 1.0. Stability is sensitive to migration out of the bin and purity to migration into the bin. The more pronounced the migration effects are, the smaller will be the values for purity and stability.

The final binning was chosen taking several criteria into account. First of all, purity and stability are required to be at least 40-50% over the full range of every cross section quantity. Then, it is required that the migration into the bins, which are not the next neighbours, are suppressed as much as possible. Dedicated control plots of the migration matrices, describing not only the amount of migration but also the transition probability for each bin, are shown in Appendix C. Finally, it is required that the uncertainty distributions do not significantly spike in any of the bins to suppress non-physically large uncertainties due to a statistical deficit in the sample that is used to assess the uncertainty.

In previous CMS analyses at $\sqrt{s} = 7$ TeV [23–25], the same criteria for the choice of binning were used. Moreover, for all cross section quantities which were already investigated in [23, 25], the identical binning was chosen. In comparison to that analysis, purity and stability are on average about 20% higher due to the Double Kinematic Fit and the minimal χ^2-probability requirement (see Chapter 7.2.4). Therefore, the precision of the final result is expected to be improved.
In addition, it was possible to add an additional bin for high transverse momenta of the top quarks ($400\,\text{GeV} < p_T^t < 500\,\text{GeV}$) in comparison to the analysis at $\sqrt{s} = 7$ TeV. Furthermore, the binning of all other cross section quantities involving also the transverse momentum of the top quarks ($p_T^{\text{lead t}}$, $p_T^{\text{sublead t}}$ and p_T^t ($t\bar{t}$ com)) is chosen to be the same as for p_T^t in order to enable a direct comparison.

Figure 8.1 to 8.4 show the obtained values for purity and stability of all cross section quantities for the combined ℓ+jets channel and the binning used in this thesis. All

[b] Events which are "generated in bin i" are also required to be part of the corresponding phase space definition (for $i > 0$) .
The minimal requirement for events "reconstructed anywhere" is that the Double Kinematic Fit provides a solution for these events which is passing the minimum χ^2-probability requirement.

values are derived from $t\bar{t}$ signal events of the MadGraph+pythia prediction.

Typical values for the purity and stability of the extrapolated parton level cross section quantities are 60-70%. For the visible phase space cross section quantities, the stability is on average higher due to the restriction of the phase space to the visible kinematic acceptance while the purity is on average lower because of events migrating from outside the visible particle phase space definition into the sample of reconstructed events.
For N_{jets}, the natural integer bin width of $\Delta N_{jets} = 1$ fulfils the criteria of $s_i \& p_i > 40 - 50\%$. Events with jet multiplicities of seven and eight are combined in one bin to minimise the statistical uncertainty.
In contrast to this, the purity obtained in the first bin of ρ_S is significantly lower due to a large amount of reconstructed events which are not part of the particle level phase space definition or comprise no additional gen jet. Therefore, the binning is finally chosen by neglecting the first bin and optimising purity and stability for the other bins.

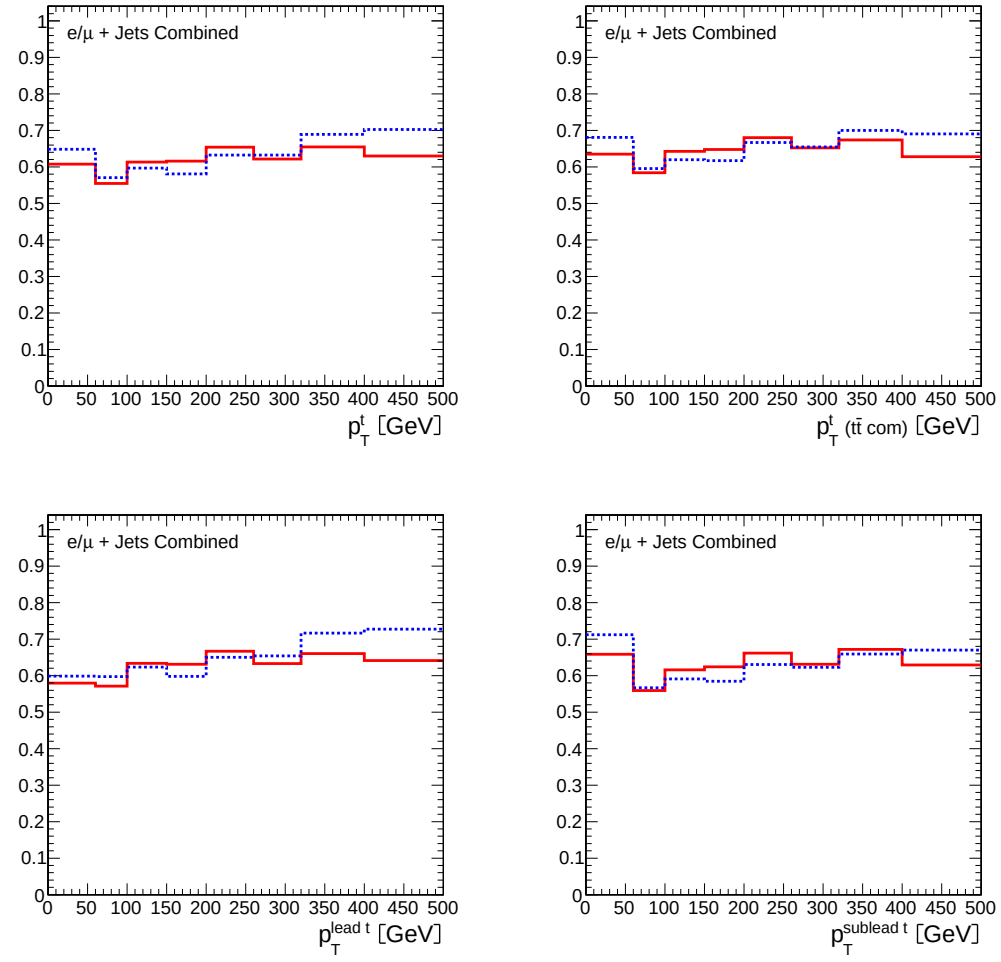

Figure 8.1: Purity (red) and stability (blue) for differential cross section quantities related to the top-quark transverse momentum in the extrapolated parton level phase space to quantify migration effects for the chosen analysis binning in the combined ℓ+jets channel.

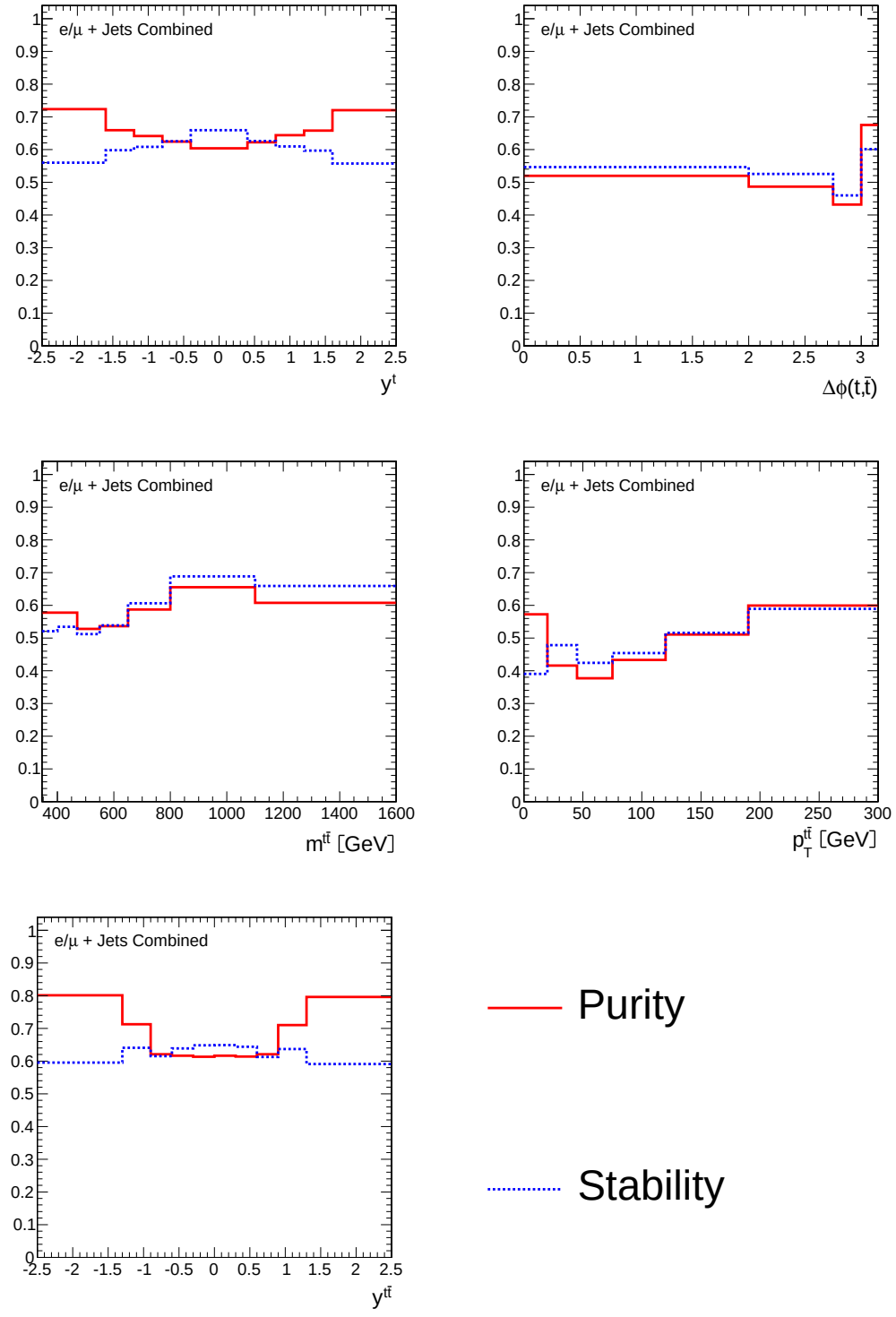

Figure 8.2: Purity (red) and stability (blue) for the differential cross section quantities related to the top-quark (top) and the $t\bar{t}$ system (middle, bottom) in the extrapolated parton level phase space to quantify migration effects for the chosen analysis binning in the combined ℓ+jets channel.

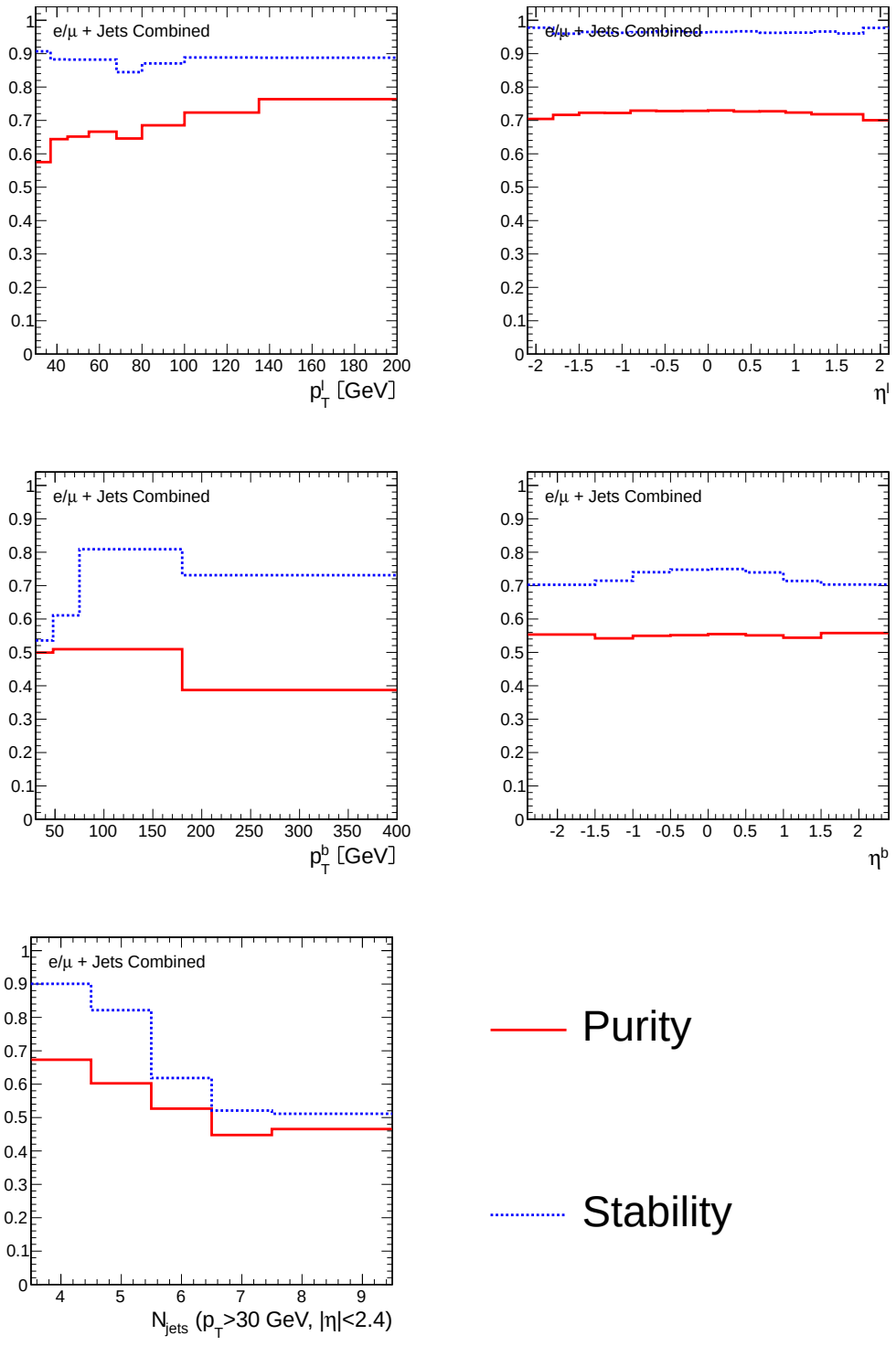

Figure 8.3: Purity (red) and stability (blue) for differential cross section quantities related to the lepton, jets and b jets in the visible particle level phase space to quantify migration effects for the chosen analysis binning in the combined ℓ+jets channel.

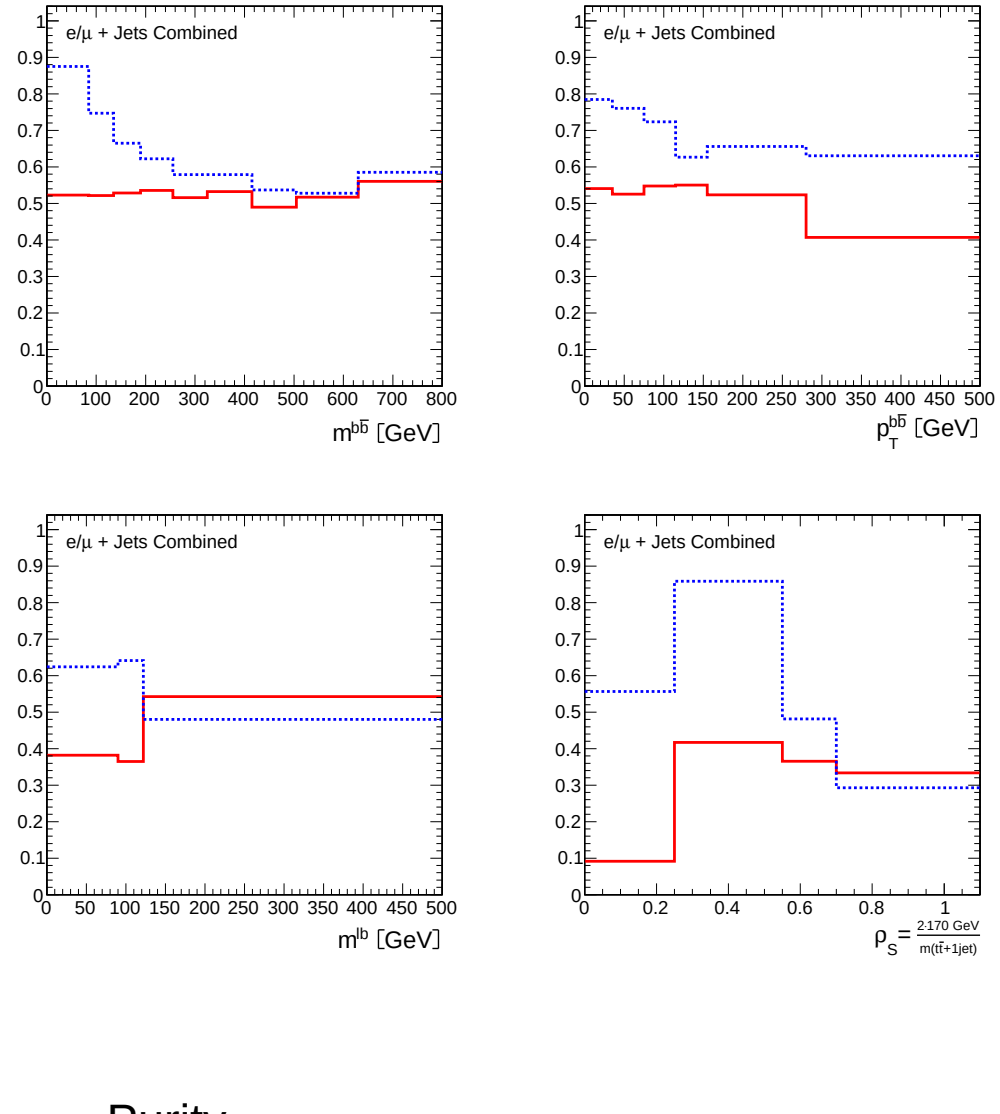

Figure 8.4: Purity (red) and stability (blue) for differential cross section quantities related to the $b\bar{b}$ system, the invariant mass of the lepton and the leptonic b jet and ρ_S in the visible particle level phase space to quantify migration effects for the chosen analysis binning in the combined ℓ+jets channel.

Chapter 9

Unfolding Procedure

As discussed in Chapter 8.4, migration effects lead to a correlation between the observed number of events in each bin. To obtain the underlying true distribution from the observed number of events in data ($\vec{N}_{\text{data}}$), these effects have to be corrected.

Analogously to the Equations 8.8 and 8.9 for the inclusive cross section, the expected contribution of background events from non-$t\bar{t}$ physics processes is subtracted. Furthermore, the contribution of $t\bar{t}$ events originating from non-signal final states is corrected for in a multiplicative approach using the $t\bar{t}$ signal fraction $f_{t\bar{t}\ \text{SG}}$ defined in equation 8.9. This background correction is done individually for each bin i:

$$(N_{\text{SG data}})_i = \left(\left(N^{\text{events,data}}_{\text{reco,selected}}\right)_i - \left(N^{\text{non}-t\bar{t}\ \text{events,sim}}_{\text{reco,selected}}\right)_i\right) \cdot f^i_{t\bar{t}\ \text{SG}}. \tag{9.1}$$

The event yield after background correction ($\vec{N}_{\text{SG data}}$) is used as input for the unfolding procedure which is finally correcting migration-, efficiency- and acceptance-effects.

The unfolding problem is explained in Chapter 9.1, followed by a description of the regularised unfolding method in Chapter 9.2 and an overview about the choice of the regularisation strength in Chapter 9.3. Finally, the unfolding setup and the chosen regularisation strength are intensively tested in Chapter 9.4 using pseudo data.

9.1 The Unfolding Problem

The background-corrected event yield ($\vec{N}_{\text{SG data}}$) is connected with the underlying true distribution ($\vec{N}_{\text{true}}$) via the response matrix $\mathbf{A^{resp}}$:

$$\vec{N}_{\text{SG data}} = \mathbf{A^{resp}}\ \vec{N}_{\text{true}}. \tag{9.2}$$

Every entry (i,j) of the response matrix $\mathbf{A^{resp}}$ is defined as the number of events which are generated in bin i and reconstructed in bin j divided by all events generated in bin j:

$$A^{\text{resp}}_{ij} = \frac{N^{\text{gen in bin j}}_{\text{reco in bin i}}}{N^{\text{gen in bin j}}}. \tag{9.3}$$

Events which are not part of the corresponding phase space definition, are filled in the underflow bin (j=0). $\mathbf{A^{resp}}$ contains a description of the migration effects through its two-dimensional form as well as the information about efficiencies and acceptances and is obtained from the simulation after applying all correction factors as detailed in Chapter 6.7. The product of efficiency and acceptance for a single bin j of the true spectrum is obtained from $\mathbf{A^{resp}}$ by summing over all reconstructed bins i:

$$(\epsilon \cdot \mathrm{Acc})_{\mathrm{j}} = \sum_i \mathrm{A}_{\mathrm{ij}}^{\mathrm{resp}} = \frac{\sum_i \mathrm{N}_{\text{reco in bin i}}^{\text{gen in bin j}}}{\mathrm{N}^{\text{gen in bin j}}} = \frac{\mathrm{N}_{\mathrm{reco}}^{\text{gen in bin j}}}{\mathrm{N}^{\text{gen in bin j}}}. \tag{9.4}$$

Solving Equation 9.2 for $\vec{\mathrm{N}}_{\mathrm{true}}$ is called the unfolding problem. The most obvious solution would be to obtain the true distribution using the inverted response matrix:

$$\vec{\mathrm{N}}_{\mathrm{true}} \;=\; (\mathbf{A^{resp}})^{-1}\; \vec{\mathrm{N}}_{\mathrm{SG\ data}}. \tag{9.5}$$

Unfortunately, the unfolding problem is in general a mathematically ill-posed problem and the result obtained from the direct inversion of the response matrix is generally unstable [181]. The obtained solution for $\vec{\mathrm{N}}_{\mathrm{true}}$ might be non-physically oscillating because a small change of $\vec{\mathrm{N}}_{\mathrm{SG\ data}}$ can have a large impact on the obtained $\vec{\mathrm{N}}_{\mathrm{true}}$ due to the correlations introduced by migrations. This is a general characteristic of the unfolding problem and might occur even if $\vec{\mathrm{N}}_{\mathrm{SG\ data}}$ is known accurately.

9.2 The Regularised Unfolding Method

As discussed previously, Equation 9.2 cannot be solved directly by inverting the response matrix. The large uncertainty values of low-statistic bins can propagate to other bins when migration effects are present and the full response matrix is directly inverted. This will lead in general to large statistical uncertainties of the obtained result, negative correlations between neighbouring bins and a non-physically oscillating result for $\vec{\mathrm{N}}_{\mathrm{true}}$.

Therefore, statistically insignificant contributions need to be suppressed when solving Equation 9.2. In order to achieve this, the solution needs to be regularised, e.g. by implying an a-priori constraint regarding the smoothness of the final result. This is achieved by a re-formulation of Equation 9.2 as χ^2-problem (χ^2_A) and the inclusion of an additional regularisation term $\left(\mathrm{Reg}\left(\vec{\mathrm{N}}_{\mathrm{true}}\right)\right)$:

$$\chi^2_{\mathrm{tot}} \;=\; \chi^2_A + \mathrm{Reg}\left(\vec{\mathrm{N}}_{\mathrm{true}}\right). \tag{9.6}$$

Consequently, the minimum of χ^2_{tot} equals an approximate solution $\vec{\mathrm{N}}_{\mathrm{true}}$ for the unfolding problem in Equation 9.2 which is additionally restricted by the regularisation condition.

The formulation of χ^2_A involves the covariance matrix (**COV**) for $\vec{\mathrm{N}}_{\mathrm{SG\ data}}$. In general, the entry (i,j) of the covariance matrix for a quantity x is defined as

$$\mathrm{cov}_{\mathrm{ij}} \;=\; \langle \left(\mathrm{x_i} - \langle \mathrm{x_i} \rangle\right)\left(\mathrm{x_j} - \langle \mathrm{x_j} \rangle\right) \rangle. \tag{9.7}$$

where $\langle x_i \rangle$ denotes the expectation value for x_i. The diagonal entries represent the squared statistical uncertainties, while off-diagonal entries indicate correlations between the bins. Here, $\mathbf{COV}_{\vec{N}_{\text{SG data}}}$ is simply constructed from the statistical uncertainty of the number of observed events in data for each bin i (ΔN^i_{Data}):

$$\left(\mathbf{COV}_{\vec{N}_{\text{SG data}}}\right)_{ij} = \delta_{ij} \cdot \Delta N^i_{\text{Data}} \cdot \Delta N^j_{\text{Data}}. \tag{9.8}$$

Furthermore, χ^2_{tot} is not directly solved for $\vec{N}_{\text{true}}$ but relative to the predicted distribution of the MC event generator ($\vec{N}^{\text{sim}}_{\text{gen}}$) by introducing w_j:

$$w_j = \frac{N_{\text{true,j}}}{N^{\text{sim}}_{\text{gen,j}}}. \tag{9.9}$$

Consequently, the response matrix is replaced by the event matrix $\mathbf{A}^{\textbf{event}}$, which contains directly the number of events:

$$A^{\text{event}}_{ij} = N^{\text{gen in bin j}}_{\text{reco in bin i}} = A^{\text{resp}}_{ij} \cdot N^{\text{gen in bin j}}. \tag{9.10}$$

Finally, χ^2_A is obtained as:

$$\chi^2_A = \left(\mathbf{A}^{\textbf{event}}\vec{w} - \vec{N}_{\text{SG data}}\right)^T \left(\mathbf{COV}_{\vec{N}_{\text{SG data}}}\right)^{-1} \left(\mathbf{A}^{\textbf{event}}\vec{w} - \vec{N}_{\text{SG data}}\right). \tag{9.11}$$

The second term in Equation 9.6 is the regularisation term $\left(\text{Reg}\left(\vec{N}_{\text{true}}\right)\right)$. It involves the regularisation parameter τ which is a measure of the regularisation strength and the curvature matrix $\mathbf{C}$ which defines the type of the smoothness condition:

$$\mathbf{C} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 1 & -2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & -2 & 1 \\ 0 & \cdots & 0 & 1 & -1 \end{pmatrix}. \tag{9.12}$$

Using this definition, $\mathbf{C}\vec{w}$ equals the discrete second derivative of $\vec{w}$ and the final regularisation term is constructed by multiplying its euclidian norm with the regularisation parameter:

$$\text{Reg}\left(\vec{N}_{\text{true}}\right) = \tau \left(\mathbf{C}\vec{w}\right)^T \left(\mathbf{C}\vec{w}\right). \tag{9.13}$$

After choosing the regularisation condition, the task of the unfolding procedure is to find a solution $\vec{w}_{\text{sol}}$ for the minimum of:

$$\chi^2_{\text{tot}} = \underbrace{\left(\mathbf{A}^{\textbf{event}}\vec{w} - (\vec{N}_{\text{SG data}})\right)^T \left(\mathbf{COV}_{\vec{N}_{\text{SG data}}}\right)^{-1} \left(\mathbf{A}^{\textbf{event}}\vec{w} - \vec{N}_{\text{SG data}}\right)}_{\chi^2_A} + \underbrace{\tau \left(\mathbf{C}\vec{w}\right)^T \left(\mathbf{C}\vec{w}\right)}_{\text{Reg}\left(\vec{N}_{\text{true}}\right)}. \tag{9.14}$$

For this thesis, the choice of $\mathbf{C}$ and the solving of χ^2_{tot} is done following the methods of the Singular Value Decomposition (SVD) as detailed in [182] and implemented in the software package TSVDUnfold [183]. The unfolded spectrum ($\vec{N}_{\text{true}}$) is then obtained from the solution $\vec{w}_{\text{sol}}$ using Equation 9.9 and the final differential cross section $\frac{d\sigma}{dx}\left(\vec{N}^{t\bar{t}\to e/\mu \text{ events,data}}_{\text{unfolded}} = \vec{N}_{\text{true}}\right)$ is derived using Equation 8.12.

Moreover, side bins[a] are ignored at reconstruction level but unfolded at generator level. This means, that events in the underflow (UF) and overflow (OF) bins are ignored for data ($N^{\text{data}}_{\text{reco in bin OF}\bigvee\text{UF}} \equiv 0$) and for the MC reconstruction level ($N^{\text{gen in bin i}}_{\text{reco in bin OF}\bigvee\text{UF}} \equiv 0 \;\forall$ bins i) of the response matrix while they are considered for the MC generator truth level ($N^{\text{gen in bin OF}\bigvee\text{UF}}_{\text{reco in bin j}} \neq 0 \;\forall$ bins j $\neq$ (OF$\bigvee$UF)) of the response matrix. Consequently, migration effects from the side bins into the measured bins are corrected. In contrast, the measured content of the side bins in data is not used for the calculation of normalised differential cross sections.
The contribution of events in reconstruction level side bins is anyhow small as the kinematic region of the measurement is chosen such that the majority of the events is contained. Therefore, the statistical uncertainty of UF and OF bins at reconstruction level is large and including them into the unfolding procedure would not improve the obtained result.
In contrast, the side-bin contribution at generator level is relevant for the cross section quantities measured in the visible particle level phase space quantities if events are reconstructed but not part of the defined phase space (see Chapter 8.1.2).

Furthermore, the normalisation of the unfolded differential cross section results as described in Equation 8.13 is done for the visible range of the measurement. This means, that the bin content of UF and OF bins is explicitly set to 0 and therefore, the integral over the measured bins of the final normalised differential cross section results is exactly one:

$$\sum_{\substack{\text{measured}\\ \text{bins i=1}}}^{N_{\text{bins}}} \left(\frac{1}{\sigma}\frac{d\sigma}{dx}\right)_{\text{bin i}} \cdot \Delta x_{\text{bin i}} = 1.0. \tag{9.15}$$

For the measurements in visible particle level phase space the unfolded result contains entries in the UF bin due to events migrating from outside the phase space into the sample of reconstructed events. As indicated in Equation 9.15, this contribution is consistently neglected for the final result.

[a] Here, side bins denote the underflow and overflow bins, which contain events that are reconstructed or generated outside the kinematic region of the measurement.

9.3 Choice of the Regularisation Strength

One important aspect of the unfolding procedure is the choice of the continuous regularisation parameter τ in Equation 9.14. The value of τ is a measure of the regularisation strength as it changes the relative size of the regularisation condition $\text{Reg}\left(\vec{\text{N}}_{\text{true}}\right)$ with respect to the χ^2_A-expression, which is related to the unfolding problem itself.

9.3.1 The Minimum-Global-Correlation Method

The choice of the optimal regularisation parameter (τ) is based on the minimum-global-correlation method proposed in [181] and implemented in [94].

Following this method, τ is individually determined for each cross section quantity by scanning the root-mean square (RMS) global correlation coefficient ($\bar{\rho}$) for all values of τ and choosing the minimum. The value of $\bar{\rho}$ is basically a measure of the average correlation of each bin with all other bins. The idea behind this method is that a regularisation that is too weak or too strong will lead to a large value of $\bar{\rho}$.

A correlation that is too weak removes basically the regularisation term so that $\chi^2_{\text{tot}} \approx \chi^2_A$ and the obtained solution has the same problems as discussed for the direct matrix inversion in Chapter 9.1:

- unstable, oscillating behaviour
- large statistical uncertainties
- large negative correlations between neighbouring bins

} regularisation is too weak

In contrast, a regularisation that is too strong can smooth not only statistical fluctuations but affect the structure of the distribution itself. In the most extreme case, the obtained solution $\vec{\text{N}}_{\text{true}}$ is forced to be very similar to the (simulated) spectrum, which was used in the regularisation condition $\text{Reg}\left(\vec{\text{N}}_{\text{true}}\right)$. In summary, a regularisation that is too strong leads to:

- bias towards the prediction used for the regularisation
- large positive correlations (also between far distant bins)

} regularisation is too strong

Both, the large negative correlations of the too weak and the large positive correlations of the too strong regularisation will result in a large value for $\bar{\rho}$. Therefore, the minimum of $\bar{\rho}(\tau)$ is chosen as optimal τ parameter for the analysis. This minimum depends on the measured distribution in data as well as on the chosen binning.

$\bar{\rho}$ is defined as the RMS of the global correlations of all bins:

$$\bar{\rho} = \sqrt{\frac{1}{\text{N}_{\text{bins}}} \sum_{\text{i}=1}^{\text{N}_{\text{bins}}} \rho_{\text{i}}^2}. \tag{9.16}$$

Here, the global correlation of bin i (ρ_i) is the maximum correlation of this bin with every possible linear combination of all other bins. The concept of ρ_i is detailed in [184–186]. ρ_i is obtained using the statistical covariance matrix (**COV**) of the final result:

$$\rho_\mathrm{i} = \sqrt{1 - \frac{1}{(\mathbf{COV})_{\mathrm{ii}} \cdot (\mathbf{COV}^{-1})_{\mathrm{ii}}}}. \tag{9.17}$$

The statistical covariance matrix of the final result is obtained via error propagation using pseudo-experiments. For each pseudo experiment, the measured event yield is overlaid with noise according to the (poissonean) statistical error ($\sqrt{\mathrm{N}_\mathrm{Data}}$) in each bin and the unfolding procedure is applied to the smeared data distribution. Finally, the covariance matrix is extracted from 1000 pseudo experiments for each distribution using Equation 9.7.

9.3.2 τ-Scan for All Cross Section Quantities

The RMS global correlation ($\bar{\rho}$) as a function of the regularisation parameter (τ) is shown in Figure 9.1 to 9.2 for the cross section quantities related to top quarks and in Figure 9.3 to 9.4 for all other cross section quantities. Typically, $\bar{\rho}$ is at a moderate level for small values of τ (weak regularisation), has a minimum at the optimal value which is finally chosen for the analysis and rises steeply for large values of τ (strong regularisation). The size of $\bar{\rho}(\tau_{min}) - \bar{\rho}(0)$ is a measure for the amount of migration. For quantities with low migration effects like the transverse momentum or the pseudorapidity of the lepton, the value of $\bar{\rho}(\tau \approx 0)$ is already small and the shallow minimum indicates that only a very weak regularisation is needed. For top-quark related quantities, the minimum is in general deeper as migration effects are more pronounced.

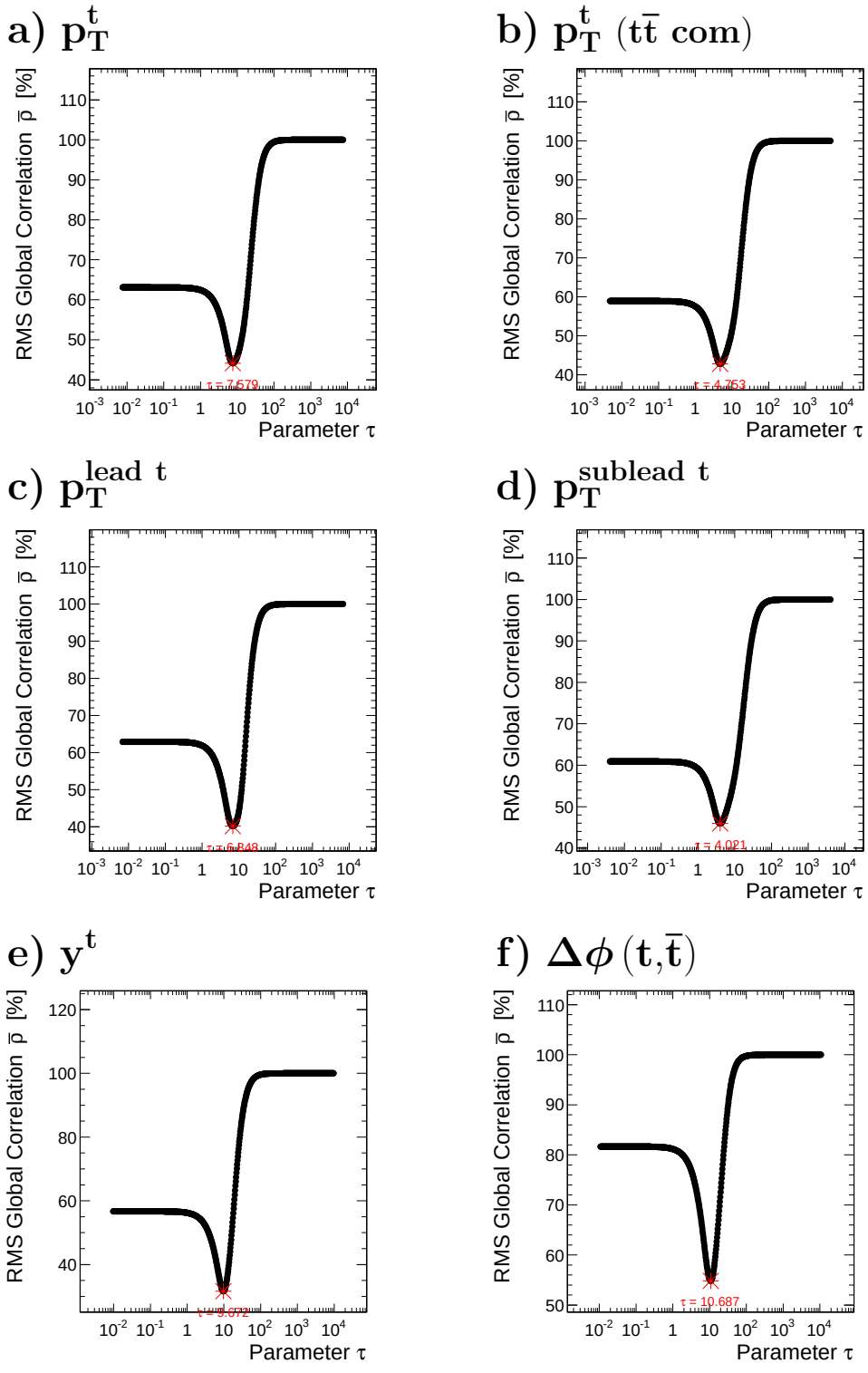

Figure 9.1: Scan of the RMS global correlation of the unfolded result as function of the regularisation parameter τ in the combined ℓ+jets channel for the top-quark quantities. The marked minimum is finally used for the measurement.

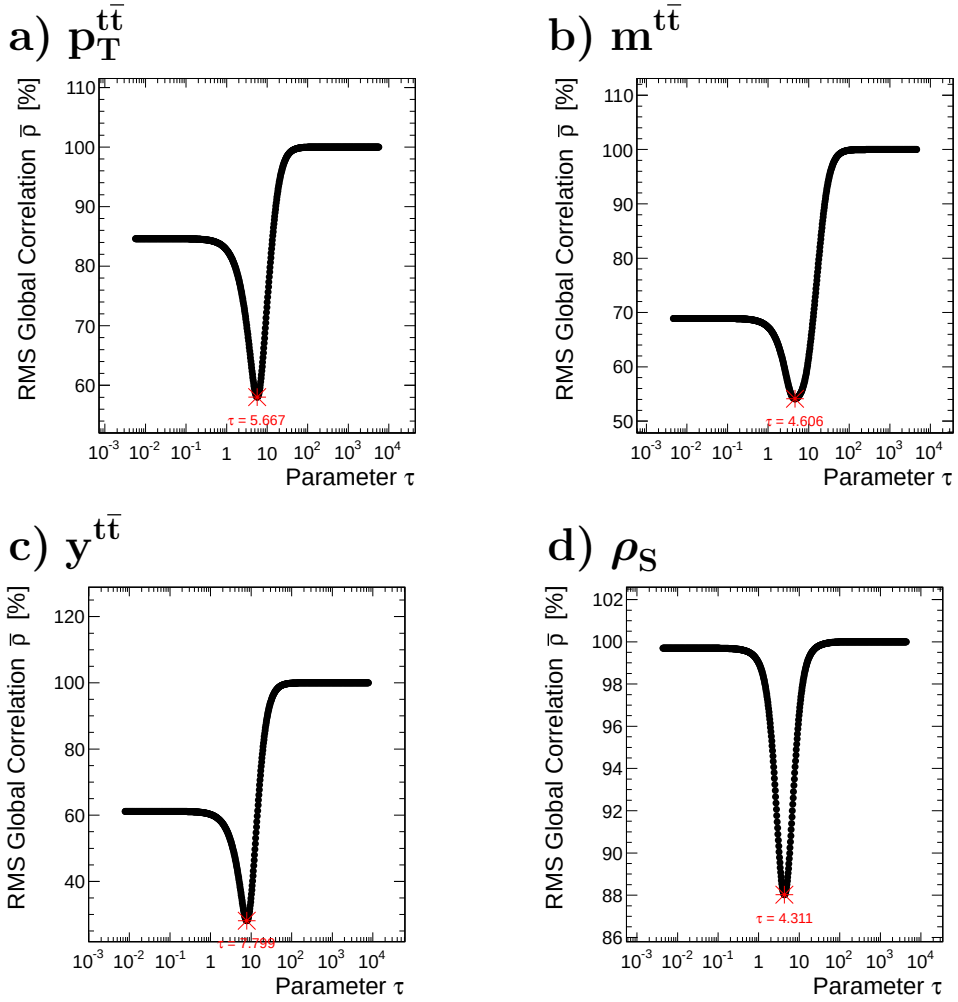

Figure 9.2: Scan of the RMS global correlation of the unfolded result as function of the regularisation parameter τ in the combined ℓ+jets channel for the quantities related to the $t\bar{t}$ system. The marked minimum is finally used for the measurement.

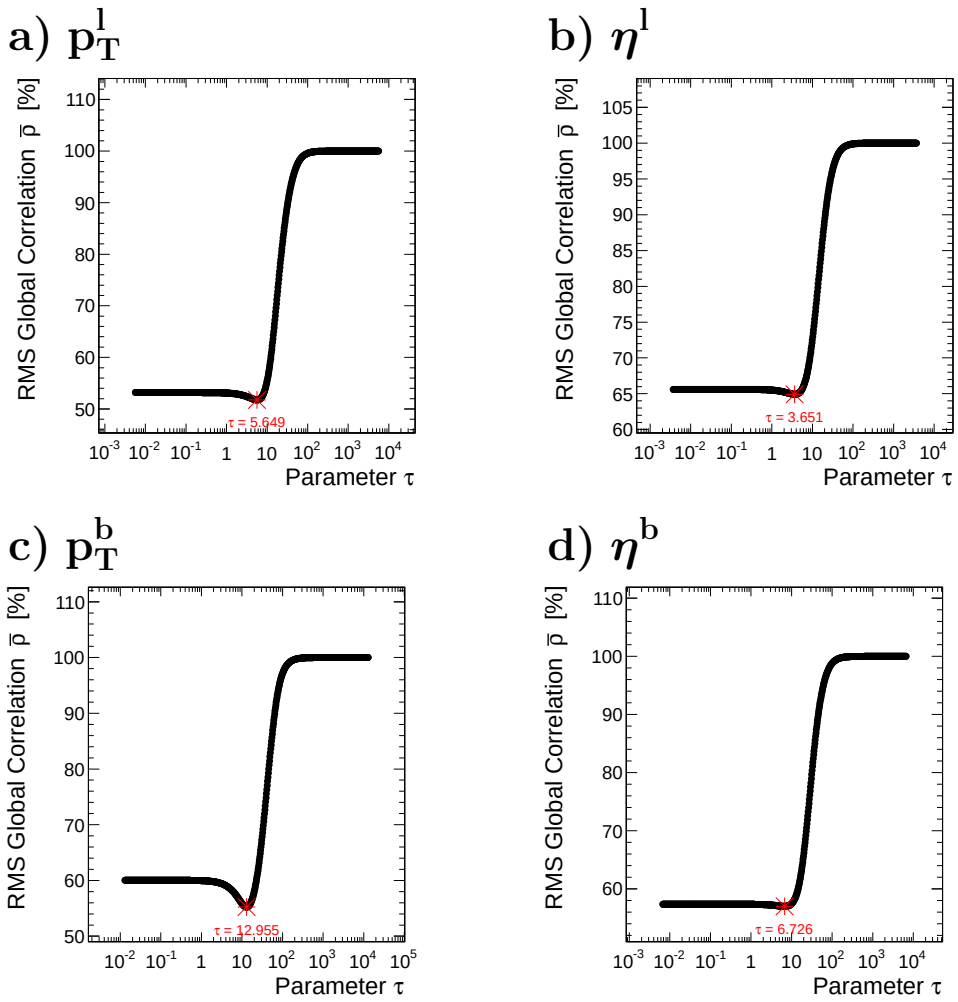

Figure 9.3: Scan of the RMS global correlation of the unfolded result as function of the regularisation parameter τ in the combined ℓ+jets channel for the lepton (a,b) and b-jet (c,d) quantities. The marked minimum is finally used for the measurement.

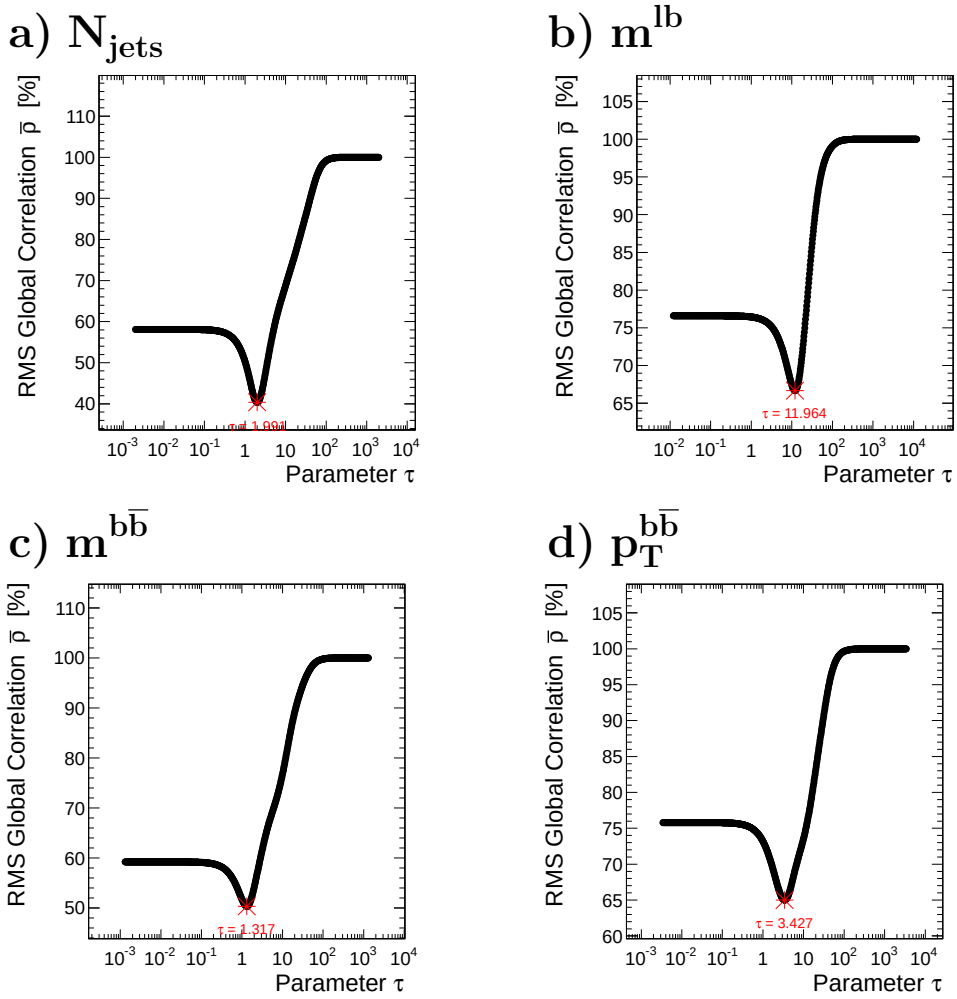

Figure 9.4: Scan of the RMS global correlation of the unfolded result as function of the regularisation parameter τ in the combined ℓ+jets channel for the jet multiplicity (a), the invariant mass of the lepton and the leptonic b jet associated to the top-quark decay (b) and the quantities of the $b\bar{b}$ system associated to the top-quark decay (c,d). The marked minimum is finally used for the measurement.

9.4 Testing the Unfolding Setup

The unfolding procedure is one of the key points of the analysis as it delivers the final differential cross sections from the reconstruction level event yields. For the regularised unfolding method explained in the previous chapters, efficiencies[b], acceptances, migrations and the input for the regularisation condition are taken from the MadGraph+pythia $t\bar{t}$ prediction.

Within the unfolding procedure itself, some knowledge about the approximate shape of the result is assumed, e.g. the regularisation condition involves some preliminary knowledge about the curvature of the obtained result. Therefore, every unfolding procedure depends to some extend on the simulation used to obtain the response matrix and the regularisation condition.

Consequently, it is important to test the unfolding procedure in order to verify the unfolding setup and the chosen regularisation strength. For this purpose, closure tests are performed to proof that the true input shape is reproduced by unfolding the reconstruction level event yield. This was done by performing the complete analysis with different sets of pseudo data which are created from MC predictions. For all closure tests, the response matrix of the central MadGraph+pythia MC prediction, which is also used to perform the data analysis, is used for the unfolding. As for the real data, the regularisation parameter is optimised for every pseudo data set by minimising the RMS global correlation. In order to use an independent dataset for the $t\bar{t}$ component, the MC predictions without MadSpin as listed in Chapter 5.3.2 are used to create the pseudo data. It was checked that the obtained shapes of the cross section quantities agree between the simulations with and without MadSpin. The only difference observed involves the quantities $p_T^{b\bar{b}}$, $m^{b\bar{b}}$ and m^{lb} which are sensitive to spin correlation effects, which are only modelled in the simulations including MadSpin.

Several pseudo data sets are created to test the unfolding procedure, involving different "true" input shapes for the cross section quantities of the $t\bar{t}$ component. Every pseudo data set ($\vec{N}_{\text{pseudo data}}$) is created from the default MC predictions ($\vec{N}^{\text{sim}}$) by adding the expected contribution of each process. The absolute normalisation for the data luminosity ($\mathscr{L}_{\text{data}}$) is obtained from the inclusive cross sections (σ^{theory}) listed in Chapter 5.3.2:

$$\vec{N}_{\text{pseudo data}} = \sum_{\substack{\text{non-}t\bar{t}\text{ default}\\ \text{MC prediction i}}} \vec{N}_i^{\text{sim}} \cdot \frac{\sigma_i^{\text{theory}} \cdot \mathscr{L}_{\text{data}}}{N_i^{\text{sim,gen}}} + \vec{N}_{t\bar{t},\text{modified}}^{\text{sim}} \cdot \frac{\sigma_{t\bar{t}}^{\text{theory}} \cdot \mathscr{L}_{\text{data}}}{N_{t\bar{t},\text{modified}}^{\text{sim,gen}}}. \quad (9.18)$$

Here, $N_i^{\text{sim,gen}}$ is the total number of events in the MC prediction i and $N_{t\bar{t},\text{modified}}^{\text{sim,gen}}$ is the total number of events in the modified $t\bar{t}$ prediction.

Different pseudo data sets are created by varying the $t\bar{t}$ component– either by using event weights as function of a specific kinematic quantity or by adding an additional

[b] All correction factors for the efficiencies are applied (see Chapter 6.7).

kind of signal. Consequently, the underlying true shape of the pseudo data and the assumed shape, which is used to perform the unfolding, differ. This reflects the situation in real data where the true distribution might be different than the MC prediction.

Finally, different kinds of tests are performed, involving pseudo data and real data to validate the unfolding setup:

- Consistency tests with the default MC prediction
  $\longrightarrow$ to verify that the unfolding setup is working
- Pseudo data closure tests with a different "true" shape of the $t\bar{t}$ component
  - obtained by modifying p_T^t shape
  - obtained by modifying $m^{t\bar{t}}$ shape

  $\longrightarrow$ to proof that there is no bias towards the MC prediction used within the unfolding procedure
- Pseudo data closure test with an additional BSM vector boson signal
  - obtained by adding a Z'$\longrightarrow t\bar{t}$ signal
    $(m_{Z'} = 1\,\text{TeV}, \Gamma_{Z'} = 100\,\text{GeV}, \sigma(Z') \cdot \text{BR}(Z' \to t\bar{t}) = 5\text{pb})$

  $\longrightarrow$ to test the sensitivity regarding a hypothetical heavy resonance
- Consistency tests by applying the regularised unfolding method iteratively
- Stability tests with different choices of the unfolding setup for pseudo data and real data
  - comparison to a SVD unfolding method with a discrete regularisation
  - test different choices of the regularisation parameter τ

  $\longrightarrow$ to verify the chosen setup and parameters

All closure tests are performed by comparing the true normalised differential cross sections of the pseudo data with the result obtained from unfolding the reconstruction level event yields of the pseudo data with the response matrix of the default MC prediction. As the true input shapes of the pseudo data are different in comparison to the default $t\bar{t}$ MC prediction, these closure tests are a direct measure of the sensitivity of the unfolding procedure regarding the (unavoidable) input from the simulation.

9.4.1 Consistency Tests

The first and simplest consistency check is to unfold pseudo data, which is created from the plain sum of all MC predictions without any smearing, shape distortions or additional signals. The reconstructed and unfolded differential normalised cross sections are compared to the original predictions and found to be identical, proving that the analysis setup is working as expected.

Additionally, it was checked that unfolding the data and applying the response matrix to the result gives back the original data distribution. This ensures, that no statistically relevant contributions are suppressed by the regularisation.

9.4.2 Closure Tests with Different Signal Shapes

Furthermore, pseudo data sets with a modified $t\bar{t}$ component are created. The applied distortions for all $t\bar{t}$ events result in a softer or harder spectrum for $m^{t\bar{t}}$ or p_T^t. While a simple linear function is used to modify the $m^{t\bar{t}}$ spectrum, an exponential approach is chosen for p_T^t. This choice for p_T^t was motivated by the preliminary results of this analysis [26]. The "down" variation of $w(x=p_T^t)$ corresponds to a twice[c] as soft spectrum as observed in [26] while the "up" variation results in a spectrum, which is harder by the same amount as the "down" variation is softer than the initial MC prediction. Details for the modification functions $w(x=p_T^t, m^{t\bar{t}})$ are given in Table 9.1.

Quantity x	Variation	Modification Function w(x)	Closure Test Results
p_T^t	down	$-1+2\cdot\exp(0.318-0.00282\cdot x)$	Figure 9.5 to 9.8
	up	$\exp(0.318-0.00282\cdot x)$	Figure 9.9 to 9.12
$m^{t\bar{t}}$	down	$1-(x-350)\cdot 0.0008$	Figure 9.13 to 9.16
	up	$1+(x-350)\cdot 0.0008$	Figure 9.17 to 9.20

Table 9.1: Details on the applied modification functions for the $t\bar{t}$ component which are used to create pseudo data sets for the unfolding closure tests. The last column refers to the Figure with the obtained results

To avoid unnatural distortions due to extreme scaling of single events, the weight w(x) is restricted to the following range:

$$0.1 \leq w(x) \leq 2.0. \tag{9.19}$$

For the closure test with a modified p_T^t spectrum, one has to consider that two top quarks are present in each event:

$$\begin{aligned} w'(x = p_T^t) \quad &= \quad \sqrt{w\left(p_T^t\right) \cdot w\left(p_T^{\bar{t}}\right)} \\ \left(\text{for } x = \ m^{t\bar{t}}: \ \ w'(x) = w(x)\right). & \end{aligned} \tag{9.20}$$

While the modification procedure changes the shape of all cross section quantities, the absolute normalisation is kept constant by applying the additional normalisation factor:

$$w_{\text{Norm}}(x) = \left(\sum_{(\text{all events i})} (w'(x_i)) \right)^{-1}. \tag{9.21}$$

Finally, the modification function is converted into an event weight wgt(x):

$$\text{wgt}(x) = w'(x) \cdot w_{\text{Norm}}(x). \tag{9.22}$$

[c] Here, twice as soft means that the slope of the ratio between the created pseudo data and the prediction used for the unfolding approach is twice as large as observed in real data.

This event weight is applied analogously to the event weights for the data driven corrections detailed in Chapter 6.7 to create the modified $t\bar{t}$ component of the pseudo data.

The corresponding results for the true and unfolded distributions of the pseudo data are shown in Figure 9.5 to 9.12 for the modification of the p_T^t spectrum and in Figure 9.13 to 9.20 for the modification of the $m^{t\bar{t}}$ spectrum. These kind of closure tests show the behaviour of an overall shape distortion of the true $t\bar{t}$ spectrum.

Good agreement is found even for large shape variations which would show up already in the event yields and are not observed in real data (see Chapter 7.4). Only for ρ_S, the closure test shows some differences for extreme scenarios. For a realistically[d] softer p_T^t, the closure test for ρ_S shows agreement within the uncertainties of the final measurement, which are much larger than the observed difference between true and unfolded pseudo data. Consequently, the regularised unfolding depends only minimally on the preliminary knowledge of the true input shape and no bias is introduced when the underlying shape in data is moderately different than in the MC prediction used for the unfolding procedure and cross section measurement.

[d] Here, realistically means that the slope of the ratio between the created pseudo data and the prediction used for the unfolding approach is as large as observed in real data.

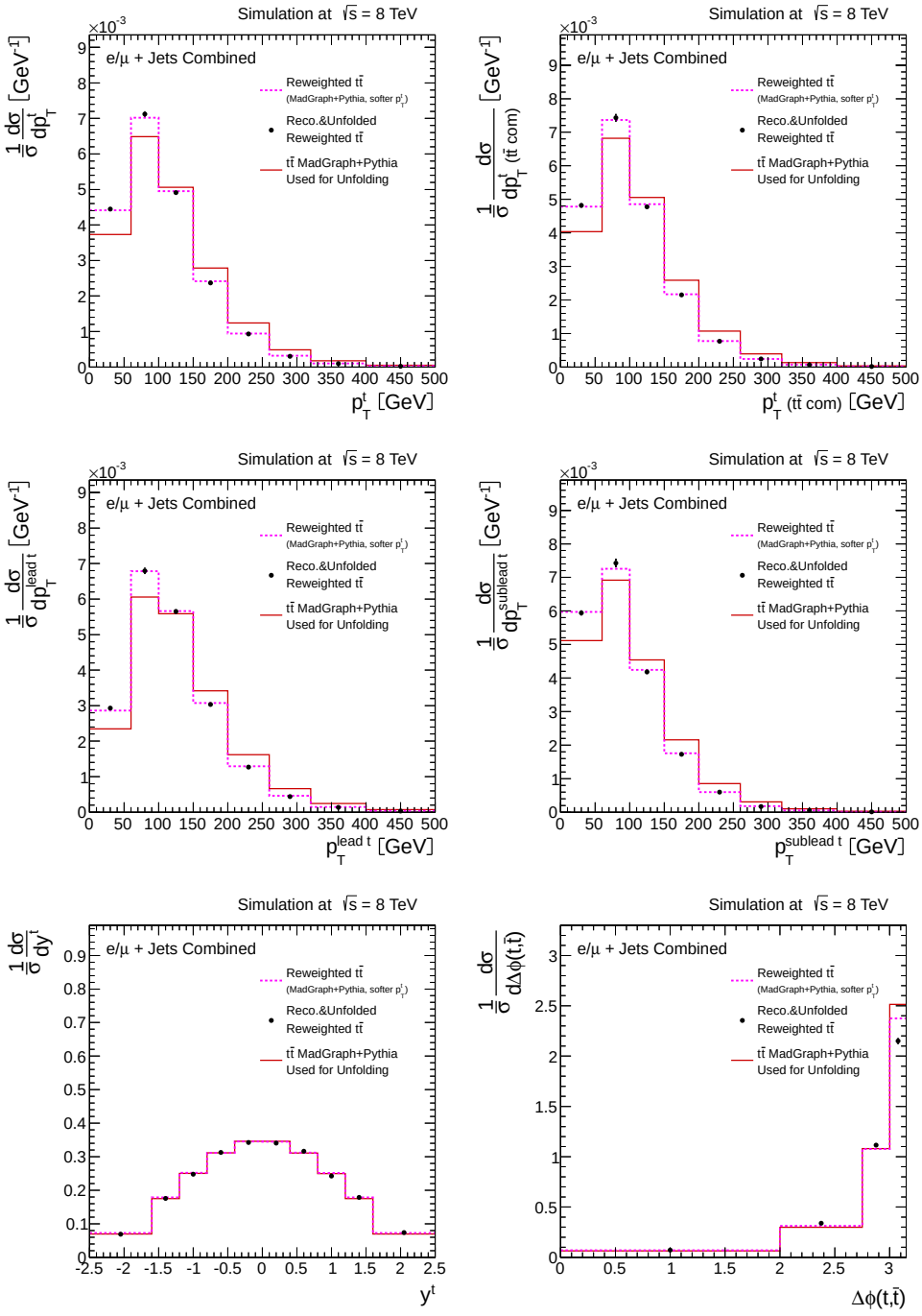

Figure 9.5: Results of the unfolding closure test with pseudo data featuring a softer p_T^t spectrum for the cross section quantities related to the top quark in the fully extrapolated parton level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

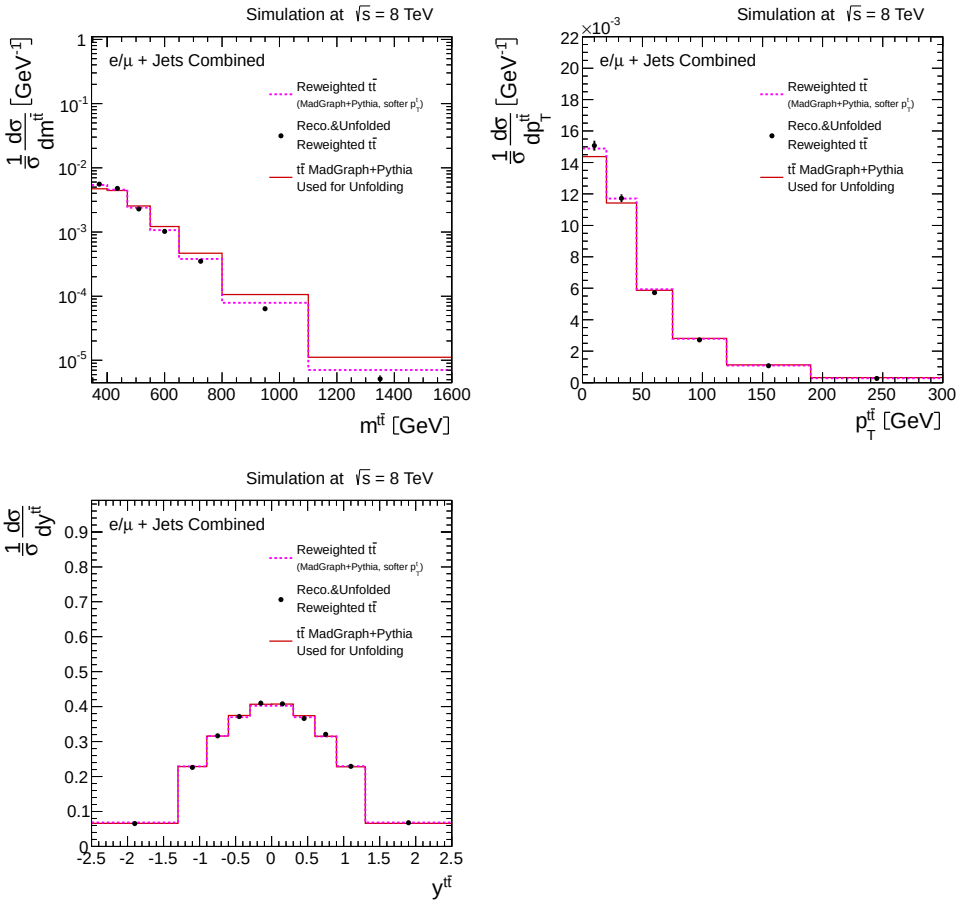

Figure 9.6: Results of the unfolding closure test with pseudo data featuring a softer p_T^t spectrum for the cross section quantities related to the $t\bar{t}$ system in the fully extrapolated parton level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

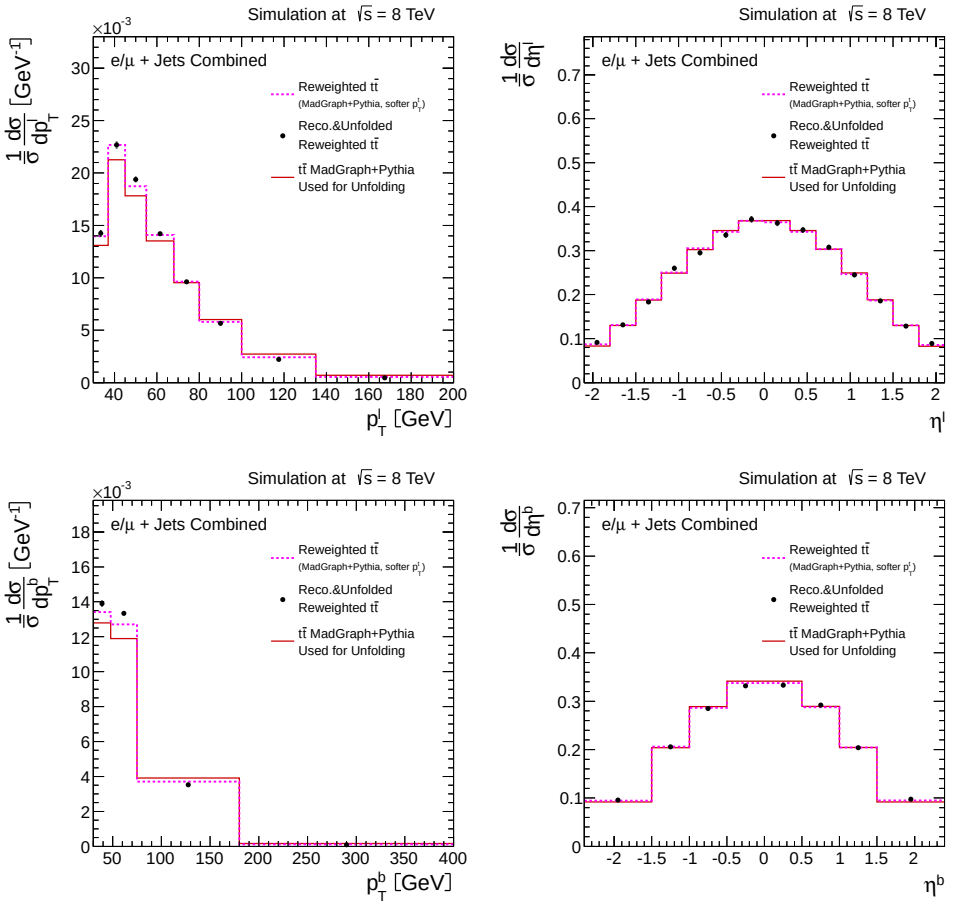

Figure 9.7: Results of the unfolding closure test with pseudo data featuring a softer p_T^t spectrum for the lepton and b-jet cross section quantities in the visible particle level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

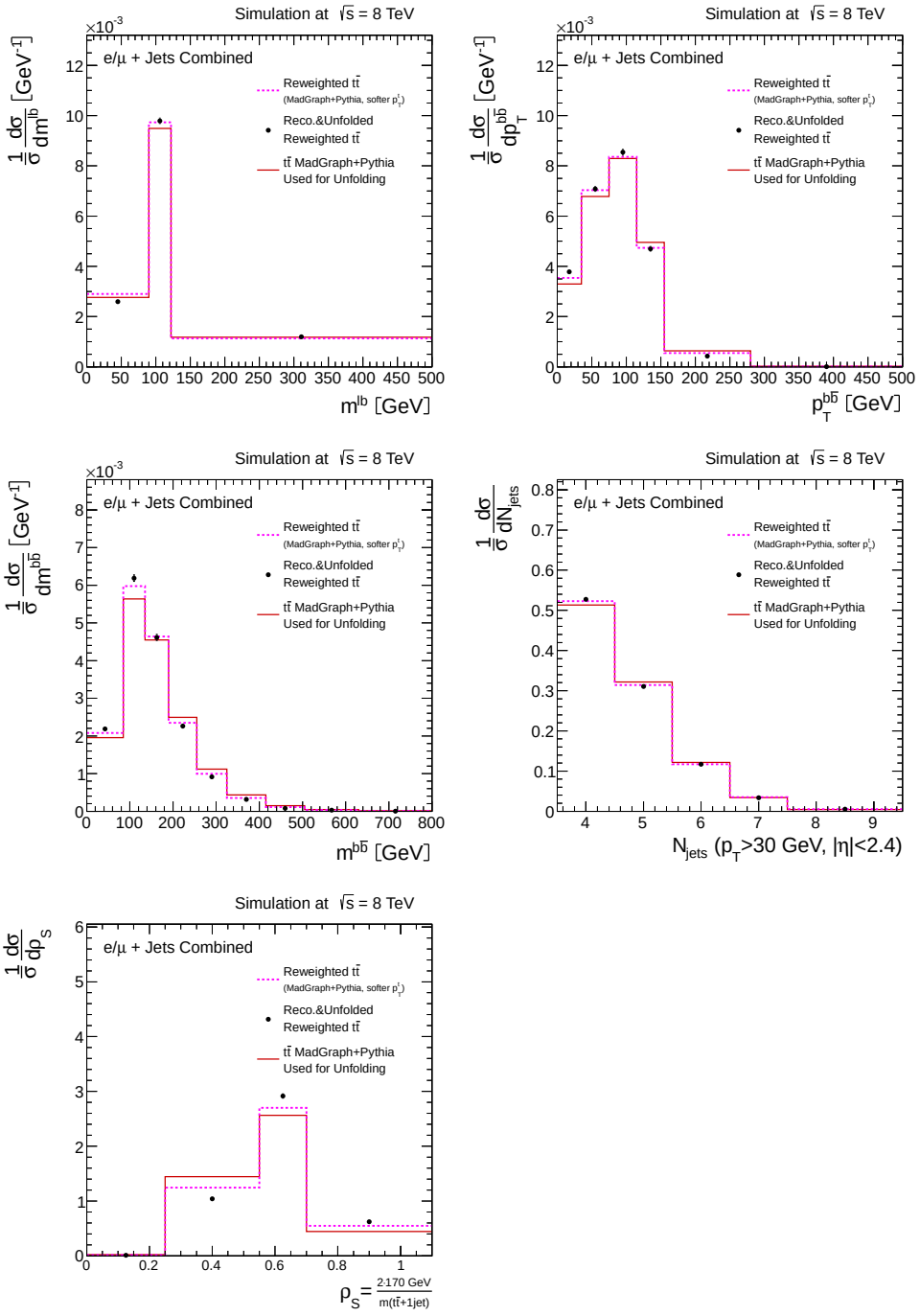

Figure 9.8: Results of the unfolding closure test with pseudo data featuring a softer p_T^t spectrum for the jet multiplicity, the $b\bar{b}$ system cross section quantities, the invariant mass of the lepton and the leptonic b bjet and ρ_S in the visible particle level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

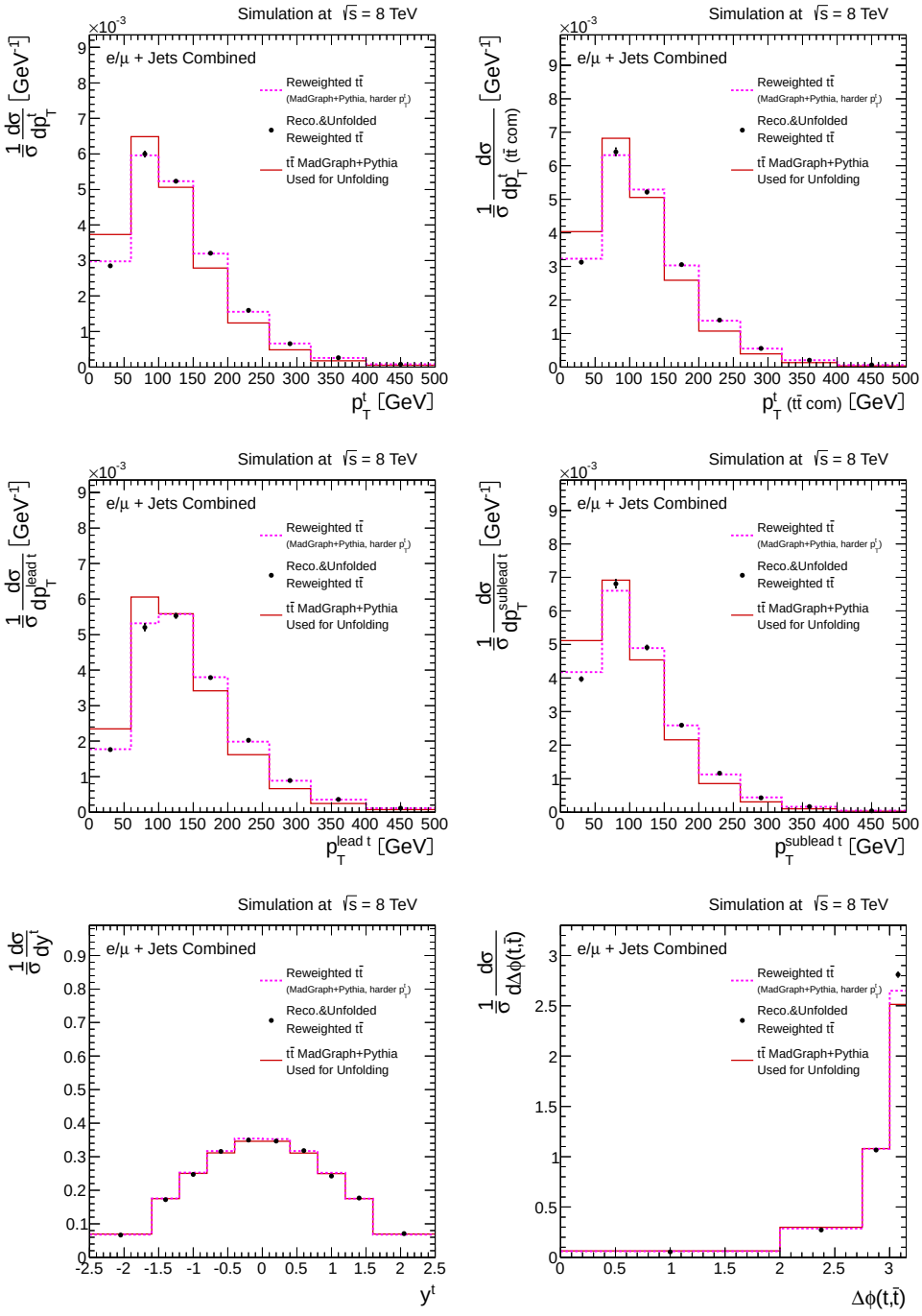

Figure 9.9: Results of the unfolding closure test with pseudo data featuring a harder p_T^t spectrum for the cross section quantities related to the top quark in the fully extrapolated parton level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

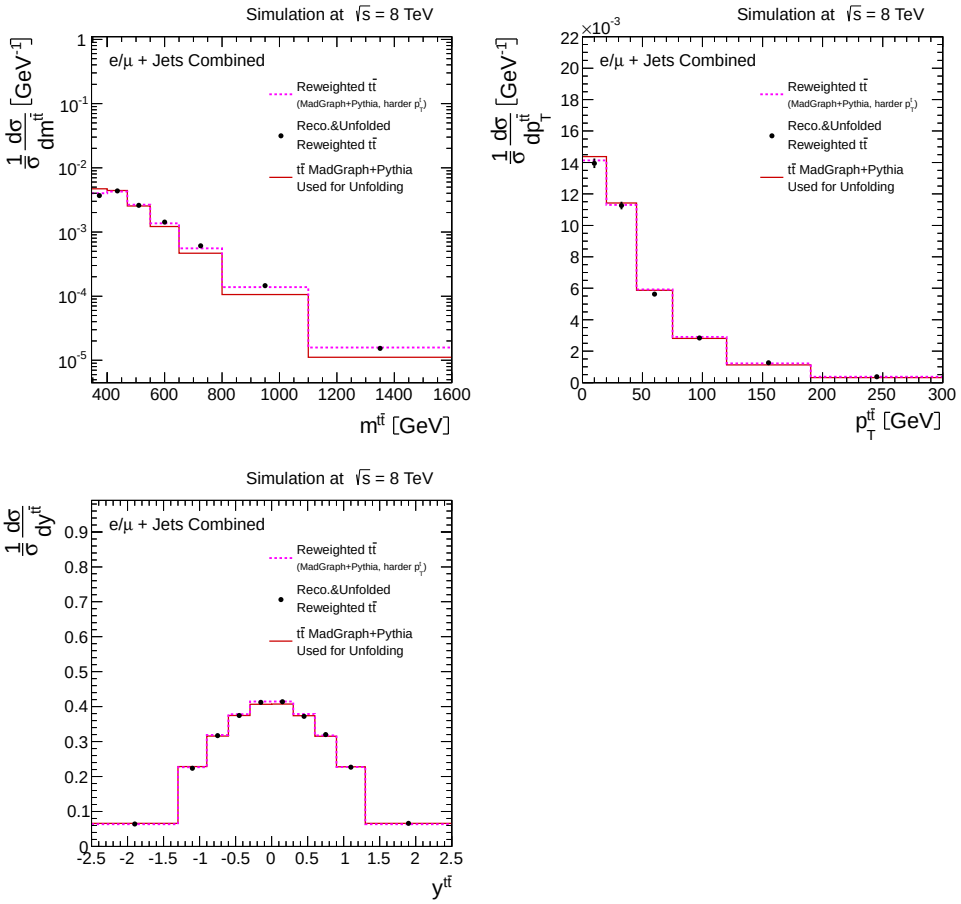

Figure 9.10: Results of the unfolding closure test with pseudo data featuring a harder p_T^t spectrum for the cross section quantities related to the $t\bar{t}$ system in the fully extrapolated parton level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

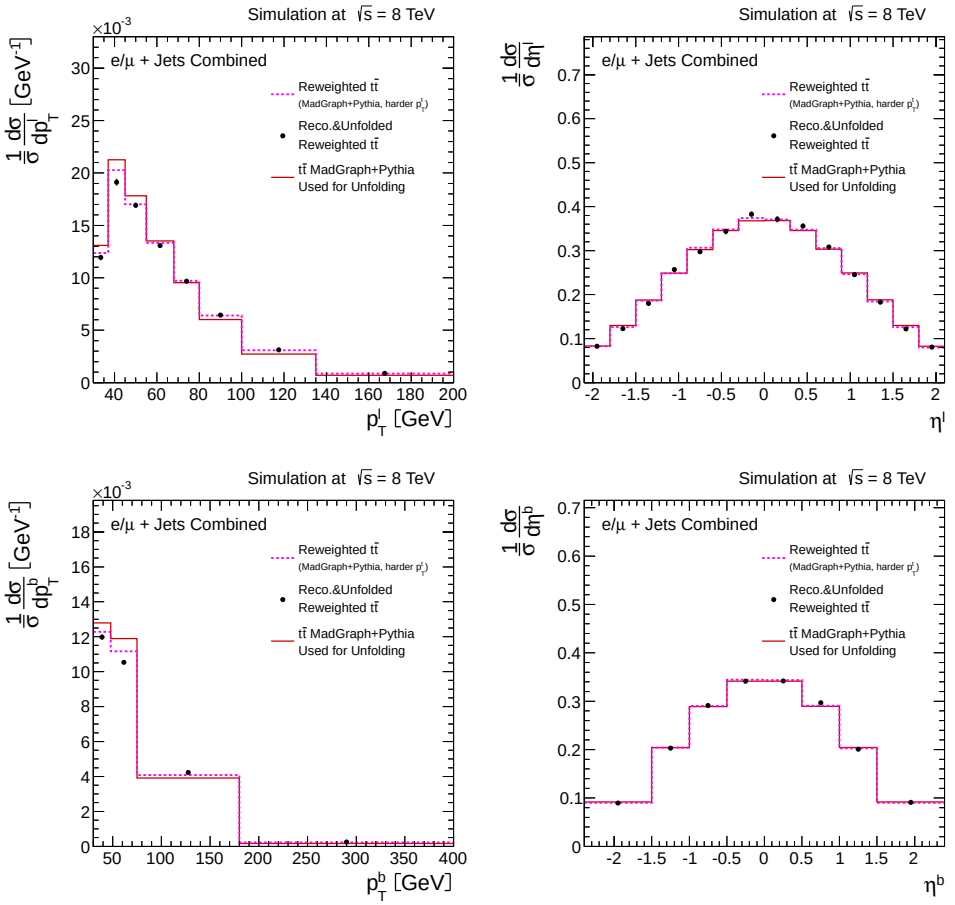

Figure 9.11: Results of the unfolding closure test with pseudo data featuring a harder p_T^t spectrum for the lepton and b-jet cross section quantities in the visible particle level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

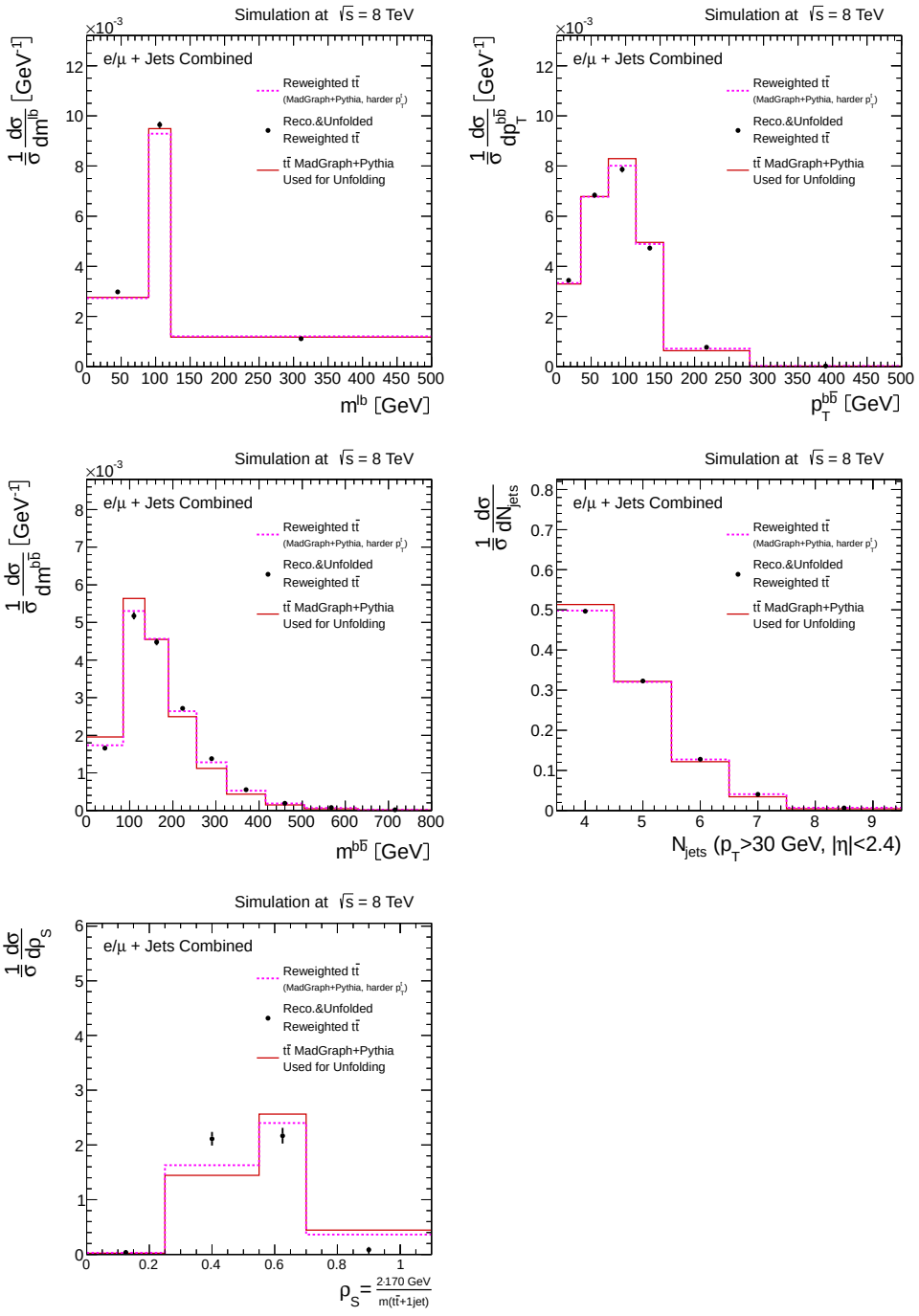

Figure 9.12: Results of the unfolding closure test with pseudo data featuring a harder p_T^t spectrum for the jet multiplicity, the $b\bar{b}$ system cross section quantities, the invariant mass of the lepton and the leptonic b bjet and ρ_S in the visible particle level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

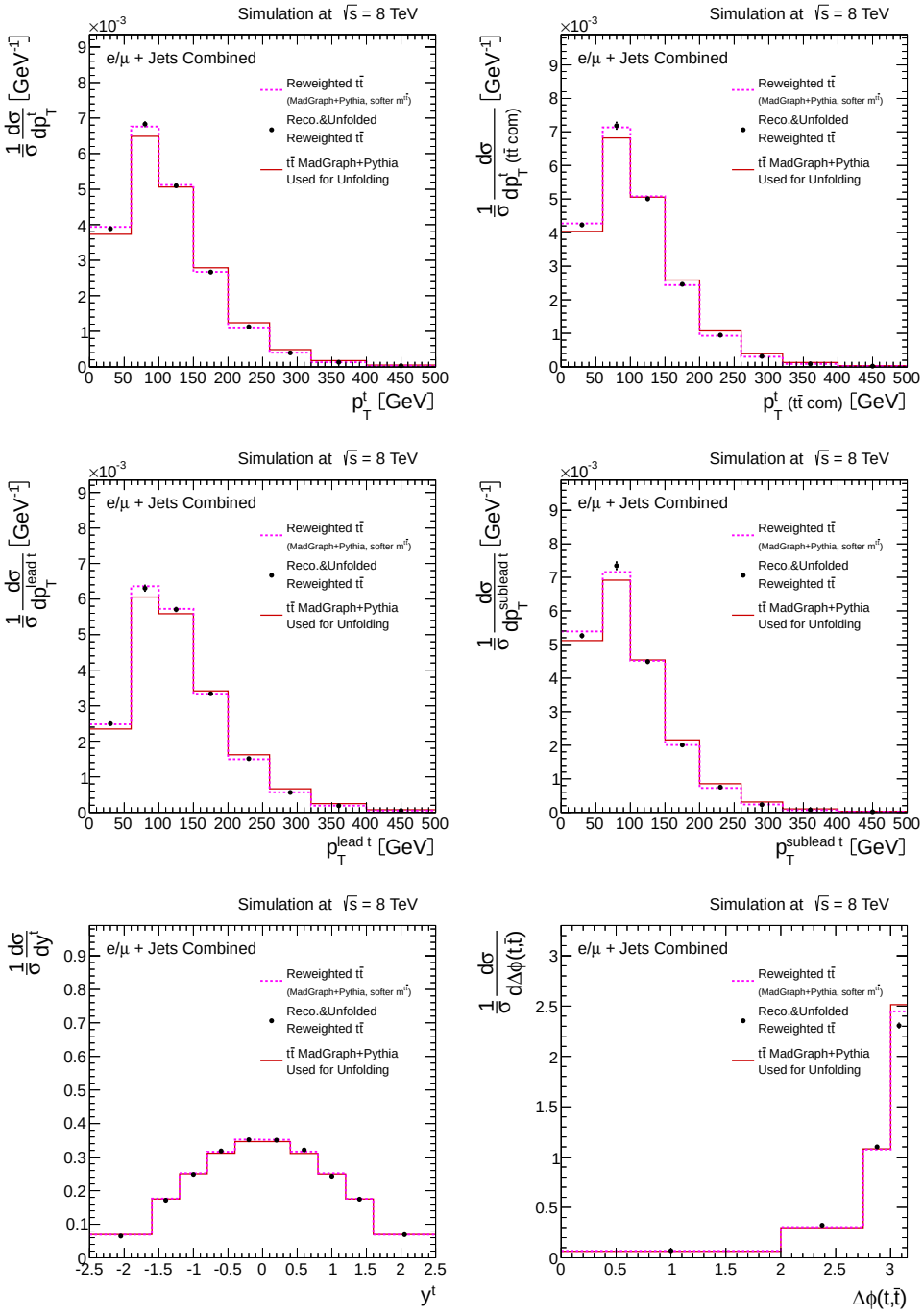

Figure 9.13: Results of the unfolding closure test with pseudo data featuring a softer $m^{t\bar{t}}$ spectrum for the cross section quantities related to the top quark in the fully extrapolated parton level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

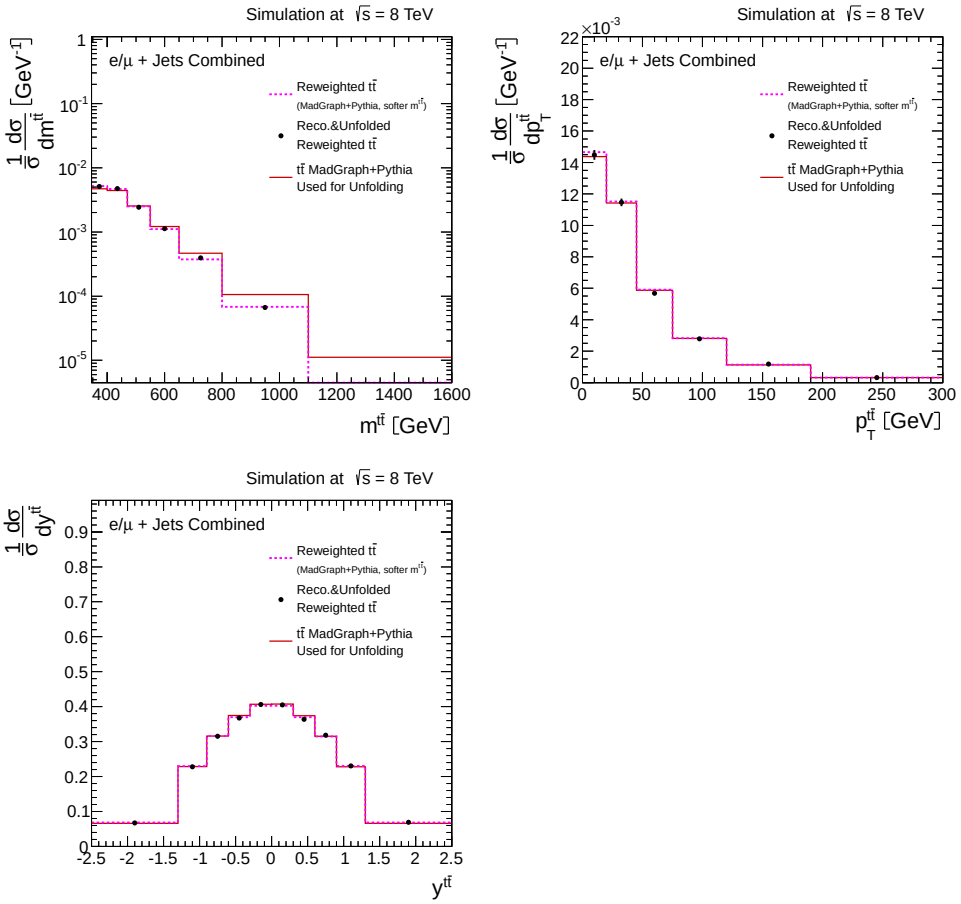

Figure 9.14: Results of the unfolding closure test with pseudo data featuring a softer $m^{t\bar{t}}$ spectrum for the cross section quantities related to the $t\bar{t}$ system in the fully extrapolated parton level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

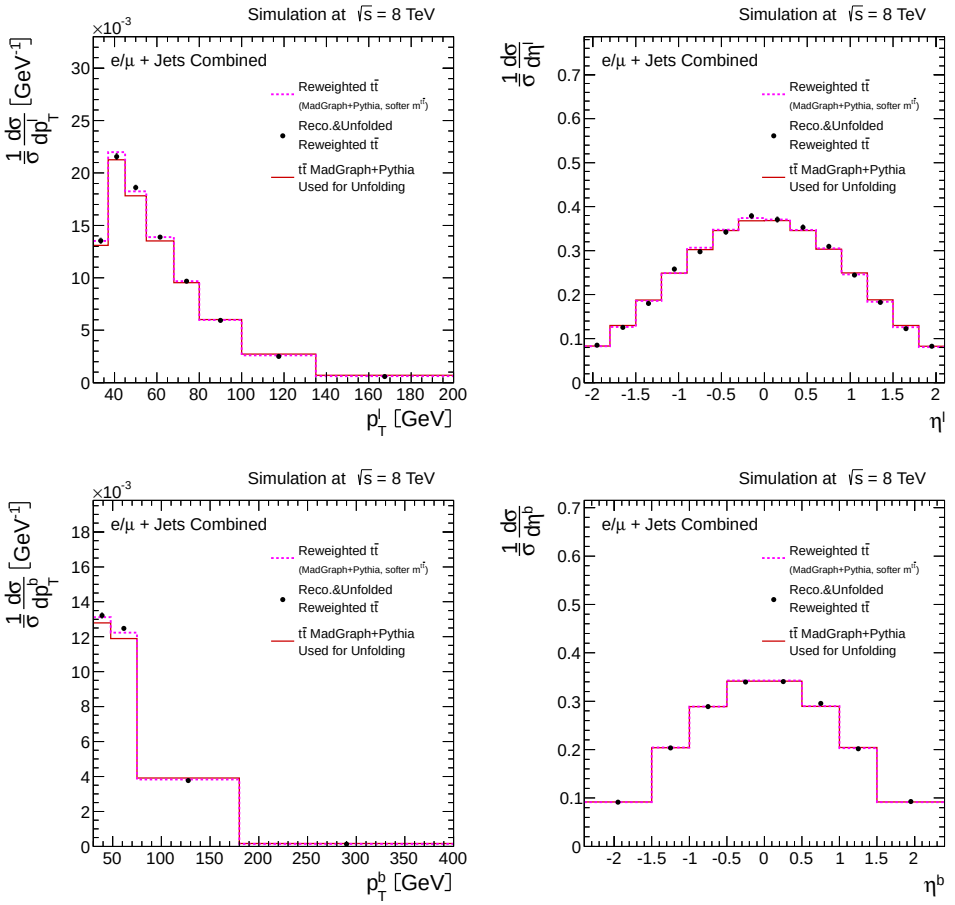

Figure 9.15: Results of the unfolding closure test with pseudo data featuring a softer $m^{t\bar{t}}$ spectrum for the lepton and b-jet cross section quantities in the visible particle level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

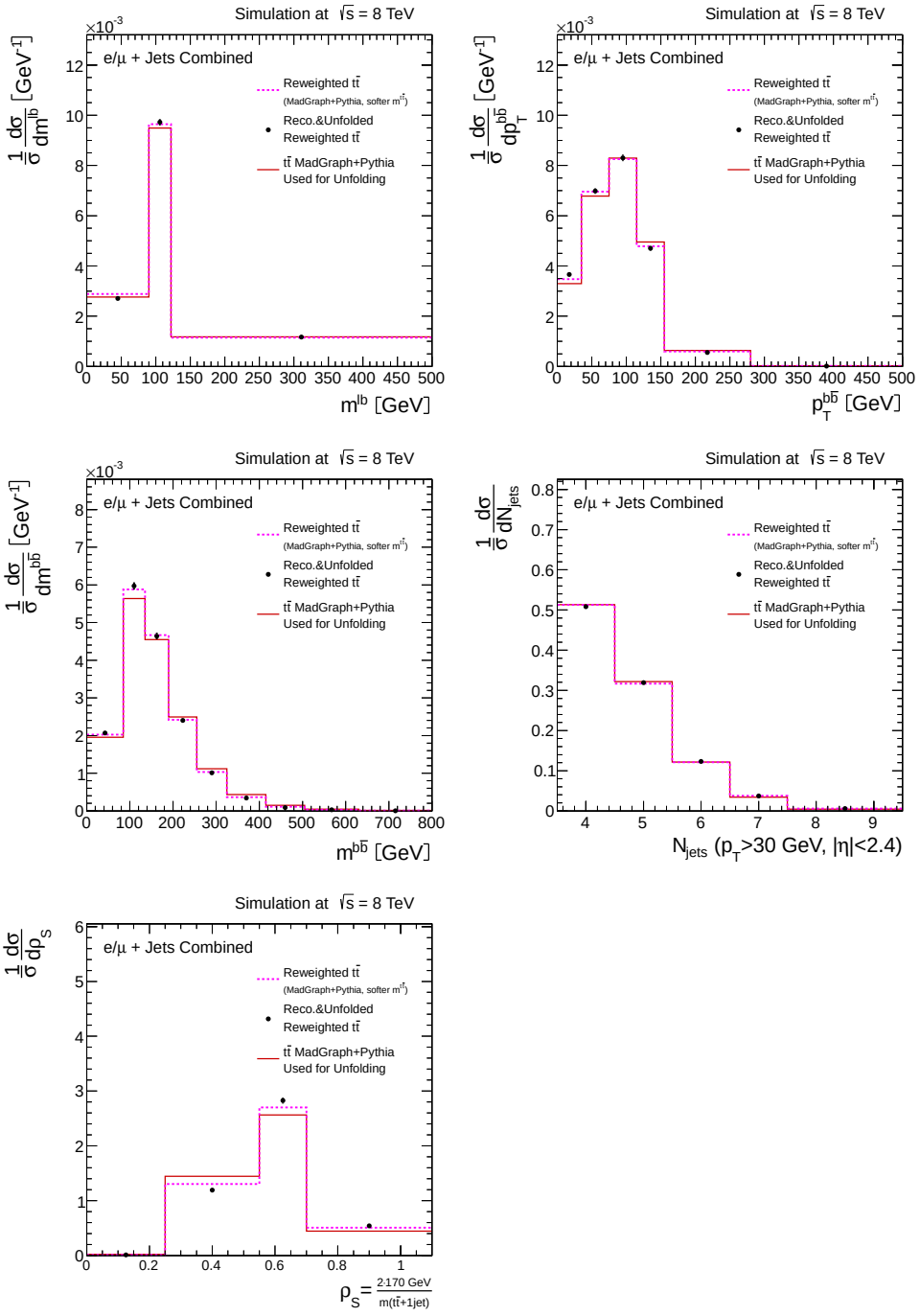

Figure 9.16: Results of the unfolding closure test with pseudo data featuring a softer $m^{t\bar{t}}$ spectrum for the jet multiplicity, the $b\bar{b}$ system cross section quantities, the invariant mass of the lepton and the leptonic b bjet and ρ_S in the visible particle level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

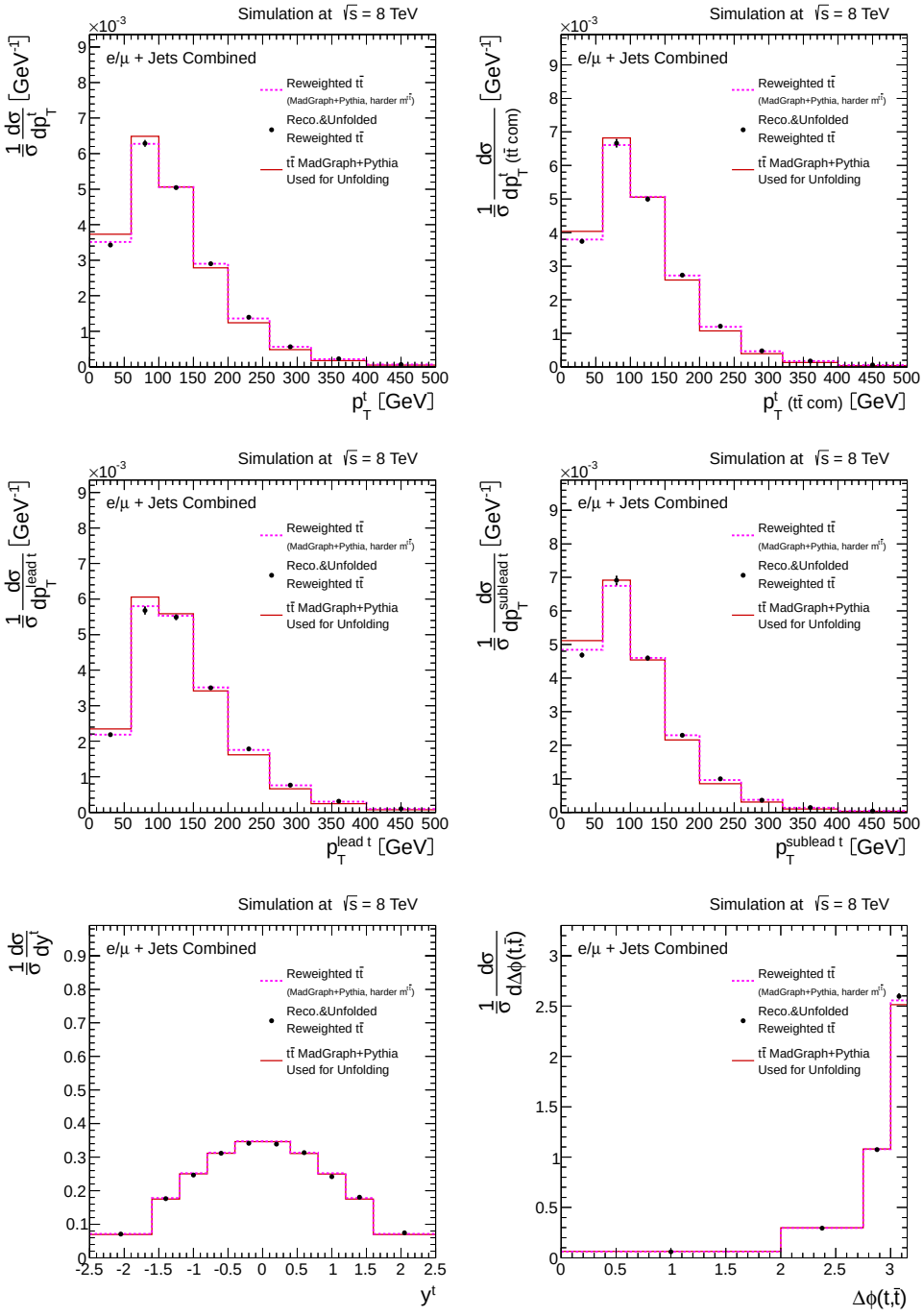

Figure 9.17: Results of the unfolding closure test with pseudo data featuring a harder $m^{t\bar{t}}$ spectrum for the cross section quantities related to the top quark in the fully extrapolated parton level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

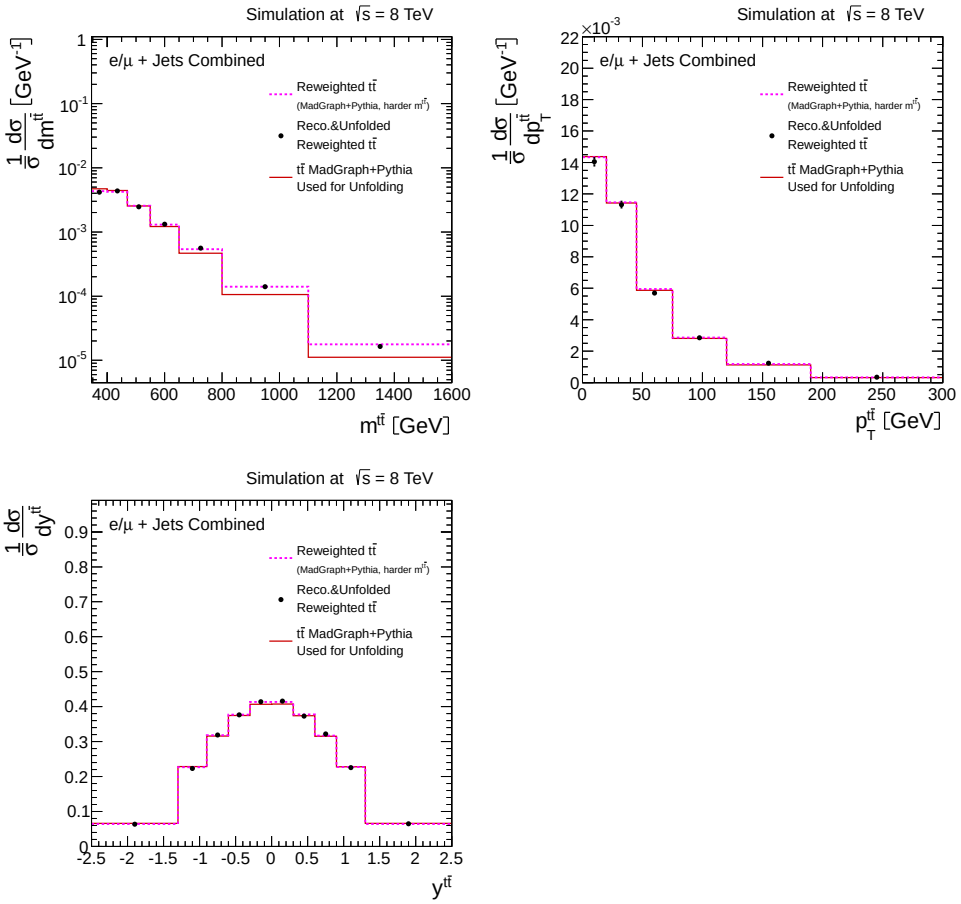

Figure 9.18: Results of the unfolding closure test with pseudo data featuring a harder $m^{t\bar{t}}$ spectrum for the cross section quantities related to the $t\bar{t}$ system in the fully extrapolated parton level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

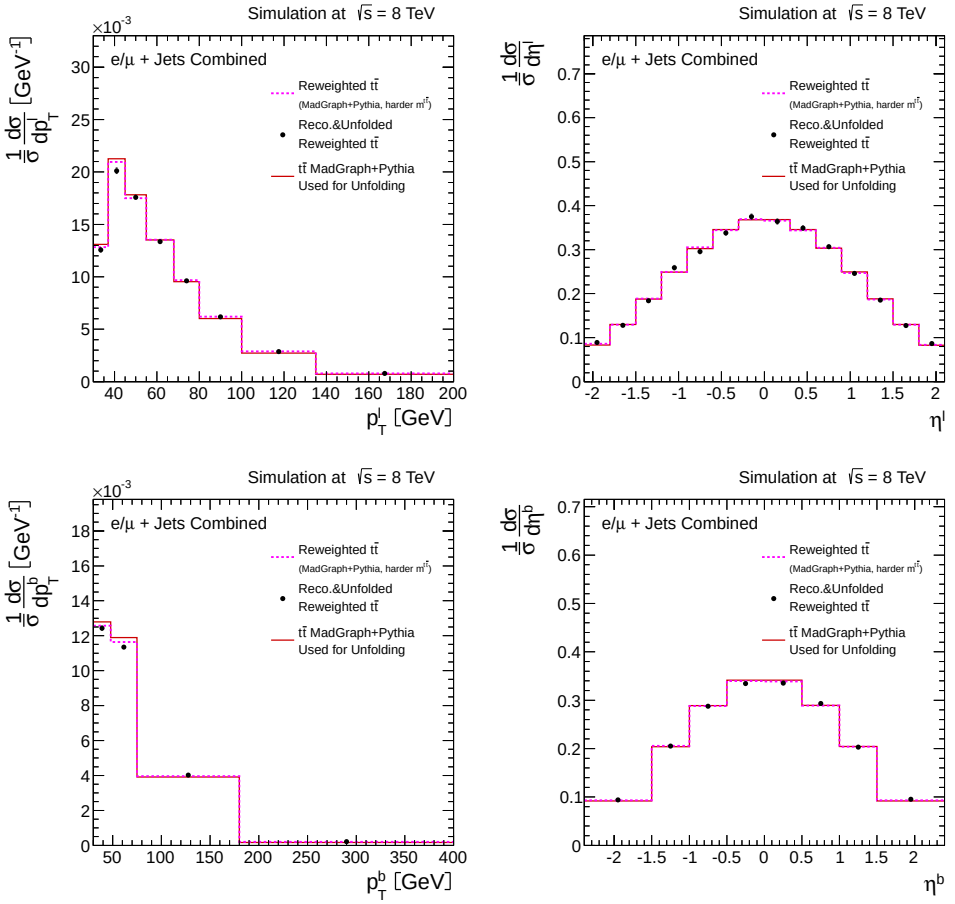

Figure 9.19: Results of the unfolding closure test with pseudo data featuring a harder $m^{t\bar{t}}$ spectrum for the lepton and b-jet cross section quantities in the visible particle level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

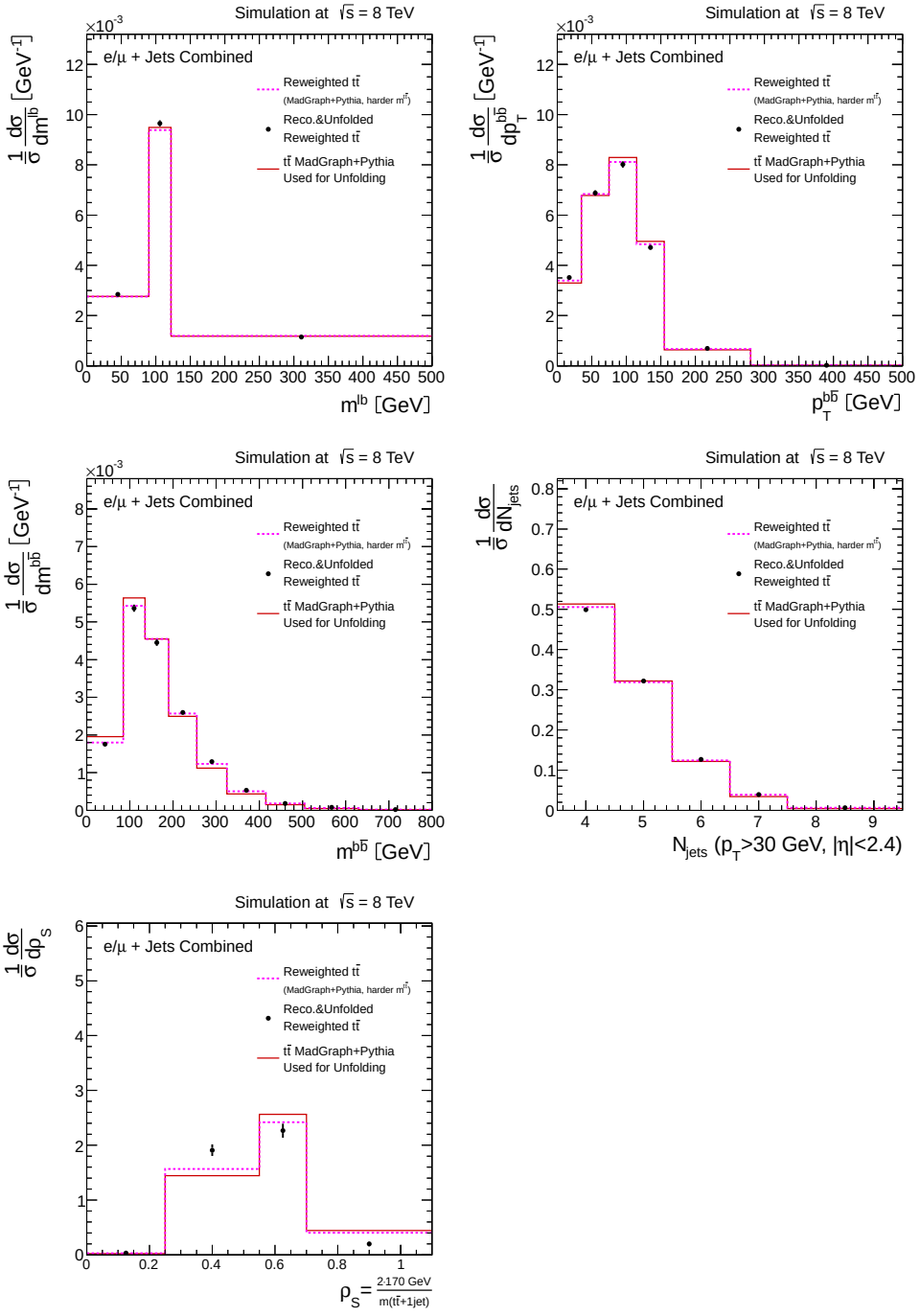

Figure 9.20: Results of the unfolding closure test with pseudo data featuring a harder $m^{t\bar{t}}$ spectrum for the jet multiplicity, the $b\bar{b}$ system cross section quantities, the invariant mass of the lepton and the leptonic b bjet and ρ_S in the visible particle level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

9.4.3 Closure Tests with an Additional Z' Signal

Another closure test is performed by creating pseudo data that contains in addition to the default processes a heavy resonance decaying into top-quark pairs [56]. This closure test gives rise to a distortion that reveals mainly in one bin of the invariant mass of the $t\bar{t}$ system.

To generate a scenario which is not too unrealistic but still in the accessible range of the analysis, a heavy and broad resonance ($m_{Z'}$ =1 TeV, $\Gamma_{Z'}$ =100 GeV) is chosen. A relatively large production cross section of $\sigma(pp \rightarrow Z') \cdot BR(Z' \rightarrow t\bar{t})$=5 pb is used. Therefore, a large impact on the cross section quantities is expected– far beyond the one observed in data for the reconstruction level event yields (see Chapter 7.4). As before, the default $t\bar{t}$ MC sample is used for the unfolding procedure to extract the normalised differential cross sections.

The corresponding results can be found in Figure 9.21 and 9.22 for the cross section quantities measured in the extrapolated parton level phase space and in Figure 9.23 and 9.24 for the cross section quantities measured in the visible particle level phase space.

Overall, the unfolding can recover the true shape of the pseudo data, which is involving an additional Z' signal, very well. Only marginal differences between the true and the measured shape of the pseudo data are observed. As studied in [23], the size of these differences depends on the absolute normalisation ($\sigma(Z') \cdot BR(Z' \rightarrow t\bar{t})$) of the additional Z' signal. For values of a few pb, the expected bias is covered by the total uncertainty of each differential measurement.

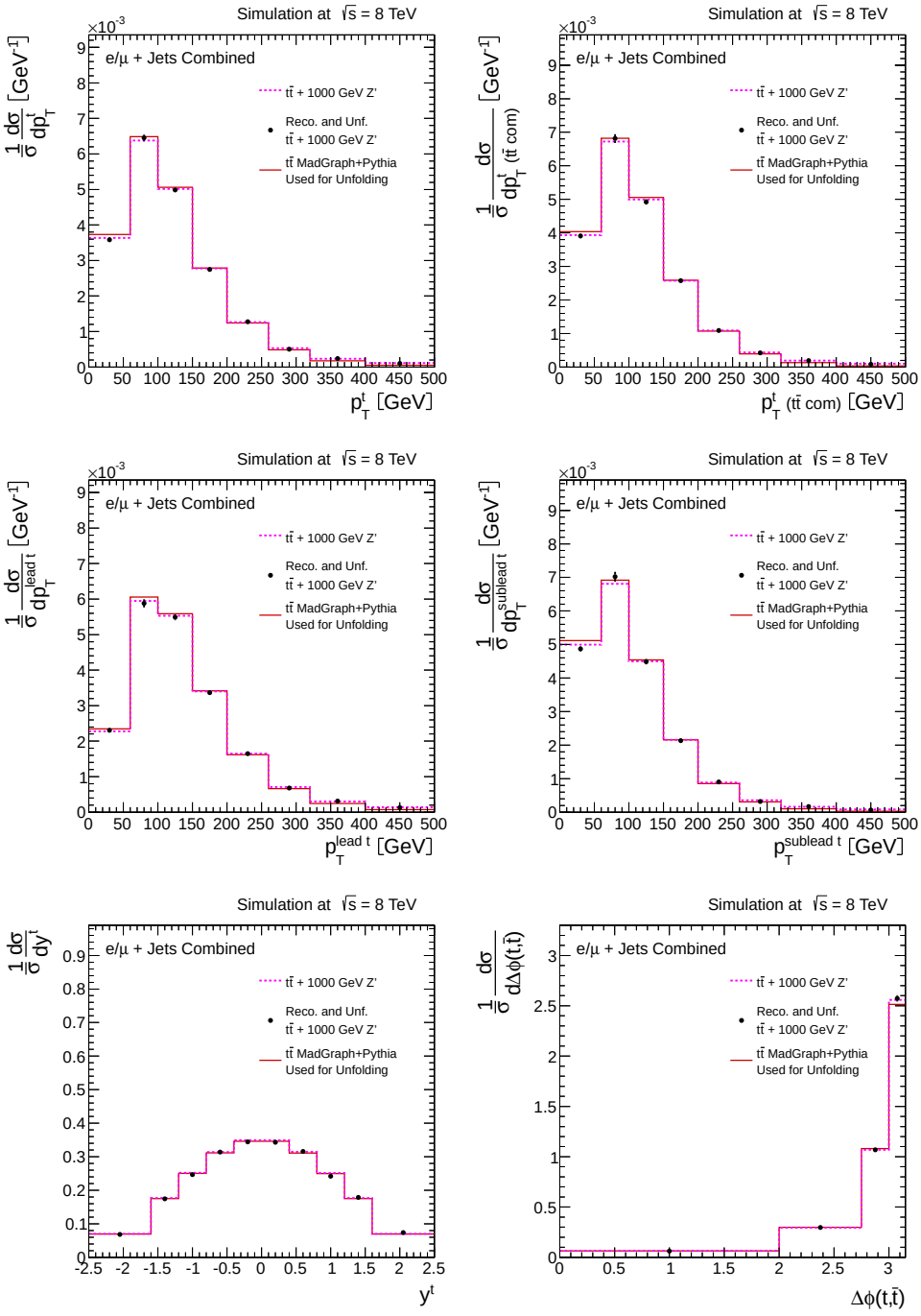

Figure 9.21: Results of the unfolding closure test with pseudo data featuring an additional Z'→tt̄ signal for the cross section quantities related to the top quark in the fully extrapolated parton level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

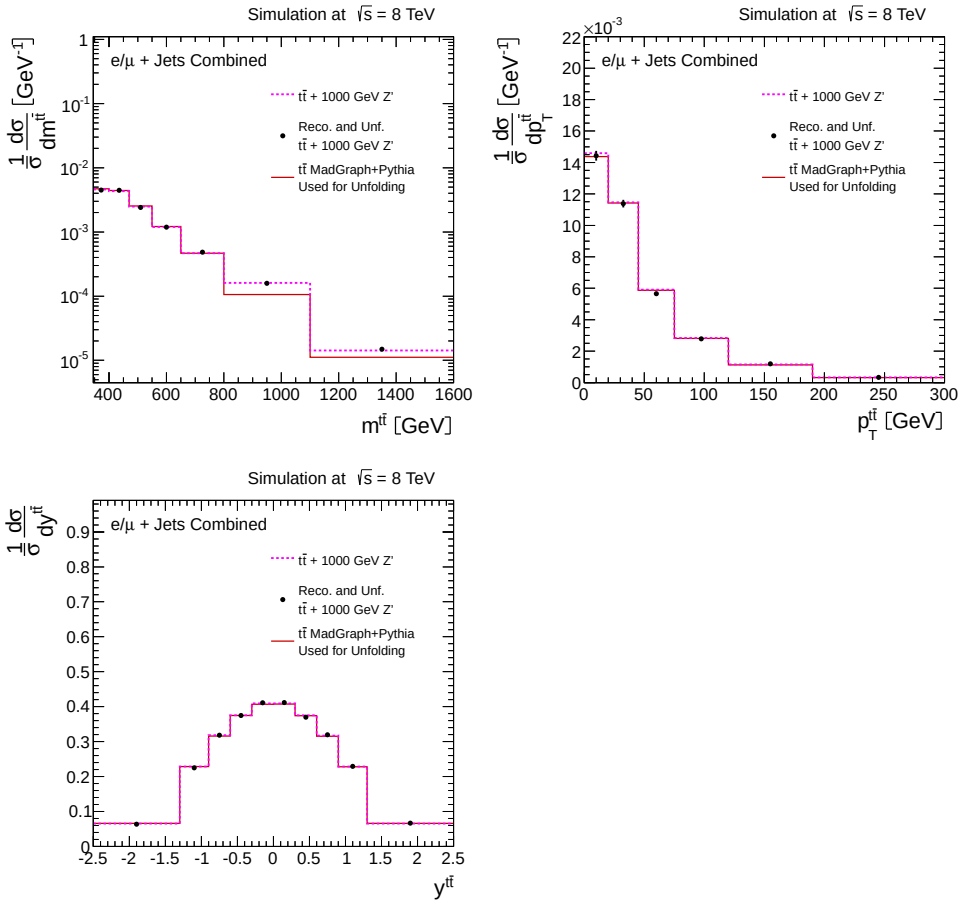

Figure 9.22: Results of the unfolding closure test with pseudo data featuring an additional Z'→$t\bar{t}$ signal for the cross section quantities related to the $t\bar{t}$ system in the fully extrapolated parton level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

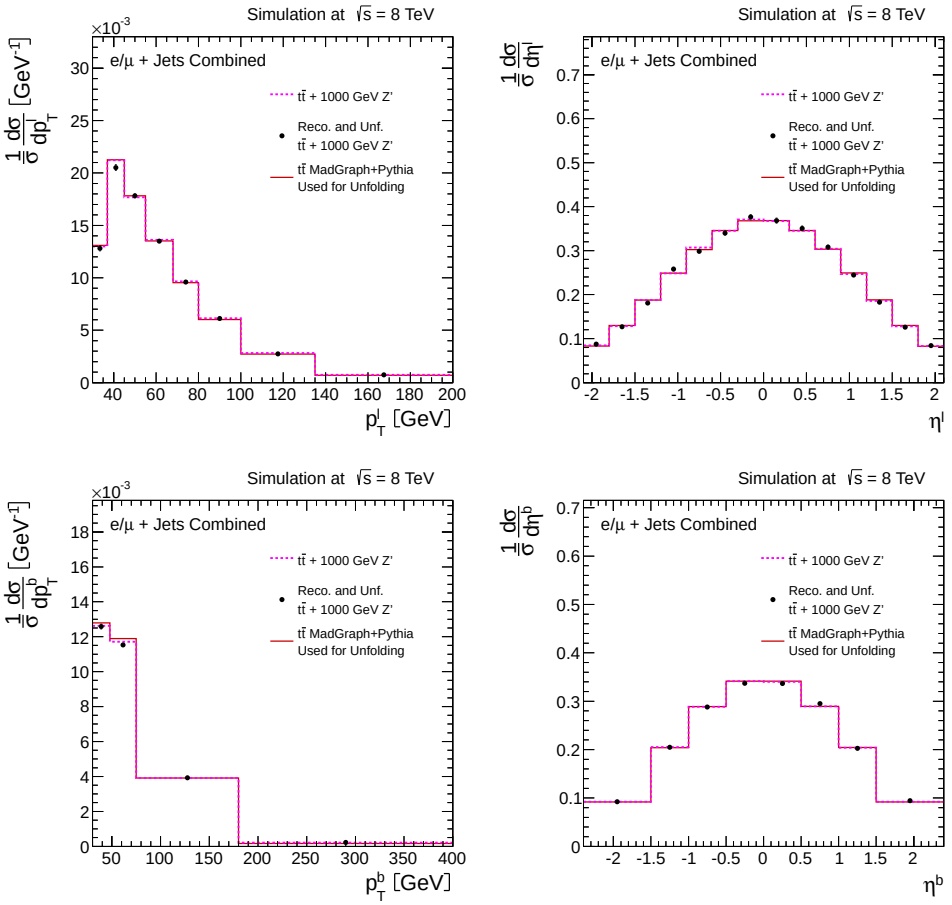

Figure 9.23: Results of the unfolding closure test with pseudo data featuring an additional Z'→$t\bar{t}$ signal for for the lepton and b-jet cross section quantities in the visible particle level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

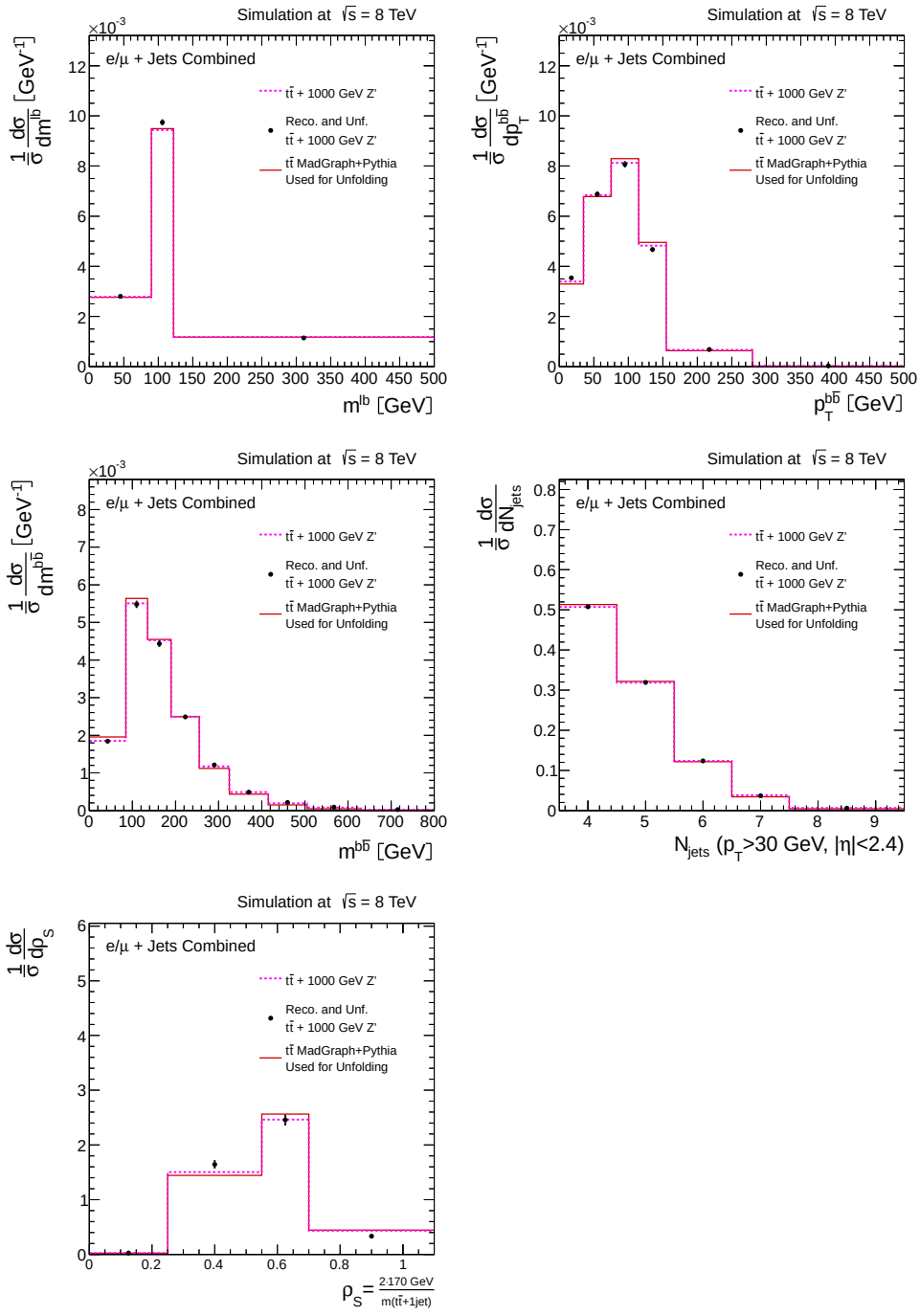

Figure 9.24: Results of the unfolding closure test with pseudo data featuring an additional Z'→tt̄ signal for for the jet multiplicity, the bb̄ system cross section quantities, the invariant mass of the lepton and the leptonic b bjet and ρ_S in the visible particle level phase space. The normalised differential cross section as extracted from unfolding the reconstruction level yield (black points) is compared to the true shape (pink line) for the pseudo data. The default MC prediction (red line) is used to obtain the response matrix for the unfolding procedure. Only the expected statistical uncertainties for the recorded luminosity are shown as error bars.

9.4.4 Iterative Unfolding

Another closure test is performed by iteratively applying the regularised unfolding procedure in data and correcting the assumed shape of the MC prediction to the measured result. This procedure is found to converge very fast. For the investigated top-quark and $t\bar{t}$ system quantities, the obtained difference for the extracted result between the first and last iteration step is small compared to the total uncertainty.

9.4.5 Testing Different Regularisation Schemes and Parameter Choices

To study the sensitivity of the analysis result regarding the choice of the specific unfolding scheme, the alternative SVD unfolding method as described in [182] and e.g. used in the ATLAS measurement of differential cross sections [22] is tested for the unfolding step of the analysis. In contrast to the regularised unfolding used for the standard analysis, the regularisation for SVD unfolding is performed using an integer parameter $k \leq N_{bins}$. Within this scheme, a large k parameter corresponds to a weak regularisation and equals a small τ parameter of the regularised unfolding method.

Furthermore, the impact of the choice of the regularisation parameter (τ) on the final result within the regularised unfolding is studied by varying τ around the optimal value (τ_d) obtained from the global correlation method.

The transverse momentum of the top quarks was chosen for these tests, as here the largest differences between MC prediction and data are observed for the reconstruction level event yields (see Chapter 7.4). Both tests are performed by repeating the analysis for p_T^t in data and pseudo data and comparing the extracted result with the result of the default analysis.

For pseudo data, τ_d is like for real data obtained by minimising the RMS global correlation of the extracted result. Moreover, the default $t\bar{t}$ MC prediction is consistently used to perform the unfolding.

The obtained results for real data as well as for pseudo data with a corrected $t\bar{t}$ component according to the softer p_T^t in data as obtained in [26] are shown in Figure 9.25.

In summary, all closure tests involving pseudo data reveal the expected behaviour regarding the choice of τ. For the default value τ_d, the true input shape of the pseudo data is reproduced by the measurement. Here, the observed difference between the extracted result and the true value is small ($< 1\%$ for the first bin) in comparison to the total uncertainty, which is in the order of 3-7% (see Chapter 10).
For large values of τ and small values of k, the regularisation is too strong, leading to a bias of the extracted result towards the MC prediction which is used to determine the response matrix and the regularisation condition of the unfolding procedure.
Furthermore, the results extracted for regularisation parameters $0 \leq \tau \leq 10 \cdot \tau_d$ of the regularised unfolding and parameters $k \geq 3$ of the SVD unfolding are observed to be very consistent to each other.

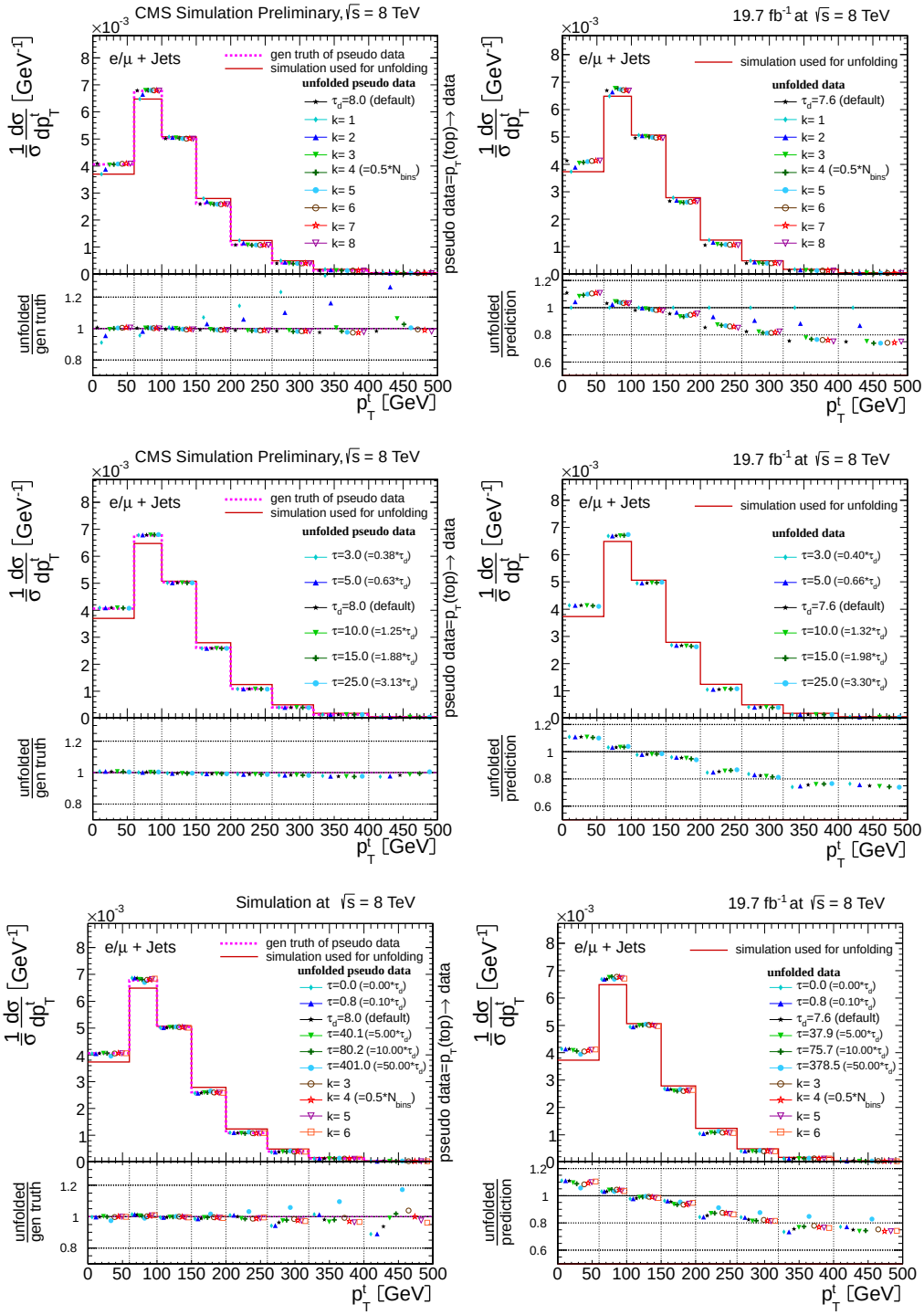

Figure 9.25: Extracted normalised differential cross section as function of p_T^t for different regularisation schemes and regularisation parameters. Top: all k parameters of the SVD unfolding method. Middle: different τ parameters around the default value (τ_d) for the regularised unfolding method as used in the analysis. Bottom: a mixture of different regularisation parameters for both unfolding methods. Left: pseudo data test with a softer p_T^t spectrum and ratio to truth of the respective pseudo data. Right: result for real data and ratio to the default MC prediction, which is used for the unfolding procedure.

Also the extracted results from real data are observed to be very similar among each other when using the SVD unfolding scheme with $k \geq 3$ and the regularised unfolding scheme with regularisation parameters around τ_d. A stable, non-oscillating result is also found without any regularisation ($\tau = 0$). This means, that migration effects within the studied event sample are already reduced to a reasonable low level by the optimised kinematic reconstruction, the selection and the choice of the analysis binning. Furthermore, in all scenarios with a regularisation that is not too strong, i.e. no bias is observed in the corresponding pseudo data closure test, a softer spectrum for p_T^t is obtained from data than predicted by MadGraph+pythia. This will be discussed in detail in Chapter 11.2.1 and 12.

Consequently, the obtained result for p_T^t in data is stable regarding the choice of the unfolding scheme and the exact choice of the regularisation parameter. Especially the kinematic region for $p_T^t < 150\,\text{GeV}$ is hardly effected ($< 2\%$) and even for larger values of p_T^t, the effect is fully covered by the uncertainties.

The obtained statistical correlation of the result is studied for several choices of the regularisation parameters in both regularisation schemes to proof that the minimal global correlation method is working. The statistical correlation matrices are obtained from the statistical covariance matrices which are calculated using pseudo experiments (see Chapter 9.3.1). In general, the entry (i,j) of the correlation matrix **CORR** is defined as:

$$\text{corr}_{ij} = \frac{\text{cov}_{ij}}{\sqrt{\text{cov}_{ii} \cdot \text{cov}_{jj}}}. \tag{9.23}$$

The obtained correlation matrices **CORR** of the extracted result in data are shown in Figure 9.26 for different choices of the continuous regularisation parameter τ in the regularised unfolding scheme and in Figure 9.27 for different choices of the discrete regularisation parameter k in the SVD unfolding scheme.

Good consistency is found between the two regularisation schemes and as expected, the minimum of the average correlation is obtained for the regularised unfolding scheme with $\tau = \tau_d$. Due to the normalisation of the final result, negative correlations of 20 to 50% are obtained between neighbouring bins while the correlation to next to neighbouring bins is small (typically $\approx 10\%$). For a weaker regularisation, the obtained negative correlations between neighbouring bins are larger and the correlations between next-to neighbouring bins is positive. In contrast, a stronger regularisation leads to large positive correlations between neighbouring bins and sizeable long-term correlations.

The comparison of the obtained statistical correlations of the result for the two regularisation schemes indicates that the obtained optimal τ parameter of the global correlation method corresponds to a fairly large k parameter (between 7 and 8) of the SVD unfolding scheme. As k=8 means that no regularisation is done at all, the regularisation strength for p_T^t, which is used for the data analysis, is not so intense.

Furthermore, the dependency of the obtained total (stat.$\oplus$syst.) uncertainty of the data measurement on the choice of the regularisation scheme and strength has been studied. For the SVD scheme with $k \geq 4$ and regularised unfolding scheme with

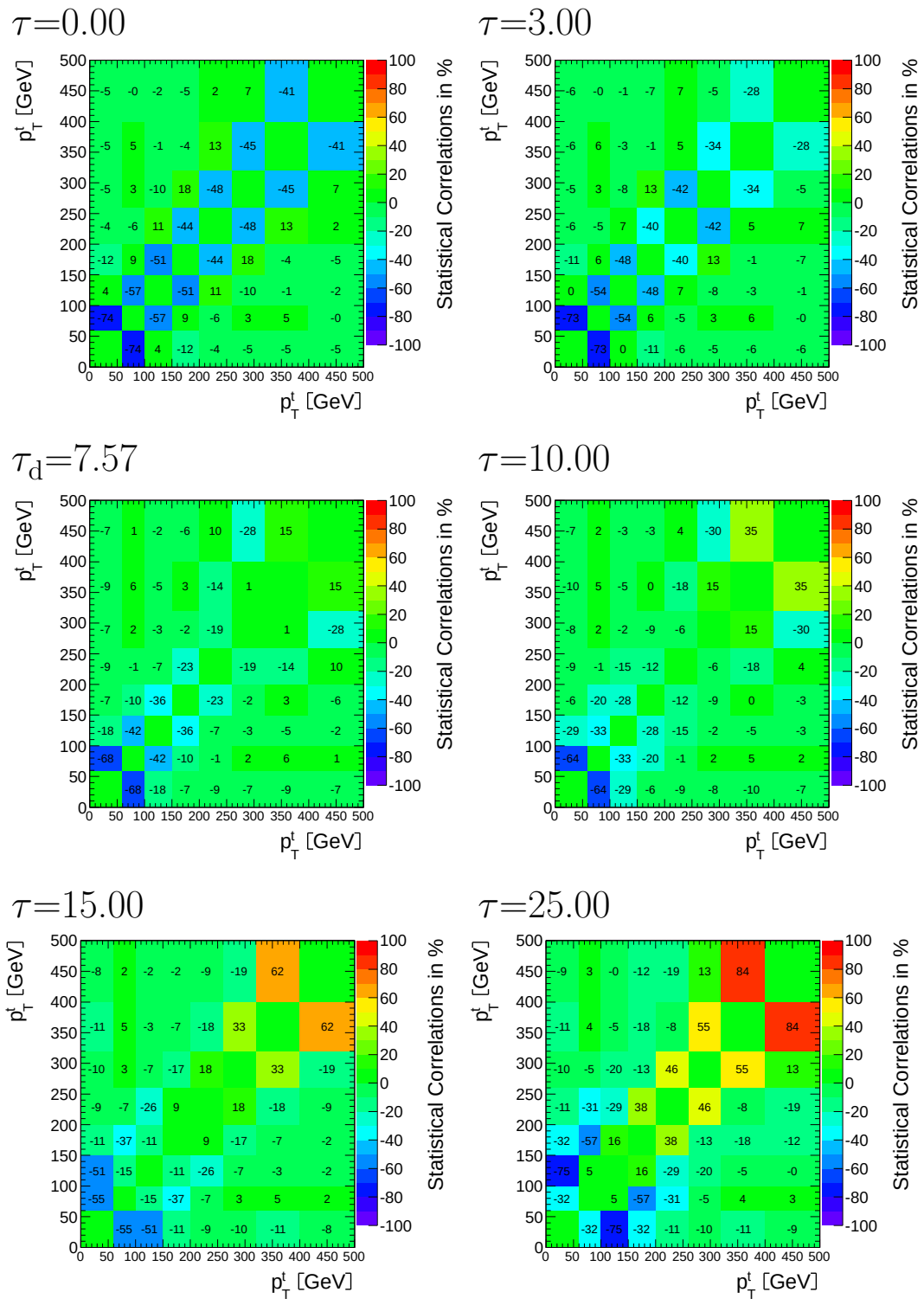

Figure 9.26: Obtained statistical correlation of the extracted normalised differential cross section in data as a function of p_T^t for different regularisation parameters τ around the standard value τ_d obtained by the minimal global correlation method in the regularised unfolding scheme.

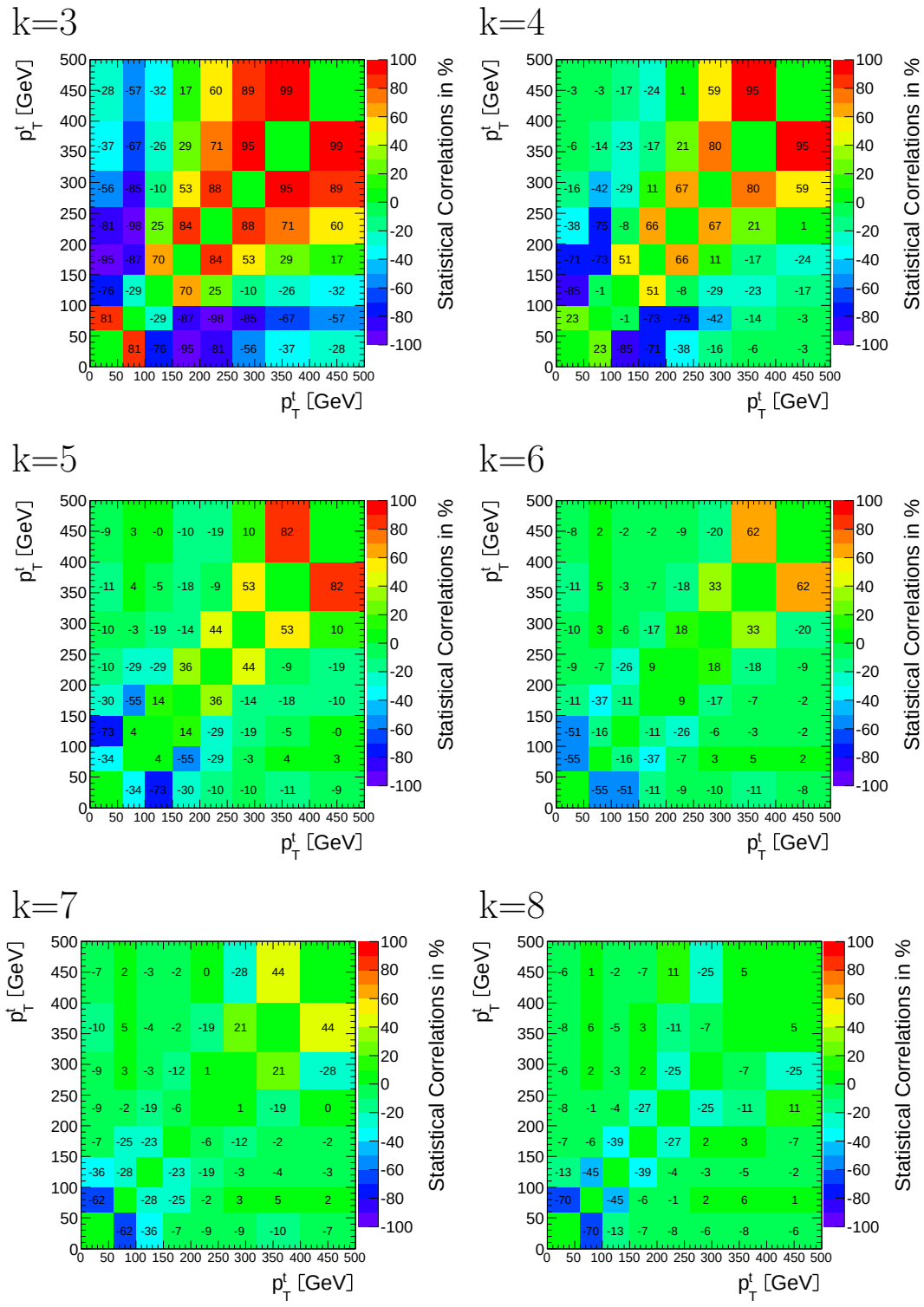

Figure 9.27: Obtained statistical correlation of the extracted normalised differential cross section in data as a function of p_T^t for different regularisation parameters k using the alternative SVD unfolding scheme.

$3 \leq \tau \leq 25$ ($\tau_{\mathrm{d}} = 7.57$), the relative uncertainty changes typically by at most 25%. On average, a smaller uncertainty is obtained for a stronger regularisation.

9.4.6 Conclusions for the Unfolding Tests

Different input shapes of the $\mathrm{t\bar{t}}$ component and an additional Z' signal in pseudo data are recovered when unfolding the reconstruction level event yield with the default MC prediction. This proves that the obtained result is independent of the choice of the MC prediction for the unfolding procedure for realistically small differences between the true shape in data and the shape used for the procedure. In this context, realistically small refers to variations in the order of observed shape differences between the data measurement and the MC prediction for the reconstruction level event yields. As expected, a small bias shows up for closure tests with pseudo data featuring an extreme, unrealistic shape difference which is not observed in real data (see Chapter 7.4).

Small remaining differences in the closure tests reflect non-perfect assumptions for the migrations, reconstruction efficiencies and the shape of the non-signal $\mathrm{t\bar{t}}$ component. This impact of the input MC prediction on the obtained cross section results is covered by the systematic uncertainties because a new response matrix is computed and used for the evaluation of every single uncertainty.

Furthermore, the transverse momentum of the top quarks is investigated further, as the largest differences between MC prediction and data are observed here for the reconstruction level event yields. A closure test for pseudo data with a shape corrected according to the shape obtained in real data [26] proofs that no bias is introduced for the chosen unfolding setup. Additionally, the result obtained from data is stable when varying the regularisation parameter (Figure 9.25 right) within the range which is expected to be unbiased from the pseudo data test (Figure 9.25 left). Moreover, also the obtained results when using the alternative SVD regularisation scheme are consistent.

In summary, all tests performed show that the regularised unfolding procedure as used in this thesis is robust, well understood and the obtained results are consistently stable when changing the assumptions moderately. No indication of a possible bias of the result towards the prediction that is used for the unfolding procedure is found.

Chapter 10

Systematic Uncertainties

Already the precision of the CMS $t\bar{t}\to\ell$+jets differential cross section analyses at $\sqrt{s} = 7$ TeV [23] was limited by systematic uncertainties. Therefore, their correct and complete evaluation is an important aspect for the assessment of the final results.

In general, there are two distinct categories of systematic uncertainties. Experimental uncertainties discussed in Chapter 10.1 are related to detector effects and reflect the experimental precision of the measured analysis input (luminosity, measured energies of the jets, etc.). In contrast, modelling uncertainties discussed in Chapter 10.2 reflect the precision of the theoretical prediction of the simulation which is e.g. used to correct migration effects or to extrapolate the measured result from the visible to the full phase space.

The calculation of the systematic uncertainties largely follows standard methods and recommendations of the TopPAG or is treated equally to the $\sqrt{s} = 7$ TeV analysis.
For all differential cross section measurements, the uncertainties are significantly reduced by the normalisation (see Chapter 8.3). Uncertainties, which change the content in all bins by the same factor (i.e. flat SFs), cancel completely, while other uncertainties are at least reduced in comparison to an absolute differential cross section measurement. Therefore, the largest contribution is expected from uncertainties which affect the shape of the differential distributions. Consequently, additional uncertainty sources with impact on the obtained shape are evaluated.

All uncertainties are quantified by repeating the data analysis with systematically varied input from the simulation, i.e. a different response matrix for the unfolding and a different prediction for the expected background contribution and $t\bar{t}$ signal fraction. The difference between the result obtained with systematically varied input and the standard analysis is quoted as systematic uncertainty. For uncertainty sources with up- and down-variation, the average value is calculated individually for each bin and cross section quantity and quoted as uncertainty.
Finally, the total uncertainty is derived by adding all single contributions individually for each bin and cross section quantity in quadrature.

A method to derive the full covariance matrix for all systematic uncertainties of the

normalised differential cross section measurements is introduced in Chapter 10.3.

10.1 Experimental Uncertainties

10.1.1 Luminosity

The luminosity is centrally measured by CMS using Van der Meer scans and the official value for the uncertainty is 2.5%(syst.)$\oplus$0.5%(stat.) [96]. For the normalised differential cross sections, the luminosity uncertainty cancels almost completely and only a small residual contribution remains due to the normalisation of the background predictions.

10.1.2 Pile-Up

The pile-up (PU) model estimates the mean number of PU interactions to be about 20 events for the $\sqrt{s} = 8$ TeV pp collision data used for this thesis. As explained in [137], the PU estimation is based on the total inelastic pp cross section, which CMS determines to be 69.4mb.

Following the official recommendation, this cross section is varied by $\pm 5\%$ to obtain different PU estimates and hence to evaluate a systematic uncertainty related to the modelling of the PU. The quoted variation includes luminosity and cross section uncertainties as well as an uncertainty related to the modelling and physics aspects of the PU simulation.

As illustrated in Figure 5.4, this variation conservatively covers all differences for the number of primary vertices distribution, which is directly sensitive to the underlying PU.

Technically, the PU uncertainty is evaluated by redoing the analysis with adapted event weights for the PU reweighting, which are obtained with the varied total inelastic pp cross section and the method detailed in Chapter 5.3.3.

10.1.3 Jet Energy Scale

The estimation of systematic uncertainty on the jet energy scale follows the recommendation of the *JetMET* group.

Technically, the reconstructed jet energy is shifted as a function of η and p_T of the jet using the provided uncertainties [187]. This shift is also propagated to the reconstructed missing transverse energy. No analysis-specific uncertainty for a special flavour composition is calculated.

10.1.4 Jet Energy Resolution

As detailed in Chapter 6.4.1, the JER in simulation needs to be corrected to equal the one observed in data. The officially recommended up and down variations for the JER SFs as listed in Table 6.1 are used to evaluate this uncertainty.

10.1.5 Data-Driven Efficiency Corrections

As described in in Chapter 6.7, the efficiency for the trigger- and lepton-selection, as well as for the b-jet identification are corrected using data to MC SFs obtained from an independent data study. For all corrections, the uncertainty of the provided SFs is used to derive the corresponding uncertainties.

All SFs are obtained as a function of p_T and η of the lepton or jet. For the determination of normalisation uncertainties, which are dominant for absolute cross sections, the SFs are varied up or down in all bins of the parametrisation simultaneously.

Shape uncertainties are addressed as the uncertainty for a flat variation of the SF mostly cancels for the normalised differential cross sections. Thus, the SFs of different kinematic regions are varied in different directions following the method described in [23]. In most of the cases, two bins are chosen such that they are divided by the median of the respective distribution in the phase space of the applied event selection. The SFs in these bins are then shifted antagonistically, i.e. simultaneously up in bin one and down in bin two and vice-versa.

Trigger and Lepton Efficiency

The uncertainties related to the lepton and trigger selection are determined by the statistical uncertainties of the T&P scale factors and an additional conservative 1% flat systematic uncertainty to account for possible differences in the obtained efficiency between Z boson and $t\bar{t}$ events, e.g. due to different p_T and η distributions or different hadronic activities. In contrast to the 1% systematic uncertainty, the statistical uncertainty is almost negligible for the majority of the bins.

For the normalisation uncertainty, both uncertainty contributions are added in quadrature and the variation is performed in the same direction for all bins of the parametrised SF.
Shape uncertainties in η are estimated by shifting the SF antagonistically up and down for $|\eta|$ above and below 0.7. To assign a shape uncertainty depending on p_T, the SF is shifted antagonistically up and down for p_T above and below 55 GeV.

All described normalisation and shape variations are done independently and added in quadrature.

B-Tagging Efficiency

The determination of the uncertainty related to the b-jet identification follows the recommendations of the BTV group.

For the normalization uncertainty, the SF is shifted by its uncertainty in the same direction independently of the kinematic region.
In order to determine a shape uncertainty, the scale factor is shifted antagonistically up and down above and below $p_T = 65\,\text{GeV}$ for a p_T depended uncertainty and independently for $|\eta|$ above and below 0.7 for an η dependent uncertainty. These variations are performed with half of the corresponding quoted uncertainty, so that the difference between the applied scale factors in the two different bins equals the full uncertainty.

All described normalisation and shape variations have been done independently and added in quadrature.
In addition, also the misidentification rate of the b-jet identification procedure is varied according to the recommendations of the BTV group. The effect on the measured cross sections was found to be negligible.

10.2 Modelling Uncertainties

10.2.1 Background Modelling

After the final selection step, the expected remaining contribution from non-$t\bar{t}$ events is small (≈5%). About half of these events are expected to originate from the production of single top quarks and another ≈20% each from $t\bar{t}$ pairs accompanied by an additional vector boson and events with W-boson production accompanied by additional jets. All other processes are negligible (see Chapter 6.8 for details).

As discussed previously, not only the rate but also the shape of the background distributions has to be understood adequately. Therefore, all systematic uncertainties like e.g. the JES uncertainty are also propagated to the background samples, forming a set of shape variations.

In addition, the total amount of background events which is expected for the integrated data luminosity using theory cross sections (discussed in Chapter 5.3.2), is varied conservatively:

- single top-quark production: ± 30%
- W+jets: ± 100%
- Z+jets: ± 100%
- WW,WZ,ZZ: ± 30%
- $t\bar{t}$ +V, V∈(W,Z,γ): ± 100%

These variations are much larger than the uncertainty of the cross section and are expected to cover also a potential mismodelling of efficiencies or flavor compositions. For

W/Z+jets events, uncertainty on the modelling of the heavy flavor content is observed to be large [188]. For the $t\bar{t}$ $+\gamma$ MC prediction, the quoted uncertainty is expected to cover also the overlap with $t\bar{t}$ $+\gamma$ events produced by the PS in the default $t\bar{t}$ sample. As the reconstructed shape for the event yields of the measured cross section quantities differs between $t\bar{t}$ signal events and the background processes, any variation of the total rate of a background process leads to a shape variation for the reconstructed event yields.

Because of the very low contribution from QCD multijet-events after the final selection step, this process is completely neglected. The negligibly low fraction of QCD multijet-events has been checked thoroughly by studying sensible quantities like the lepton isolation and is also confirmed by other CMS top-quark analysis with a similar event selection [169].

As single top events are expected to be the majority of the remaining background events, they are treated specially. In addition to the already discussed shape and rate uncertainties, a shape uncertainty resulting from a variation of the Q^2 scale (as defined in Chapter 10.2.2) is evaluated.

Furthermore, $t\bar{t}$ events with non-signal final states are treated synchronous with the $t\bar{t}$ signal events by evaluating all systematic uncertainties (including also $t\bar{t}$ modelling uncertainties like the top-quark mass, the Q^2 scale and the ME-PS matching threshold) at the same time. Finally, the effect on the reconstructed event yields is propagated to the final result through the $t\bar{t}$ signal fraction which is assessed separately in each bin (see Chapter 8.3).

10.2.2 Hard Scattering Q^2 Scale

The MC prediction depends on the choice of the hard-scattering scale Q^2 (see Chapter 5.3.2). Different default values of Q^2 are realised depending on the MC generator [189]:

- MadGraph: $Q^2_{\text{def}} = m^2_{\text{top}} + \sum p^2_{\text{T}}$
- powheg: $Q^2_{\text{def}} = m^2_{\text{top}}$

The variation of Q^2 addresses renormalisation and factorisation scale uncertainties of the ME calculation as well as the amount of ISR and FSR because the starting/evolution scale of the parton shower is changed simultaneously.

To test the sensitivity of the analysis concerning the choice of Q^2, specific samples with $Q^2_{\text{up}} = 4 \cdot Q^2_{\text{def}}$ and $Q^2_{\text{dn}} = 0.25 \cdot Q^2_{\text{def}}$ as listed in Table 5.3 for the MadGraph+pythia $t\bar{t}$ MC prediction and in Table 5.4 for the powheg+pythia single top-quark MC prediction are used.

10.2.3 Matching Threshold between Parton Shower and Matrix Element Calculation

For the $t\bar{t}$ sample, the matching threshold between the ME calculation, which is done by MadGraph, and the parton shower, which is simulated using pythia, is varied by a factor of 2 and 0.5. This probes additional jet production and ISR/FSR modelling.

Analogously to the Q^2 uncertainty, dedicated MC predictions as listed in Table 5.3 are used to evaluate the ME-PS matching threshold uncertainty.

10.2.4 Hadronisation Model

The hadronisation process is a complex mechanism (see Chapter 5.3.1) and phenomenological models are used to describe this part in MC predictions.

As different approaches are used within pythia and herwig, these two MC generators are used to address this modelling uncertainty for $t\bar{t}$ events. Due to technical limitations, e.g. the p_T or angular based ordering of the parton shower, not every ME generator can be interfaced to pythia and herwig. To date, only the combinations MadGraph+pythia, mc@nlo+herwig and powheg+pythia result in physically consistent predictions.

In order not to pick up additional higher order modelling differences between MadGraph+pythia and mc@nlo+herwig, the absolute hadronisation uncertainty is evaluated comparing the analysis results obtained with the two NLO ME generators powheg+pythia and mc@nlo+herwig. The obtained absolute difference in each bin is conservatively addressed as two-sided uncertainty ($\pm$) individually for each cross section quantity.

10.2.5 W-Boson Branching Ratio

The LO BRs of the MadGraph+pythia $t\bar{t}$ MC prediction are corrected to the measured values in data (see Chapter 6.7). The uncertainties of the BR measurement in data are propagated to the measured cross sections but found to be negligibly small.

10.2.6 Parton Density Functions

For the simulation of the $t\bar{t}$ signal with MadGraph+pythia, the CTEQ6L1 [120] PDF set is used. The impact of the PDF model uncertainty on the measurement has been evaluated by reweighting the obtained distributions according to the maximum shape variation of the PDF uncertainties.

First, the effect of the PDF choice on the shape of the generator truth distribution is studied for each cross section quantity. As no uncertainty for the LO CTEQ6L1 PDF set is provided, all variations of the NLO CT10 [190] PDF set are used. A relative

change of the shape $\mathrm{SF}^{\mathrm{PDF\ up/down}}_{\mathrm{bin\ i}}(\mathrm{x})$ is calculated for all variations j individually for each cross section quantity x and each bin i.
In the first step, the minimum ($\mathrm{MIN}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x})$) and maximum ($\mathrm{MAX}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x})$) variation for each bin of every normalised differential distribution is calculated:

$$\mathrm{MAX}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x}) = \max_{\mathrm{j}} \left(\frac{1}{\sigma_{\mathrm{t\bar{t}}}} \frac{\mathrm{d}\sigma}{\mathrm{dx}} \right)_{\mathrm{bin\ i}} (\mathrm{CT10,\ variation\ j}) \tag{10.1}$$

$$\mathrm{MIN}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x}) = \min_{\mathrm{j}} \left(\frac{1}{\sigma_{\mathrm{t\bar{t}}}} \frac{\mathrm{d}\sigma}{\mathrm{dx}} \right)_{\mathrm{bin\ i}} (\mathrm{CT10,\ variation\ j})\,. \tag{10.2}$$

From these, the maximum shape variations ($\mathrm{UP}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x})$,$\mathrm{DOWN}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x})$) are build by varying the normalised cross section linearly from $\mathrm{MIN}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x})$ in the first to $\mathrm{MAX}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x})$ in the last bin and vice versa:

$$\begin{aligned} \mathrm{DOWN}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x}) &= \mathrm{MIN}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x}) \\ &\quad + \left(\mathrm{MAX}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x}) - \mathrm{MIN}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x}) \right) \cdot \frac{\text{bin i} - \text{first bin}}{\text{last bin} - \text{first bin}} \end{aligned} \tag{10.3}$$

$$\begin{aligned} \mathrm{UP}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x}) &= \mathrm{MAX}^{\mathrm{PDF\ max}}_{\mathrm{bin\ i}}(\mathrm{x}) \\ &\quad - \left(\mathrm{MAX}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x}) - \mathrm{MIN}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x}) \right) \cdot \frac{\text{bin i} - \text{first bin}}{\text{last bin} - \text{first bin}}. \end{aligned} \tag{10.4}$$

To transfer the variations of the CT10 PDF set to the standard MC prediction, relative SFs are derived:

$$\mathrm{SF}^{\mathrm{PDF\ up/down}}_{\mathrm{bin\ i}}(\mathrm{x}) = \frac{\mathrm{UP/DOWN}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x})}{\left(\frac{1}{\sigma_{\mathrm{t\bar{t}}}} \frac{\mathrm{d}\sigma}{\mathrm{dx}} \right)_{\mathrm{bin\ i}} (\mathrm{CT10, default})}. \tag{10.5}$$

In the next step, systematically varied MC predictions are obtained by weighting individually each cross section quantity x of the default $\mathrm{t\bar{t}}$ MC prediction according to the relative change $\mathrm{SF}^{\mathrm{PDF}}_{\mathrm{bin\ i}}(\mathrm{x})$ determined in the previous steps.
Finally, the PDF uncertainty is evaluated by repeating the analysis with the systematically varied $\mathrm{t\bar{t}}$ MC predictions: modified generator truth distribution, modified reconstruction level event yield and modified response matrix.

10.2.7 Top-Quark Mass

To evaluate the uncertainty due to the assumed top-quark mass of 172.5 GeV in the simulation, the whole analysis chain is performed using specific $\mathrm{t\bar{t}}$ MC samples, which are created with two different top-quark masses of 171.5 GeV and 173.5 GeV (see Table 5.3). In comparison to the result of the single most precise CMS measurements of the top-quark mass [169, 174], this is a conservative estimation.

10.3 Definition of the Systematic Covariance Matrices

The correlation between the bins of the final normalised differential cross sections are a helpful information for analyses using these measured distributions as input. Examples for such follow-up analyses are PDF fits [191, 192] or the extraction of the top-quark mass from differential cross sections [49], which is discussed in more detail in Chapter 13. Furthermore, the full set of correlations is required to perform a χ^2-test to compare different theory calculations to this measurement.

As illustrated in Equation 9.23, the correlation is calculated from the covariance matrix, which is defined by Equation 9.7. For the statistical uncertainties, the covariance matrices are obtained from pseudo experiments within the unfolding procedure as explained in Chapter 9.3.1 and already done in [23].
As the total uncertainty is dominated by the systematic uncertainties, also the total covariance matrix is expected to have a large contribution from systematic uncertainty sources. To derive the covariance matrices for the systematic uncertainties, a simple statistical approach [193] explained in the following is chosen.

For every bin i, the calculated final normalised differential cross section is given as:

$$a_i = \left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{\text{bin i}}. \tag{10.6}$$

The absolute difference between the nominal result and the result obtained for a systematic variation sys is given as:

$$d_i^{\text{sys up/dn}} = a_i^{\text{sys up/dn}} - a_i^{\text{nominal}}. \tag{10.7}$$

For almost all uncertainty sources, the final absolute value Δ_i^{sys} of the systematic uncertainty sys is obtained from the symmetrisation of an up and down variation:

$$\Delta_i^{\text{sys}} = \frac{1}{2}\cdot\left[|d_i^{\text{sys up}}| + |d_i^{\text{sys down}}|\right]. \tag{10.8}$$

Consequently, $\frac{\Delta_i^{\text{sys}}}{a_i^{\text{nominal}}}$ is the relative uncertainty in bin i for a given systematic variation sys. Due to the definition of the symmetrisation, the information of the sign of the variation is lost for all symmetrised uncertainties. While this information is not needed to obtain the uncertainties, it is important to derive the full covariance matrix. Per convention, the sign of the up variation is used further on:

$$d_i^{\text{sys}} = \Delta_i^{\text{sys}} \cdot \text{sign}\left(d_i^{\text{sys up}}\right). \tag{10.9}$$

Furthermore, the symmetrisation procedure does not conserve the normalisation of the systematically varied distributions $a_i^{*\ \text{sys up/down}}$:

$$\int_{\text{bins i}} a_i^{*\ \text{sys up/down}} = \int_{\text{bins i}} a_i^{\text{nominal}} \pm d_i^{\text{sys}} \neq 1. \tag{10.10}$$

Therefore, it is ensured by redetermining the normalised differential cross section $a^{*\ \mathrm{sys\ up/down}}_{i,\mathrm{norm}}$ for the symmetrised up and down variation:

$$a^{*\ \mathrm{sys\ up/down}}_{i,\mathrm{norm}} = \frac{a^{*\ \mathrm{sys\ up/down}}_{i}}{\int\limits_{\mathrm{bins\ i}} a^{*\ \mathrm{sys\ up/down}}_{i}}. \tag{10.11}$$

Then, $a_{i,\mathrm{norm}}$ is used to re-derive the symmetrised absolute uncertainty d^{sys}_{i}:

$$d^{*\ \mathrm{sys}}_{i} = \Delta^{*\ \mathrm{sys}}_{i} \cdot \mathrm{sign}\left(d^{*\ \mathrm{sys\ up}}_{i}\right). \tag{10.12}$$

This leads to an iterative approach which is found to converge already after one iteration and the final systematic covariance matrix for a single systematic uncertainty sys is then given as:

$$\mathrm{cov}^{\mathrm{sys}}_{i,j} = d^{*\ \mathrm{sys}}_{i} \cdot d^{*\ \mathrm{sys}}_{j}. \tag{10.13}$$

The described procedure is applied for all systematic uncertainties and the total systematic covariance matrix is obtained by the sum of all separate results:

$$\mathrm{cov}^{\mathrm{sys\ tot}}_{i,j} = \sum_{\mathrm{all\ sys}} \mathrm{cov}^{\mathrm{sys}}_{i,j}. \tag{10.14}$$

For the final results, the uncertainty values before the iterative process are quoted and only the obtained covariance matrices are used to quantify correlations of the final normalised differential cross sections.

Potential bottlenecks of this statistical ansatz are the assumption of Gaussian uncertainties and the choice of the sign for d^{sys}_{i}. It was checked that this procedure changes the total systematic uncertainty in each bin only slightly ($\leq$0.5% on average) and that the choice of the sign for d^{sys}_{i} has no impact on the obtained covariance matrices.

Combining the the obtained covariance matrices for the statistic and systematic uncertainties the total covariance matrix is obtained:

$$\mathrm{cov}^{\mathrm{tot}}_{i,j} = \mathrm{cov}^{\mathrm{stat}}_{i,j} + \mathrm{cov}^{\mathrm{sys\ tot}}_{i,j}. \tag{10.15}$$

Finally, the total correlation matrix of the measured normalised differential cross sections is given by:

$$\mathrm{corr}^{\mathrm{tot}}_{i,j} = \frac{\mathrm{cov}^{\mathrm{tot}}_{i,j}}{\sqrt{\mathrm{cov}^{\mathrm{tot}}_{i,i} \cdot \mathrm{cov}^{\mathrm{tot}}_{j,j}}}. \tag{10.16}$$

Chapter 11

Cross Section Results

All results for the measured cross sections are discussed in this chapter. While the measured inclusive cross section is presented in Chapter 11.1, the measured normalised differential cross sections are presented in Chapter 11.2.1 for the top-quark and $t\bar{t}$ system quantities in the extrapolated parton level phase space and in Chapter 11.2.2 for the final state quantities in the visible particle level phase space.
Quantifying the agreement between different predictions and the measurement, a χ^2 test is presented in Chapter 11.2.3 for all cross section quantities and predictions. A discussion on the typical uncertainty values for all systematic uncertainty sources is given in Chapter 11.2.4.
A comparison with similar measurements and additional predictions is done in Chapter 11.2.5.
For completeness, the results are listed in detail in Appendix D and the complete covariance matrices and correlations can be found in Appendix E.

11.1 Inclusive Cross Section

Before studying differential cross sections, it is an important step to calculate the inclusive cross section to prove a good understanding of the data, the applied event selection and the modelled background processes. Additionally, the obtained result can serve as a cross-check of the published CMS results [194, 195].

11.1.1 Input and Result

All ingredients and the results of the inclusive cross section measurements for the electron and muon final states and the combined measurement are listed in Table 11.1. As defined in Chapter 8.2, the measurement is performed for the extrapolated total cross section and the visible cross section. Efficiency and acceptance are defined with respect to the visible particle level phase space. Therefore, the result for the visible cross section ($\sigma_{t\bar{t}\rightarrow X,\mathrm{vis}}$) corresponds to the phase space as defined in Chapter 8.1.2. All values are obtained for the full event selection after the Double Kinematic Fit event

reconstruction and the minimal χ^2-probability requirement.
Similar results are obtained for the muon and electron final state. The uncertainties are discussed in the next chapter.

	e+jets	μ+jets	e/μ+jets (combined)
N_{data}	26,843	24,927	51,770
$N^{MC}_{non-t\bar{t}\ BG}$	1,131	1,093	2,225
$t\bar{t}$ Signal Fraction $f_{t\bar{t}\ SG}$	0.924	0.924	0.924
Efficiency ϵ	0.206	0.198	0.202
Acceptance Acc	0.156	0.153	0.155
Luminosity $\mathscr{L}$ [fb^{-1}]	19.7	19.7	19.7
BR($t\bar{t} \rightarrow$X)	0.146	0.146	0.292
$\sigma_{t\bar{t}\rightarrow X,vis}$ [pb]	5.6	5.8	11.5 ± 1.6
$\sigma_{t\bar{t},tot}$ [pb]	257	252	254 ± 26

Table 11.1: All inputs and results of the inclusive $t\bar{t}$ cross section measurements. The visible inclusive cross section $\sigma_{t\bar{t}\rightarrow X,vis}$ includes the BR and is therefore defined specifically for each final state X=e+jets, μ+jets or e/μ+jets. The quoted error for the combined final state includes the statistical uncertainty and all systematic uncertainty sources.

11.1.2 Uncertainties

The value for all single uncertainty sources of the inclusive cross section measurement in the total and visible particle level phase space in the combined e/μ+jets final state is shown in Table 11.2.

The main uncertainties originate from hadronisation, the JES and the identification efficiency of b jets. Only uncertainties related to the $t\bar{t}$ modelling differ between the total and the visible phase space measurement because only these uncertainties affect the acceptance.

Beyond other inclusive cross section measurements of CMS [194, 195], also uncertainties related to the hadronisation and the W-boson BR are addressed. While the uncertainty related to the BR is small, a sizeable uncertainty is found for the hadronisation. Therefore, the uncertainties are presented once including all uncertainties and once without BR and hadronisation uncertainty for a direct comparison with other CMS results.

Moreover, a large difference between the hadronisation uncertainty for the total and the visible inclusive cross section measurement is observed. This difference is related to different predictions of the POWHEG+PYTHIA and MC@NLO+HERWIG MC event-generators which are used to address the hadronisation uncertainty. Efficiency, acceptance and $t\bar{t}$ signal fraction for different MC event-generators are listed in Table 11.3. The predicted acceptance differs for MADGRAPH+PYTHIA, POWHEG+PYTHIA

Uncertainty Type	**Obtained Uncertainty Value**	
	Total Inclusive Cross Section	Visible Particle Level Cross Section
Statistical	**0.5 %**	**0.5 %**
Total Systematic (excluding Hadronisation&W-boson BR)	**10.2 %** (7.9 %)	**14.0 %** (8.2 %)
Experimental Uncertainties		
Luminosity	2.7 %	2.7 %
PU	0.2 %	0.2 %
JES	5.0 %	5.0 %
JER	1.7 %	1.7 %
Trigger & lepton eff. SF (normalisation)	1.0 %	1.0 %
Trigger & lepton eff. SF (shape, η)	0.1 %	0.1 %
Trigger & lepton eff. SF (shape, p_T)	<0.1 %	<0.1 %
B-jet identification eff. SF (normalisation)	4.1 %	4.1 %
B-jet identification eff. SF (shape, η)	<0.1 %	<0.1 %
B-jet identification eff. SF (shape, p_T)	<0.1 %	<0.1 %
B-jet misidentification eff. SF	0.6 %	0.6 %
Model Uncertainties		
Single top-quark BG (normalisation)	0.7 %	0.7 %
Single top-quark BG (Q^2 scale)	0.2 %	0.2 %
VV BG (normalisation)	<0.1 %	<0.1 %
V+jets & $t\bar{t}$ +V BG (normalisation)	2.0 %	2.0 %
Q^2 scale (for $t\bar{t}$)	1.5 %	3.2 %
ME-PS matching threshold (for $t\bar{t}$)	0.9 %	0.3 %
m_{top} (for $t\bar{t}$)	1.4 %	0.1 %
W-boson BR	1.2 %	0.1 %
PDF	<0.1 %	<0.1 %
Hadronisation	6.4 %	11.3 %

Table 11.2: Uncertainties of the inclusive proton-proton $t\bar{t}$ production cross section measurement for the total and the visible particle level phase space. "V" is an abbreviation for a vector boson V$\in$ (W,Z,γ).

and MC@NLO+HERWIG. Furthermore, the predicted efficiency is equal for MADGRAPH+PYTHIA and POWHEG +PYTHIA, differing from the prediction of MC@NLO+HERWIG. Additionally, a different $t\bar{t}$ signal fraction is found for the MC@NLO+HERWIG prediction in contrast to the MADGRAPH+PYTHIA and POWHEG+PYTHIA prediction.

For the inclusive cross section in the visible phase space, the difference in efficiency matters. In contrast, the difference of the product of efficiency and acceptance matters for the total inclusive cross section. The hadronisation uncertainty is calculated comparing the result obtained with POWHEG+PYTHIA and MC@NLO+HERWIG. As the difference in the predicted efficiency is reduced by the difference in the predicted acceptance, the

	MadGraph +pythia	powheg +pythia	mc@nlo +herwig	Relative Difference
ϵ	0.202	0.198	0.227	+15%
Acc	0.155	0.147	0.139	–5%
$\epsilon \cdot$Acc	0.031	0.029	0.031	+7%
$f_{t\bar{t}\ SG}$	0.924	0.925	0.940	+2%

Table 11.3: Efficiency (ϵ), acceptance (Acc) and $t\bar{t}$ signal fraction ($f_{t\bar{t}\ SG}$) for the combined ℓ+jets final state as predicted by different MC event generator. Acceptance and efficiency are defined with respect to the visible particle level phase space. Data-driven SFs are included. The last column is the relative difference of the mc@nlo+herwig prediction with respect to the powheg+pythia prediction.

hadronisation uncertainty is lower for the extrapolated total inclusive cross section.

The applied data-driven SFs (see Chapter 6.7) are a possible origin for differences between the MC predictions with pythia and herwig, which finally contribute to the large value of the hadronisation uncertainty. All SFs are obtained using MC predictions where pythia is involved for the parton showering. In most cases, cross checks are performed using MC predictions with herwig for the parton showering. Nevertheless, these tests are done within specific phase spaces and event compositions which might differ from the event sample used for this analysis. Therefore, the obtained SFs might be not completely correct for MC predictions using herwig for the parton showering.

One topic recently investigated in this context is the efficiency for the misidentification of c-flavoured jets as b jets (ϵ_c^{tag}). As it is difficult to measure this efficiency directly in data [196], the to date description is to use the efficiency SFs for b jets with twice the quoted uncertainty. Triggered also by studies performed within this analysis [197], the differences between ϵ_c^{tag} are found to be large between MC predictions using pythia and herwig for the parton showering. Possible explanations are the different abundances of the c-flavoured hadrons between pythia and herwig [198], the different modelling of the fragmentation process or an ambiguous flavour identification procedure. This topic is studied further and a method to measure ϵ_c^{tag} in data is in development [199].

11.1.3 Comparison to Theory Predictions and Other Measurements

The result for the total inclusive cross section within this thesis is compared to different CMS measurements [194, 195] and theory predictions in Figure 11.1.

In contrast to this thesis, the CMS analyses used for comparison are optimised for the inclusive cross section measurement but use only a fraction of the complete $\sqrt{s} = 8$ TeV dataset recorded by CMS. Furthermore, the uncertainties related to the hadronisation and the W-boson BR are not included in the total uncertainty shown to allow a direct comparison with the other CMS results.

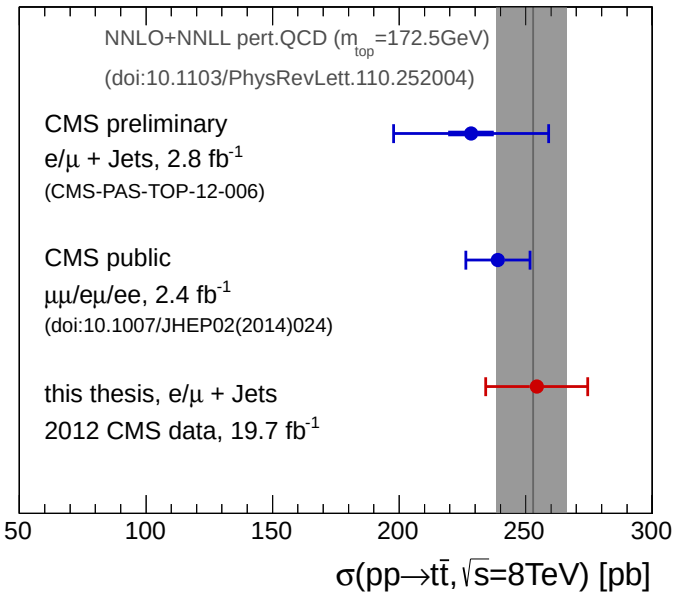

Figure 11.1: Comparison of the total inclusive proton-proton cross section for $t\bar{t}$ production as obtained in this thesis (red point) with the predicted value of the NNLO+NNLL perturbative QCD calculation [44] (grey band) and other published CMS results (blue points).

In Figure 11.1, the quoted theory prediction of

$$\sigma\left(pp \longrightarrow t\bar{t}, \sqrt{s} = 8 \text{ TeV}\right) = 252.89\ ^{+6.39}_{-8.64}\,(\text{scale})\ \pm 11.67(\alpha_S \oplus \text{PDF}) \qquad (11.1)$$

corresponds to the current most precise value at NNLO+NNLL perturbative QCD accuracy [44] obtained for a top-quark mass of $m_{top} = 172.5\,\text{GeV}$, a choice for factorisation and renormalisation scale of $\mu_F = \mu_R = m_{top}$ and the MSTW2008nnlo68cl PDF set using [200]. The PDF and α_S uncertainties are evaluated following interim recommendations of the PDF4LHC Working Group [201]. These choices and uncertainties will be recommended by the TopLHCWG to be used for published results by ATLAS and CMS.

A good consistency of the measured value with both experimental results and the theoretical prediction is observed within uncertainties while the measured value in this thesis is slightly higher than other CMS results. A similar agreement is found between the measured visible inclusive cross section and the prediction obtained from the quoted theory cross section, the PDG BR [32] and the acceptance predicted by MadGraph+pythia.
The precision achieved is better than for the preliminary ℓ+jets CMS result but lower than the CMS result of the dileptonic final state. The dileptonic final state has the advantage of an intrinsically lower background rate and smaller JES uncertainties as only two jets appear in the final state at LO. For this thesis, a tighter event selection is applied to overcome the higher background rate of the ℓ+jets final state, resulting in a

larger uncertainty due to the applied b-jet identification.
The precision of the theory prediction is similar to the most precise CMS measurement in the dileptonic final state.

The measured inclusive cross section is a valid cross check for the published results and proves the good understanding of the analysed $\mathrm{t\bar{t}}$ dataset which is need to derive normalised differential cross sections.

11.2 Normalised Differential Cross Sections

Normalised differential cross sections $\frac{1}{\sigma}\frac{\mathrm{d}\sigma}{\mathrm{dX}}$ are obtained following the cross section calculation explained in Chapter 8.3 and the unfolding procedure detailed in Chapter 9.2.

The results for the separate final states are found to be in good agreement with each other and with the combined result. Therefore, only the combined ℓ+jets measurement is presented.

All obtained results are compared to the MC predictions of MadGraph+pythia (CTEQ6L1), mc@nlo+herwig (CTEQ6M), powheg+pythia (CT10) and powheg +herwig (CT10). For further details see also Table 5.1 in Chapter 5.3.2. All quoted average and maximum values for the measured quantities refer to the MadGraph+pythia prediction. Wherever possible, the results are in addition compared to higher order differential perturbative QCD predictions. These predictions are of NLO+NNLL precision [202–204] for $m^{t\bar{t}}$ and $p_T^{t\bar{t}}$ and of approx. NNLO precision [205] for p_T^t and y^t. All predictions are binned and normalised using the same procedure as for the data measurement. The simulated number of events is high enough to neglect corresponding statistical uncertainties. Therefore, all differences between MC predictions are expected to originate from different modelling approaches (choices of factorisation and renormalisation scale, choices of the PDF, precision of the ME calculation, PS modelling, hadronisation model, etc.). For a better readability, no uncertainty bands for the MC predictions are included in the final results. As an approximation, the statistical uncertainty of the MC predictions is a factor of two to three smaller than the statistical uncertainty of the data. Furthermore, also the size of the systematic uncertainty is expected to be roughly of the same size as for the data measurement which is dominated by $t\bar{t}$ modelling uncertainties. This assumption is e.g. found to be approximately true in [23], where conservative uncertainty bands are included for the mc@nlo+herwig MC prediction. Additionally, the choice of different model parameters in the MadGraph+pythia prediction is studied in Appendix G.

For a better comparison of the compatibility between all predictions and the obtained result in data, the ratio with respect to the measured value in data is shown for all results. The statistical and systematic uncertainty of the measurement are shown for each cross section quantity as band in the ratio.

Overall, the SM predictions can describe the data reasonably well, while some differences are observed for particular distributions and predictions. All single measurements are discussed in detail for the top-quark and $t\bar{t}$ system quantities measured in the extrapolated parton level phase space in Chapter 11.2.1 and for all final state quantities measured in the visible particle level phase space in Chapter 11.2.2. An extended discussion about the physical meaning for most of the quantities can be found in Chapter 3.1.3 and [23].

11.2.1 Top-Quark and $t\bar{t}$ System Quantities

All results obtained in the extrapolated parton level phase space, which is defined in Chapter 8.1.1, are shown in Figure 11.2 for the $t\bar{t}$ system quantities and in Figure 11.3 for the top-quark quantities.

The transverse momentum of the $t\bar{t}$ system ($p_T^{t\bar{t}}$, Figure 11.2a) is sensitive to higher order effects as it balances the transverse momentum of all additionally radiated partons. In the binning of the analysis, the distribution is falling steeply. The expected average value is $\langle p_T^{t\bar{t}} \rangle \approx 55\,\text{GeV}$ and the measurement ranges up to $p_T^{t\bar{t}} = 300\,\text{GeV}$, limited by the resolution and large uncertainties for boosted topologies at large values of $p_T^{t\bar{t}}$.
All MC predictions are similar to each other and describe the measured $p_T^{t\bar{t}}$ distribution in data with uncertainties of typically $5-8\%$. For $p_T^{t\bar{t}} < 80\,\text{GeV}$ a trend towards larger values of $p_T^{t\bar{t}}$ in data is observed.
Furthermore, the NLO+NNLL perturbative QCD calculation [204] predicts a much softer spectrum and cannot describe the data distribution.

The invariant mass of the $t\bar{t}$ system ($m^{t\bar{t}}$, Figure 11.2b) reflects the centre-of-mass energy of the underlying partonic collision $\sqrt{\hat{s}}$. Furthermore, $m^{t\bar{t}}$ is sensitive to BSM effects, e.g. the production of a heavy particle decaying into $t\bar{t}$ final states, which would show up as peak in the $m^{t\bar{t}}$ distribution.
The measurement ranges from $m^{t\bar{t}} = 345\,\text{GeV}$, which equals the minimum of $m_{\text{min}}^{t\bar{t}} \approx 2 \cdot m_{\text{top}}$ where the top quarks are produced at rest, to $m^{t\bar{t}} = 1.6\,\text{TeV}$, where the top quarks are already highly boosted. In the binning of the analysis, the distribution is steeply falling. The expected average value is $\langle m^{t\bar{t}} \rangle \approx 490\,\text{GeV}$.
All predictions are found to describe the $m^{t\bar{t}}$ distribution in data reasonably well within the achieved precision of typically $4-10\%$. The MadGraph+pythia and powheg+herwig event generators deliver a slightly better prediction than the mc@nlo+herwig and powheg+pythia event generators, which predict on average larger values for $m^{t\bar{t}}$.
Moreover, the NLO+NNLL perturbative QCD prediction [203] is similar to the predictions of mc@nlo+herwig and powheg+pythia and describes therefore the data also within one to two standard deviations.

The rapidity of the $t\bar{t}$ system ($y^{t\bar{t}}$, Figure 11.2c) is a measure for the boost of the $t\bar{t}$ system along the beam axis. For the LO $t\bar{t}$ production process, it is determined by the momentum asymmetry of the interacting initial state partons.
A symmetric distribution around the maximum at $y^{t\bar{t}} = 0$, which falls off for larger values of $|y^t|$, is obtained. The expected average value of its magnitude is $\langle |y^{t\bar{t}}| \rangle \approx 0.7$ and the maximum of the measurement is $|y^{t\bar{t}}| \leq 2.5$, limited by the low number of events for high values of $|y^{t\bar{t}}|$ and the $|\eta|$-acceptance cuts of the lepton and b-jet selection.
All predictions are found to describe the $y^{t\bar{t}}$ distribution in data reasonably well with typical uncertainties of $2-5\%$. The MadGraph+pythia event generator provides a slightly better description of the data than the mc@nlo+herwig, powheg+pythia and powheg+herwig event generators, which predict a slightly less central distribu-

tion.

The transverse momentum of the top quarks (p_T^t, Figure 11.3a) is measured in the detector rest frame. Therefore, the measured value reflects the momentum of the produced top-quarks in the $t\bar{t}$ rest frame convoluted with the p_T of the $t\bar{t}$ system. The transverse top-quark momentum in the $t\bar{t}$ rest frame depends on the modelling of ISR and FSR and is discussed in more detail in Chapter 3.1.3. The p_T^t distribution rises steeply from $p_T^t = 0\,\text{GeV}$ to a maximum at $p_T^t = 80\,\text{GeV}$ and is then steeply falling with a tail to higher values. The maximum of the measured range is $p_T^t = 500\,\text{GeV}$, while the expected average value is $\langle p_T^t \rangle \approx 120\,\text{GeV}$.
A softer spectrum is observed in data than predicted by most of the MC generators. The MadGraph+pythia, powheg+pythia and mc@nlo+herwig predictions agree with each other for $p_T^t < 200\,\text{GeV}$. For $p_T^t > 200/250\,\text{GeV}$ the mc@nlo+herwig prediction is found to agree slightly better with data while the agreement of the MadGraph+pythia prediction is slightly worse. In contrast, the powheg+herwig prediction is in good agreement with data. The overall softer shape for p_T^t in data with respect to the MadGraph+pythia, powheg+pythia and mc@nlo+herwig predictions is not covered by the uncertainties of the measurement of typically $3-8\%$. In the first bin, the difference is about 10% while the uncertainty of the measurement is about 4%.
The approx. NNLO perturbative QCD calculation [205] provides a good description of the data for p_T^t. Especially for $p_T^t < 300\,\text{GeV}$, the prediction is in better agreement with data than the MadGraph+pythia, powheg+pythia and mc@nlo+herwig MC generators.

To directly access the transverse momentum of the top quark, independently from the modelling of the initial boost of the $t\bar{t}$ system, the transverse momentum of the top quark is also measured in the rest frame of the $t\bar{t}$ system (p_T^t ($t\bar{t}$ com), Figure 11.3b). The general shape of the distribution is almost equal to p_T^t obtained in the detector rest frame. Only the expected average value of $\langle p_T^t\ (t\bar{t}\ \text{com}) \rangle \approx 110\,\text{GeV}$ is slightly lower because the boost of the $t\bar{t}$ system is not included.
The agreement with the MC predictions is similar as discussed for p_T^t. Again, a softer spectrum is observed in data in comparison to the MadGraph+pythia, powheg+pythia and mc@nlo+herwig predictions while the powheg+herwig prediction is in good agreement with data. However, the uncertainties of the measurement are slightly larger. For p_T^t ($t\bar{t}$ com) $> 300\,\text{GeV}$, the prediction of MadGraph+pythia differs even more from data and other predictions. Moreover, no higher order QCD prediction is available for p_T^t ($t\bar{t}$ com).

The normalised differential cross section is separately measured for the top quark with the higher ($p_T^{\text{lead t}}$, Figure 11.3c) and the top quark with the lower transverse momentum ($p_T^{\text{sublead t}}$, Figure 11.3d). This separates the top quarks by the amount of additional radiation and gives sensitivity to potential higher order effects that are expected to be more pronounced for $p_T^{\text{sublead t}}$. The general shape of the distributions is similar to the combined p_T^t measurement and only the maximum is shifted. The expected average values are $\langle p_T^{\text{lead t}} \rangle \approx 135\,\text{GeV}$ and $\langle p_T^{\text{sublead t}} \rangle \approx 100\,\text{GeV}$.
For both measurements the conclusion is similar as for the combined p_T^t. Both spectra

are softer in data than predicted by the MadGraph+pythia, powheg+pythia and mc@nlo+herwig event generators and the observed difference is not covered by the uncertainties of the measurement. The effect is more pronounced for the top quark with lower momentum. Again, the powheg+herwig prediction describes data in both cases. The MadGraph+pythia, powheg+pythia and mc@nlo+herwig predictions are similar for $p_T^{\text{lead t}}/p_T^{\text{sublead t}} < 200/250\,\text{GeV}$ while differences are observed for higher values. mc@nlo+herwig provides a slightly better description of the data while the agreement between data and the MadGraph+pythia prediction is worse. It should be noted that the powheg+pythia prediction follows MadGraph+pythia for $p_T^{\text{lead t}}$ while it follows mc@nlo+herwig for $p_T^{\text{sublead t}}$.
An extended discussion of all measurements related to the transverse momentum of the top quarks together with additional studies is given in Chapter 12.

The rapidity of the top quarks (y^t, Figure 11.4a) is a convolution of the rapidity of the top quarks in the $t\bar{t}$ rest frame with the rapidity of the $t\bar{t}$ system. Therefore, y^t is slightly broader than $y^{t\bar{t}}$.
The distribution is symmetric around the maximum at $y^t = 0$ and falls off towards larger values of $|y^t|$. The expected average value is $\langle|y^t|\rangle \approx 0.8$ and the maximum of the measurement is $|y^t| \leq 2.5$, limited by the same arguments as $y^{t\bar{t}}$.
The rapidity of the top quarks in data is well described by all predictions (including the approx. NNLO perturbative QCD prediction [205]) within the precision of the measurement of typically $2-4\%$. Within the expected precision, all predictions are found to agree, while the MadGraph+pythia and approx. NNLO predictions are slightly more central than the mc@nlo+herwig, powheg+pythia and powheg+herwig predictions.

The difference in azimuthal angle of the two top quarks ($\Delta\phi\,(t,\bar{t})$) is shown in Figure 11.4b). In the LO $t\bar{t}$ production process, the top quarks are expected to be back-to-back in ϕ. Therefore, values $\Delta\phi\,(t,\bar{t}) < \pi$ are a consequence of additionally produced partons in higher order processes. Furthermore, processes beyond the SM can affect the $\Delta\phi\,(t,\bar{t})$ distribution as detailed in [62].
The distribution has a maximum at $\Delta\phi\,(t,\bar{t}) = \pi$ and falls steeply for smaller values with a tail down to $\Delta\phi\,(t,\bar{t}) = 0$. The expected average value is $\langle\Delta\phi\,(t,\bar{t})\rangle \approx 2.7$.
The observed result in data is described by all MC predictions within two standard deviations of the uncertainty of the measurement, which is typically $4-7\%$. All MC predictions agree with each other. A small trend towards smaller values of $\Delta\phi\,(t,\bar{t})$ in data is observed. This is consistent with the observed trend in data towards larger values of $p_T^{t\bar{t}}$ for $p_T^{t\bar{t}} < 80\,\text{GeV}$ because larger values for $p_T^{t\bar{t}}$ and smaller values for $\Delta\phi\,(t,\bar{t})$ can both be explained by more radiation.

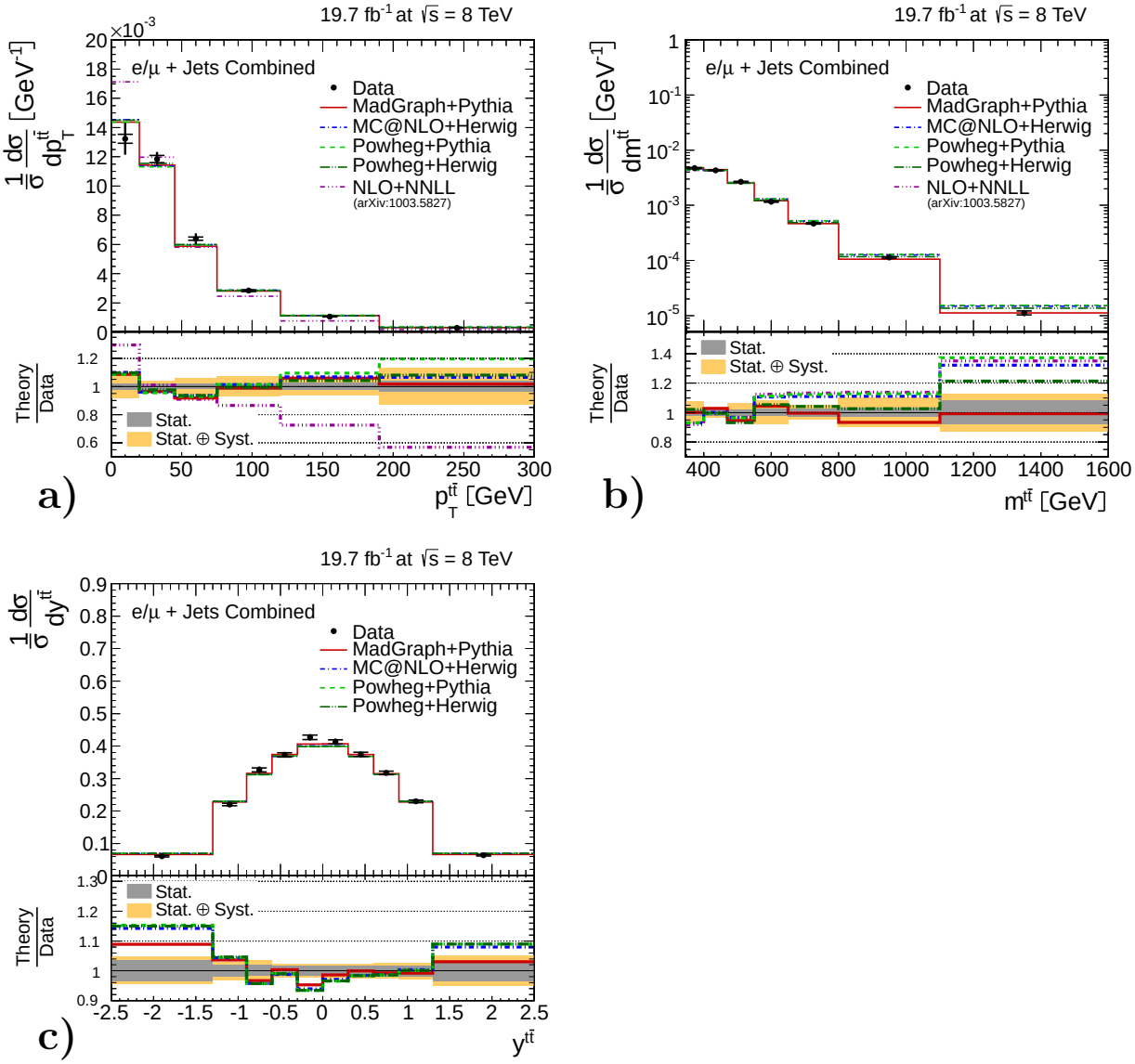

Figure 11.2: Normalised differential cross section for the transverse momentum (a), invariant mass (b) and rapidity (c) of the $t\bar{t}$ system in the extrapolated parton level phase space. The inner error bars correspond to the statistical uncertainty and the outer error bars to the combined statistical and systematic uncertainty of the measurement. Furthermore, the grey (orange) band in the ratio corresponds to the statistical (combined statistical and systematic) uncertainty of the data measurement.

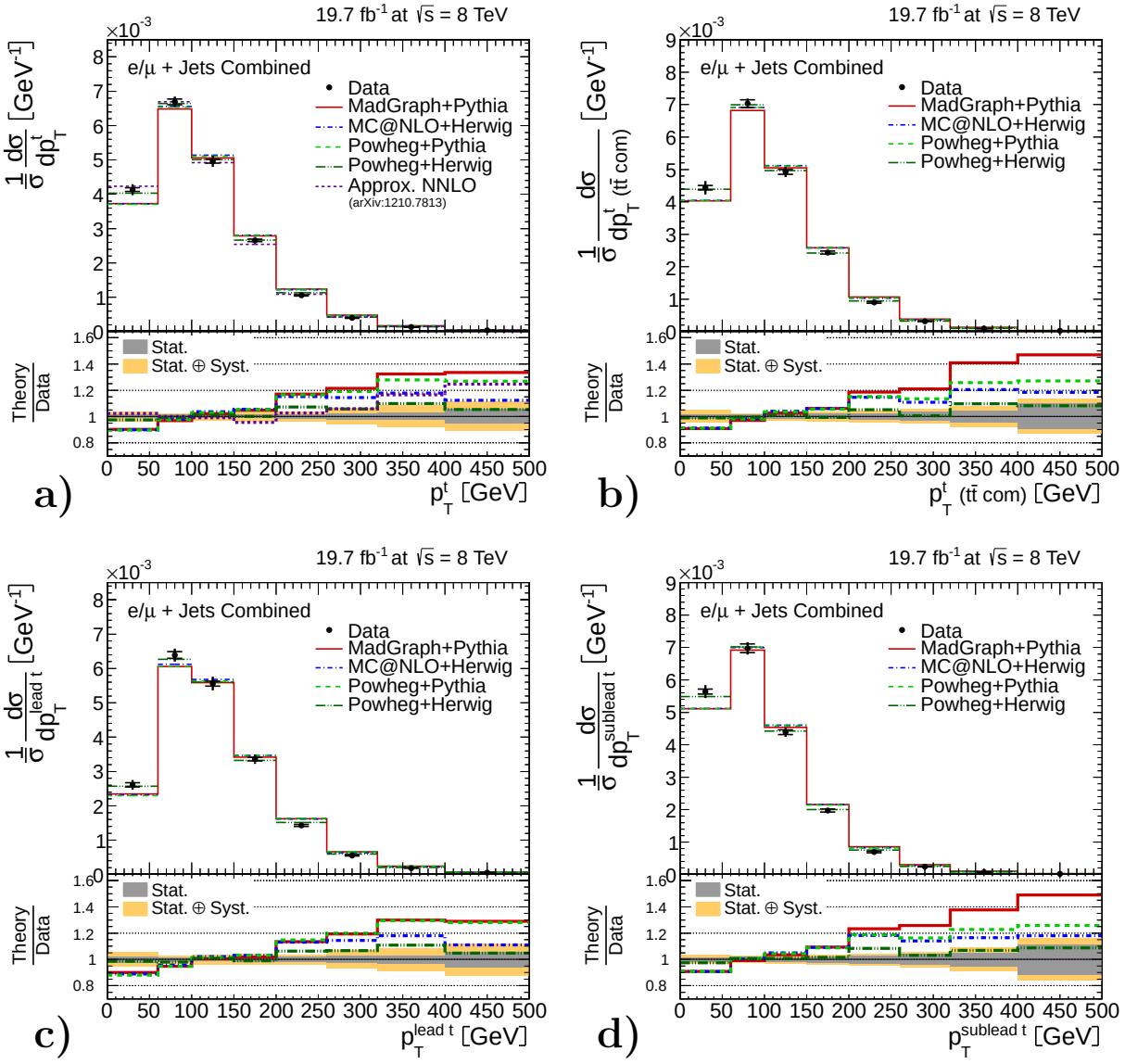

Figure 11.3: Normalised differential cross section for the transverse momentum of both top quarks in the detector (a) and $t\bar{t}$ rest frame (b), the transverse momentum of the top quark with the higher (c) and lower (d) p_T in the detector rest frame. in the extrapolated parton level phase space. The inner error bars correspond to the statistical uncertainty and the outer error bars to the combined statistical and systematic uncertainty of the measurement. Furthermore, the grey (orange) band in the ratio corresponds to the statistical (combined statistical and systematic) uncertainty of the data measurement.

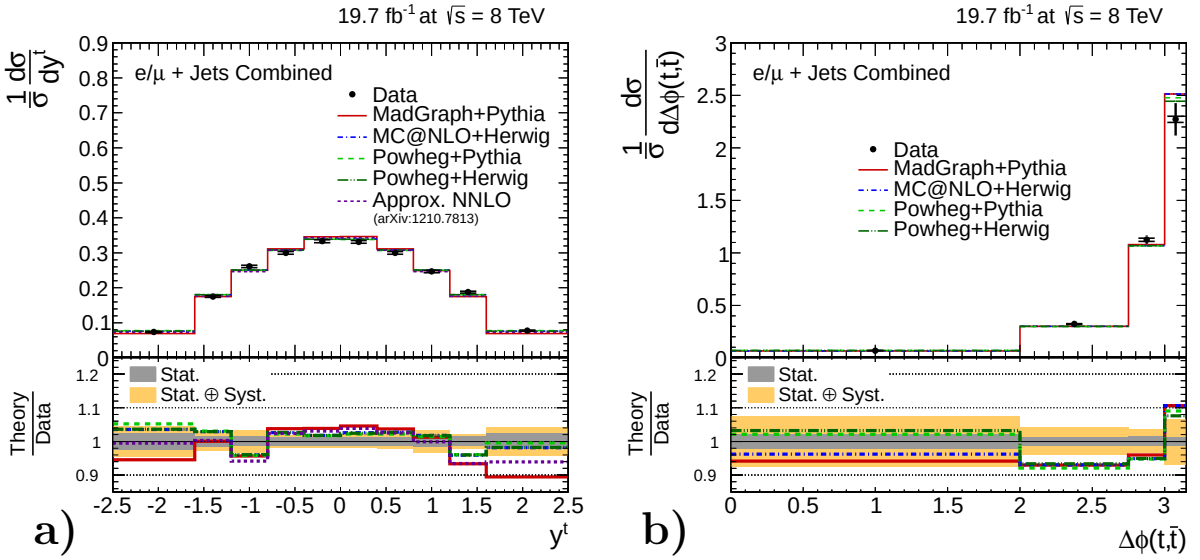

Figure 11.4: Normalised differential cross section for the rapidity of both top quarks (a) and their difference in azimuthal angle (b) in the extrapolated parton level phase space. The inner error bars correspond to the statistical uncertainty and the outer error bars to the combined statistical and systematic uncertainty of the measurement. Furthermore, the grey (orange) band in the ratio corresponds to the statistical (combined statistical and systematic) uncertainty of the data measurement.

11.2.2 Final State Quantities

All results obtained in the visible particle level phase space as defined in Chapter 8.1.2 are shown in Figure 11.5 for the lepton and the single b jets and in Figure 11.6 for the $\mathrm{b\bar{b}}$ system, the jet multiplicity, the invariant mass of lepton and leptonic b jet and ρ_{S}. The results for specific measurements of the visible particle level phase space cannot be considered as independent because they all involve the modelling of the lepton and b-jet kinematics ($\mathrm{p_T}$, η) due to the phase space definition. Nevertheless, they are an important test for the consistency of the observed results and can help to test different approximations that are used in different MC event generators.

The transverse momenta of the lepton ($\mathrm{p_T^l}$, Figure 11.5a) and the b jets ($\mathrm{p_T^b}$, Figure 11.5c) reflect the momentum of the initial top quarks (in addition to the properties of the top-quark decay). Therefore, the obtained shapes are similar to $\mathrm{p_T^t}$ but shifted to lower values. On average, the momentum of the lepton is lower than the momentum of the b jets, which are resulting directly from the top-quark decay. Furthermore, the restriction to the visible phase space results in a harder spectrum for the normalised $\mathrm{p_T^b}$ and $\mathrm{p_T^l}$ distributions.
Consistent with the measured transverse momentum of the top-quarks the measured momentum distributions of the final state lepton and b jets are found to be also slightly softer in data than predicted by the MadGraph+pythia, powheg+pythia and mc@nlo+herwig event generators. The powheg+herwig prediction describes data best. The observed agreement between data and the MC predictions is similar to $\mathrm{p_T^t}$ while the significance of the trend to lower values is reduced. This can be e.g. seen when comparing the χ^2/ndof values for the different quantities as discussed in Chapter 11.2.3. The MadGraph+pythia and powheg+pythia MC predictions agree while mc@nlo+herwig and powheg+herwig predict slightly softer distributions for $\mathrm{p_T^l}$ and provide therefore an improved description of data. The typical precision of the measurements is $2-7\%$ for $\mathrm{p_T^l}$ and $\mathrm{p_T^b}$ for the chosen binning.

Similar to the transverse momentum, also the pseudo rapidity of the final state lepton (η^{l}, Figure 11.5b) and b jets (η^b, Figure 11.5d) reflect the underlying distribution of the top quarks.
Consequently, a similar data-to-MC trend as for $\mathrm{y^t}$ is observed. All predictions describe the data within the precision of the measurements while the powheg+pythia, mc@nlo+herwig and powheg+herwig predictions are slightly less central, describing the observed shape in data slightly better. The achieved precision is typically $2-4\%$ for η^{l} and $2-3\%$ for η^b for the chosen binning.

Furthermore, the invariant mass ($\mathrm{m^{b\bar{b}}}$, Figure 11.6a) and transverse momentum ($\mathrm{p_T^{b\bar{b}}}$, Figure 11.6b) of the $\mathrm{b\bar{b}}$ system are determined for the two b jets originating from the decay of the top-quarks. The correct modelling of the $\mathrm{b\bar{b}}$ pair resulting from the decay of a top-quark pair is e.g. relevant for a precise measurement of the $\mathrm{H \to b\bar{b}}$ final state. Both measured distributions in data are described by all predictions within the precision of typically $4-15\%$ for $\mathrm{m^{b\bar{b}}}$ and $\mathrm{p_T^{b\bar{b}}}$. All predictions agree for $\mathrm{p_T^{b\bar{b}}} < 150\,\mathrm{GeV}$. For $\mathrm{p_T^{b\bar{b}}} > 150\,\mathrm{GeV}$ mc@nlo+herwig provides the best description of data and the

differences between the different predictions is up to 50%. However, the uncertainties of the measurement are of the same size as the differences between the predictions. For $m^{b\bar{b}}$, the MC predictions are more similar and only a small difference is observed for $m^{b\bar{b}} > 400\,\text{GeV}$ where MadGraph+pythia and powheg+herwig provide a slightly better description because the predicted spectrum is somewhat softer. However, the precision of the measurement at large values of $m^{b\bar{b}}$ does not allow for a reliable conclusion.

The invariant mass of the lepton and the b jet originating from the leptonically decaying top quark (m^{lb}, Figure 11.6c) is sensitive to the top-quark mass.
All predictions are in good agreement with each other and describe the measured spectrum in data within the obtained uncertainty of typically $5-10\%$. A trend towards smaller values of m^{lb} is observed in data.

The jet multiplicity (N_{jets}, Figure 11.7a) is a measure of higher order contributions. For LO $t\bar{t}$ production, four jets are expected in the ℓ+jets final states. In a simplified picture, the production of additional jets is suppressed by a factor of α_S per jet. This would lead to a spectrum which drops roughly by an order of magnitude for each additional jet produced. On the other hand, higher order interferences, the effect of the phase space definition (especially of the jet acceptance) and the possible clustering of several final state quarks in the same jet result in a spectrum falling off less steeply. In each presented MC prediction, low jet multiplicities are expected to be modelled predominantly by the ME calculation while high jet multiplicities are expected to be predominantly modelled by the parton showering.
For data and prediction about half of the events contain at least one additional jet (i.e. $N_{jets} > 4$) with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$. Good agreement is observed between data and all predictions for $N_{jets} \leq 5$. Especially at higher jet multiplicities, the mc@nlo+herwig event generator predicts a lower jet multiplicity than observed in data while the MadGraph+pythia, powheg+pythia and powheg+herwig predictions describe the data well for $N_{jets} \leq 7$ within a typical precision of $12-15\%$. For even higher values of N_{jets}, the MC predictions differ by up to a factor of three but all agree within two standard deviations with the measurement. The observed qualitative agreement is consistent with results of the CMS [206, 207] and ATLAS [208] collaborations. The numerical values cannot be directly compared as different phase space definitions are used.

Finally, ρ_S as defined in Equation 3.16 is shown in Figure 11.7b. This quantity is expected to be very sensitive to the top-quark mass [49] and depends on higher order processes because of the additional jet which is involved in the definition.
All MC predictions, which are consistently produced with $m_{top} = 172.5\,\text{GeV}$, are in good agreement with each other and the data. However, the observed uncertainties of typically $15-30\%$ are large due to the modelling of the extra jet (with $p_T > 50\,\text{GeV}$ and $|\eta| < 2.4$). Nevertheless, the extraction of the top-quark mass from this normalised differential distribution is discussed in detail in Chapter 13.

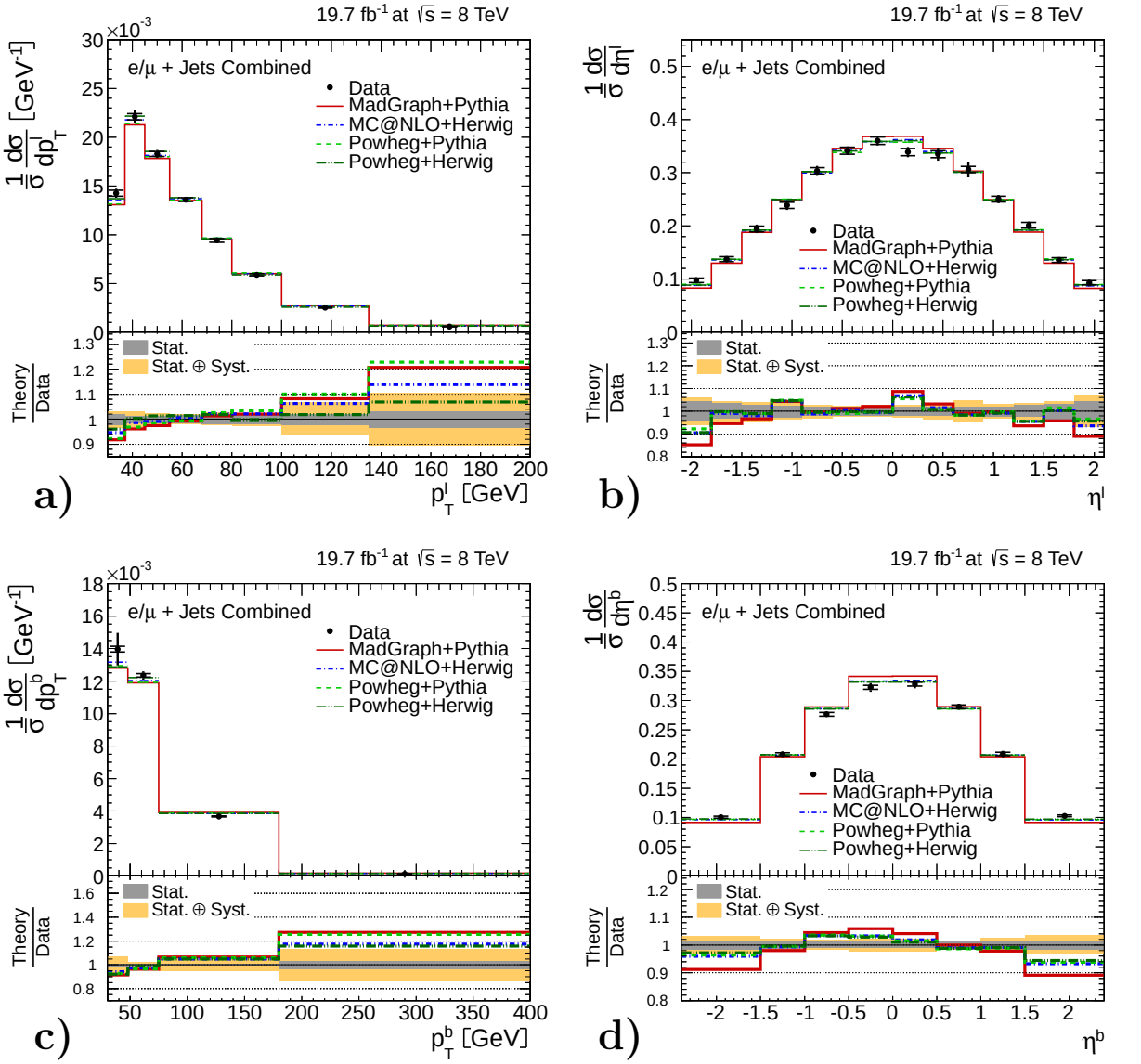

Figure 11.5: Normalised differential cross section for the transverse momentum (a,c) and pseudorapdity (b,d) of the $t\bar{t}$ final state lepton (a,b) and b jets (c,d) in the visible particle level phase space. The grey (orange) band in the ratio corresponds to the statistical (combined statistical and systematic) uncertainty of the data measurement.

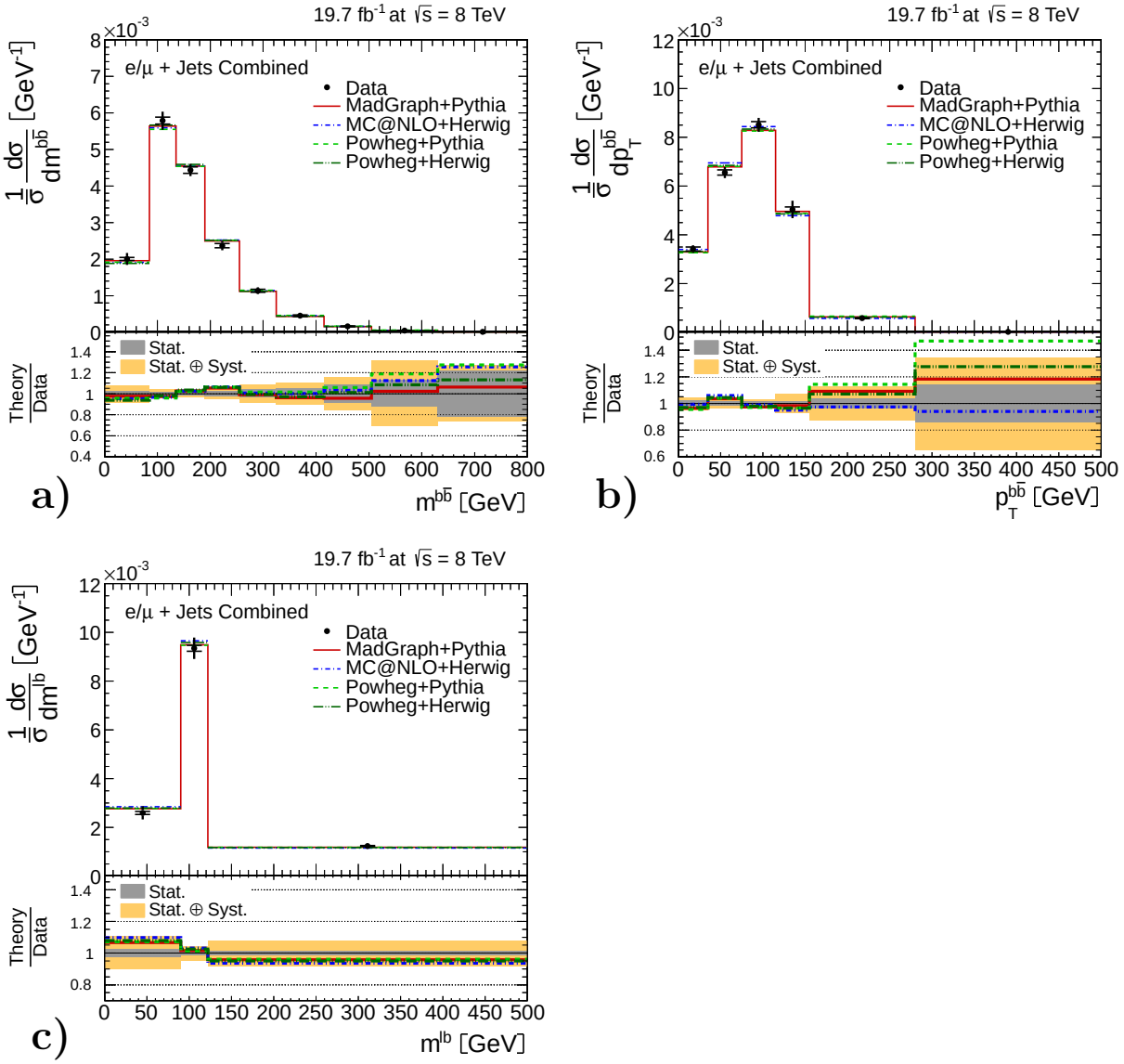

Figure 11.6: Normalised differential cross section for the invariant mass (a) and transverse momentum (b) of the $b\bar{b}$ system and the invariant mass of the lepton and the leptonic b jet (c) in the visible particle level phase space. The grey (orange) band in the ratio corresponds to the statistical (combined statistical and systematic) uncertainty of the data measurement. The lepton and the b jets are required to originate directly from the $t\bar{t}$ decay.

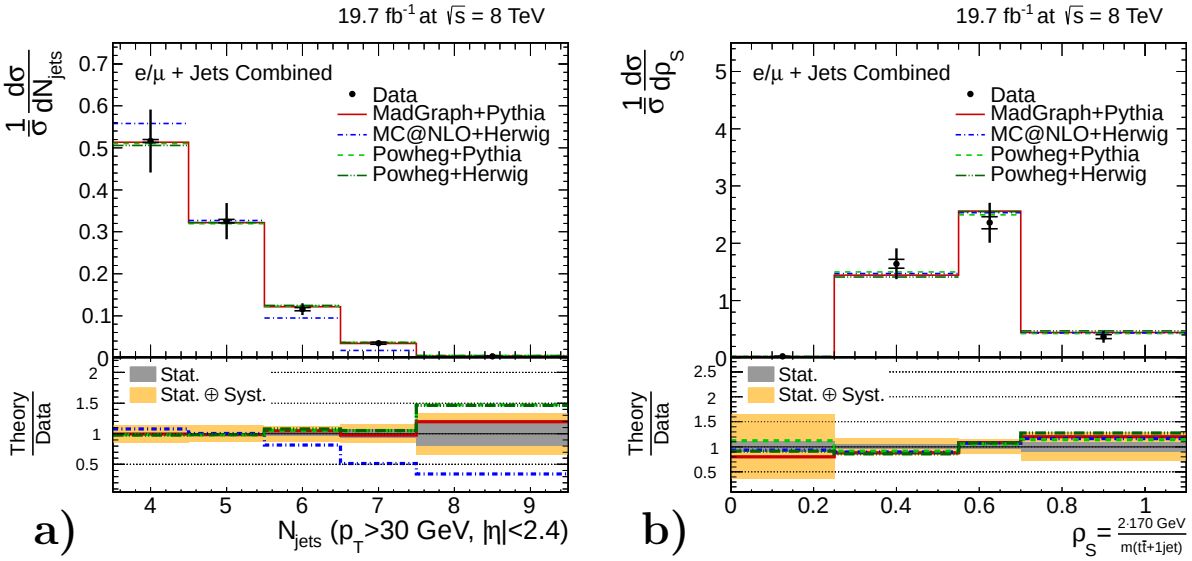

Figure 11.7: Normalised differential cross section for the jet multiplicity (a, for $N_{jets} \geq 4$ due to the event selection) and ρ_S(b) in the visible particle level phase space. The grey (orange) band in the ratio corresponds to the statistical (combined statistical and systematic) uncertainty of the data measurement.

11.2.3 χ^2 Comparison Between Data and Different Predictions

To estimate the compatibility between data and the different theory predictions, a χ^2/ndof expression is calculated for each measured normalised differential cross section ($\mathrm{diff_i}$):

$$\frac{\chi^2_{\mathrm{theory}}}{\mathrm{ndof}} = \frac{1}{\mathrm{N_{bins}} - 1} \sum_{\mathrm{bin\ i=1}}^{\mathrm{N_{bins}}} \left(\frac{\mathrm{diff_i^{data}} - \mathrm{diff_i^{theory}}}{\Delta \mathrm{diff_i^{data}}} \right)^2 . \tag{11.2}$$

In this approach, the uncertainties of the data measurement ($\Delta\mathrm{diff_i^{data}}$) are considered but note that all correlations between bins are neglected in this estimate. The degrees of freedom are reduced by one due to the normalisation.

The results can be found in Table 11.4 for the extrapolated parton level phase space and in Table 11.5 for the visible particle level phase space.

quantity	$\frac{\chi^2_{\mathrm{theory}}}{\mathrm{ndof}}$				
	MadGraph + pythia	mc@nlo + herwig	powheg + pythia	powheg + herwig	pert.QCD
p_T^t	10.13	5.83	8.59	1.07	1.79
$p_T^{\mathrm{lead\ t}}$	6.26	3.97	6.90	0.84	-
$p_T^{\mathrm{sublead\ t}}$	12.56	6.11	7.02	0.82	-
p_T^t ($t\bar{t}$ com)	11.67	4.58	5.84	0.50	-
y^t	3.96	1.49	1.35	1.44	2.64
$\Delta\phi\,(t,\bar{t})$	2.46	2.58	2.61	2.06	-
$m^{t\bar{t}}$	0.34	2.57	3.25	0.88	3.44
$p_T^{t\bar{t}}$	0.95	1.03	1.69	0.76	9.90
$y^{t\bar{t}}$	1.26	2.79	3.26	3.21	-
average	5.51	3.44	4.50	1.29	-

Table 11.4: χ^2/ndof expressions for all normalised differential cross sections measured in the extrapolated parton level phase space for all investigated predictions. The last column corresponds to a perturbative QCD calculation of NLO+NNLL precision [202] for $m^{t\bar{t}}$ [203] and $p_T^{t\bar{t}}$ [204] and approx. NNLO precision for p_T^t and y^t [205].

The obtained $\frac{\chi^2_{\mathrm{theory}}}{\mathrm{ndof}}$ values quantify the agreement of the different predictions with the data, which is qualitatively discussed in Chapter 11.2.1 and 11.2.2. The lower the value, the better data is described by the respective theory.

For most of the cross section quantities and predictions values in the order of one are obtained, indicating a good description of the data. The powheg+herwig event generator delivers the best overall description of the data with an average $\frac{\chi^2_{\mathrm{theory}}}{\mathrm{ndof}}$ value of about one. For the MadGraph+pythia, powheg+pythia and mc@nlo+herwig predictions, a similar low average value of one to two is found for the results of the

quantity	$\frac{\chi^2_{\text{theory}}}{\text{ndof}}$			
	MadGraph + pythia	mc@nlo + herwig	powheg + pythia	powheg + herwig
p_T^l	2.24	1.04	2.46	0.48
η^l	2.12	1.01	0.72	0.87
p_T^b	3.55	1.64	2.99	1.28
η^b	5.12	1.85	1.77	1.36
$p_T^{b\bar{b}}$	0.61	0.74	1.25	0.73
$m^{b\bar{b}}$	0.27	0.51	0.57	0.43
m^{lb}	0.40	1.00	0.38	0.63
N_{jets}	0.13	4.16	0.58	0.61
ρ_S	0.49	0.33	0.25	0.67
average	1.66	1.36	1.22	0.78

Table 11.5: χ^2/ndof expressions for all normalised differential cross sections measured in the visible particle level phase space for all investigated predictions.

visible particle level phase space. The largest values of four to twelve are obtained for the MadGraph+pythia, powheg+pythia and mc@nlo+herwig predictions in all measurements related to the transverse momentum of the top quark and for the NLO+NNLL prediction for $p_T^{t\bar{t}}$. Furthermore, a large value (> 4) is also obtained for the mc@nlo+herwig prediction for N_{jets} and the MadGraph+pythia prediction for η^b.

An alternative approach including all correlations between the bins of the measurement in the χ^2-expression is discussed in Appendix I. This alternative approach results in qualitatively similar conclusions, especially for the measurements related to the transverse momentum of the top quarks. In addition, the difference between data and the MadGraph+pythia prediction for the normalised differential cross sections relative to the uncertainties of the measurement and the uncertainties of the prediction is shown for separate bins in Appendix H.

11.2.4 Typical Values and Dominant Sources of the Systematic Uncertainties

Precise values for the total systematic uncertainties for all cross section quantities and bins are listed in Appendix D. For a better understanding of the results and the limitation of the measurement, the uncertainty sources with the largest contributions are identified.
The typical value for a specific uncertainty source depends on the actual distribution and bin under consideration. Especially the normalisation results in an uncertainty which is often vanishing at a specific turning point of the spectrum and becomes larger for more distant bins. Therefore, a median is calculated for each uncertainty source and cross section quantity. Here, the median value is defined by the central bin when ordering the uncertainty contributions of a given source and cross section quantity for all bins by size. In case of two central bins, the median is calculated from the average of the two values. This definition is robust against outliers while the actual value is a good representative for an average uncertainty value.

Summarising the typical uncertainties for all eighteen cross section quantities, the minimum, maximum and average median for all uncertainty sources are listed in Table 11.6 for the results in the extrapolated parton level phase space and in Table 11.7 for the results in the visible particle level phase space. They represent a typical order for a certain systematic uncertainty for the combined normalised differential cross sections.

For individual bins or even distributions the uncertainties can be higher or lower. For example the JES is a relevant uncertainty for all momentum-related quantities (e.g. p_T^t) while it is almost negligible for angle-related quantities (e.g. y^t).
Nevertheless, the table indicates that the $t\bar{t}$ model uncertainties (especially hadronisation, m_{top}, Q^2 and ME-PS matching threshold) and the JES are the most relevant uncertainties for normalised differential cross sections while all uncertainties not depending on the shape of a distribution, e.g. the luminosity or any kind of normalisation uncertainty, almost cancel. This is already the case for the CMS differential cross section analysis at $\sqrt{s} = 7$ TeV but in comparison to previous results [23, 26], the total size of the uncertainties is slightly reduced.

Larger uncertainties are obtained for the cross section quantities measured in the visible particle level phase space than for the the quantities measured in the extrapolated parton level phase space. One reason are additional migration effects from outside the defined phase space into the reconstructed event sample. Furthermore, some quantities like ρ_S or larger jet multiplicities are very challenging because their modelling depends a lot on higher order processes even beyond $t\bar{t}$ production at NLO QCD precision, which leads e.g. to larger $t\bar{t}$ modelling uncertainties. Furthermore, the analysis is optimised for the top-quark and $t\bar{t}$ quantities as e.g. parton level resolutions and constraints are used for the event reconstruction (see Chapter 7).

Experimental Uncertainties

Uncertainty Source	**average**	**minimum**	**maximum**
	(median of the relative uncertainty)		
Luminosity	<0.1%	<0.1%	0.1%
PU	0.3%	0.1%	0.9%
JES	1.6%	0.6%	3.7%
JER	0.5%	0.2%	1.1%
Trigger & lepton eff. SF (normalisation)	<0.1%	<0.1%	<0.1%
Trigger & lepton eff. SF (shape,η)	0.1%	<0.1%	0.2%
Trigger & lepton eff. SF (shape,p_T)	0.1%	<0.1%	0.2%
B-Jet identification eff. SF (normalisation)	0.1%	<0.1%	0.1%
B-Jet identification eff. SF (shape,η)	0.2%	<0.1%	0.7%
B-Jet identification eff. SF (shape,p_T)	0.6%	<0.1%	1.0%
B-Jet misidentification eff. SF	0.1%	<0.1%	0.1%

Model Uncertainties

Uncertainty Source	**average**	**minimum**	**maximum**
	(median of the relative uncertainty)		
Single top-quark BG (normalisation)	0.1%	<0.1%	0.1%
Single top-quark BG (Q^2 Scale)	0.3%	0.1%	0.5%
VV BG (normalisation)	<0.1%	<0.1%	<0.1%
V+jets & $t\bar{t}$ +V BG (normalisation)	0.6%	0.2%	1.0%
Q^2 scale (for $t\bar{t}$)	1.8%	0.7%	2.7%
ME-PS matching threshold (for $t\bar{t}$)	1.3%	0.6%	2.1%
m_{top} (for $t\bar{t}$)	1.0%	0.5%	2.0%
BR (for $t\bar{t}$)	<0.1%	<0.1%	<0.1%
PDF	<0.1%	<0.1%	0.1%
Hadronisation	1.9%	0.8%	4.4%

Table 11.6: Minimum, maximum and average medians for all sources of systematic uncertainties of the normalised differential cross section measurements in the **extrapolated parton level phase space** for the combined ℓ+jets results. "V" is an abbreviation for a vector boson V$\in$ (W,Z,γ).

Experimental Uncertainties

Uncertainty Source	**average** (median of the relative uncertainty)	**minimum** (median of the relative uncertainty)	**maximum** (median of the relative uncertainty)
Luminosity	<0.1%	<0.1%	0.1%
PU	0.2%	0.1%	0.3%
JES	3.2%	0.3%	7.3%
JER	1.1%	0.2%	5.8%
Trigger & lepton eff. SF (normalisation)	<0.1%	<0.1%	<0.1%
Trigger & lepton eff. SF (shape,η)	0.1%	<0.1%	0.6%
Trigger & lepton eff. SF (shape,p_T)	0.1%	<0.1%	0.5%
B-Jet identification eff. SF (normalisation)	0.2%	<0.1%	0.3%
B-Jet identification eff. SF (shape,η)	0.3%	<0.1%	1.2%
B-Jet identification eff. SF (shape,p_T)	0.7%	<0.1%	1.8%
B-Jet misidentification eff. SF	0.1%	<0.1%	0.2%

Model Uncertainties

Uncertainty Source	**average** (median of the relative uncertainty)	**minimum** (median of the relative uncertainty)	**maximum** (median of the relative uncertainty)
Single top-quark BG (normalisation)	0.2%	0.1%	0.5%
Single top-quark BG (Q^2 Scale)	0.4%	0.1%	1.5%
VV BG (normalisation)	<0.1%	<0.1%	<0.1%
V+jets & $t\bar{t}$ +V BG (normalisation)	1.2%	0.3%	2.8%
Q^2 scale (for $t\bar{t}$)	1.9%	0.6%	5.8%
ME-PS matching threshold (for $t\bar{t}$)	1.4%	0.5%	5.3%
m_{top} (for $t\bar{t}$)	2.2%	0.5%	7.9%
BR (for $t\bar{t}$)	<0.1%	<0.1%	<0.1%
PDF	0.9%	<0.1%	8.0%
Hadronisation	3.8%	0.3%	12.5%

Table 11.7: Minimum, maximum and average medians for all sources of systematic uncertainties of the normalised differential cross section measurements in the **visible particle level phase space** for the combined ℓ+jets results. "V" is an abbreviation for a vector boson V$\in$ (W,Z,γ).

11.2.5 Comparisons to Other Measurements and Predictions

To validate the results of the differential measurements, to illustrate the improvements with respect to previous differential $t\bar{t}$ cross section measurements and to assess the achieved precision, several comparisons are performed.

For this purpose, the normalised differential cross sections at $\sqrt{s} = 8$ TeV are compared to CMS results at $\sqrt{s} = 7$ TeV in the same ℓ+jets final state and the $\sqrt{s} = 8$ TeV CMS results of the dileptonic final state. Furthermore, the CMS and ATLAS results at $\sqrt{s} = 7$ TeV are compared to each other. Moreover, the normalised differential predictions of several MC predictions are compared to study differences in detail. Additionally, the MC predictions of MadGraph+pythia with the Perugia UE tune (at $\sqrt{s} = 8$ TeV), which can describe some particular measurements better, are shown.

A comparison is only feasible for measurements which are performed in the same phase space. Consequently, only the results for the top quarks and $t\bar{t}$ system, being obtained in the extrapolated parton level phase space, can be evaluated.

Not all measurements are done in the same analysis binning because detector resolutions and migration effects depend e.g. on the detector, the investigated final state, the event selection and the event reconstruction. To compare results in different analysis binnings, horizontal bin centre corrections (BCCs) are applied to all data points [209]. These horizontal BCCs place the measured data points at the x value where the bin average cross section of the applied analysis binning equals the cross section of an unbinned continuous prediction. This facilitates the comparison of analyses with different binnings when applying BCCs with respect to the identical unbinned prediction. Furthermore, it is checked in all cases that small differences in the predictions originating for example from subtle changes of a MC prediction at different centre-of-mass energies or different shapes of the predictions have only a very small effect on the obtained bin centre corrected x values. Moreover, the MadGraph+pythia prediction used to derive the horizontal BCCs and ratios for the CMS results at $\sqrt{s} = 7$ TeV does not include the modelling of spin correlations using MadSpin. The impact on the compared cross section quantities is studied in detail and found to be negligible.

Additionally, the comparison between all CMS measurements is done relative to the MC prediction to highlight differences between data and the MadGraph+pythia prediction. Also for the BCCs, the MadGraph+pythia prediction is chosen because it is the default $t\bar{t}$ prediction in CMS and is therefore used to perform all measurements within this thesis.

Comparison to CMS Results at $\sqrt{s} = 7$ TeV

The underlying physics for the LHC pp dataset recorded at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV is identical. Therefore, the relative performance of the theory predictions with respect to the data is expected to be similar. Consequently, any deviation from the predictions, which is caused by a mismodelling of the $t\bar{t}$ production mechanism, is expected to reveal itself at both centre-of-mass energies. As 7 and 8 TeV are not too different, the qualitative agreement of data over (MC) prediction is expected to be similar for the results obtained in the extrapolated parton level phase space.

Consequently, the normalised differential cross section results at $\sqrt{s} = 8$ TeV (this thesis) are compared to the CMS results at $\sqrt{s} = 7$ TeV in the same final state [25], to which the work of this thesis also contributed. The corresponding values for the data, which are obtained with similar methods as in this thesis, the MadGraph+pythia prediction and the BCCs are taken from [23, 25].

A good consistency is observed for all quantities. Also for $\sqrt{s} = 7$ TeV, the transverse momentum of the top quarks (p_T^t, Fig. 11.8a) is found to be softer in data than predicted by MadGraph+pythia. Furthermore, it is clearly visible that the precision is higher for the $\sqrt{s} = 8$ TeV measurement presented in this thesis than for the $\sqrt{s} = 7$ TeV results. As the uncertainties have been already systematically limited for the $\sqrt{s} = 7$ TeV results [23], a further reduction of uncertainties with a larger dataset is only expected, if the larger statistics can be used to reduce systematic uncertainty sources, e.g. by reducing migration effects. The comparisons for the combined ℓ+jets final state are shown in Figure 11.8.

Comparison to CMS Results in the Dileptonic Final State

The normalised differential cross sections measured in this thesis are compared to the complementary and statistical independent CMS results of the dileptonic final state at $\sqrt{s} = 8$ TeV [27]. As the visible phase space is different for the two $t\bar{t}$ final states, only the top quark and $t\bar{t}$ quantities can be compared. The MadGraph+pythia MC prediction used to calculate the ratios and the horizontal BCCs is identical for both channels and refers to the prediction listed in Chapter 5.3.2.

All comparisons are shown in Figure 11.9 and 11.10. A good consistency between the different final state analyses is observed for all quantities. Especially all measurements related to the transverse momentum of the top quark are completely consistent and reveal a softer spectrum in data than predicted by MadGraph+pythia.
The precision achieved for the preliminary results of the dileptonic final state is lower than the typical 3-7% of the measurements presented in this thesis and typically, a binning of lower granularity is chosen.

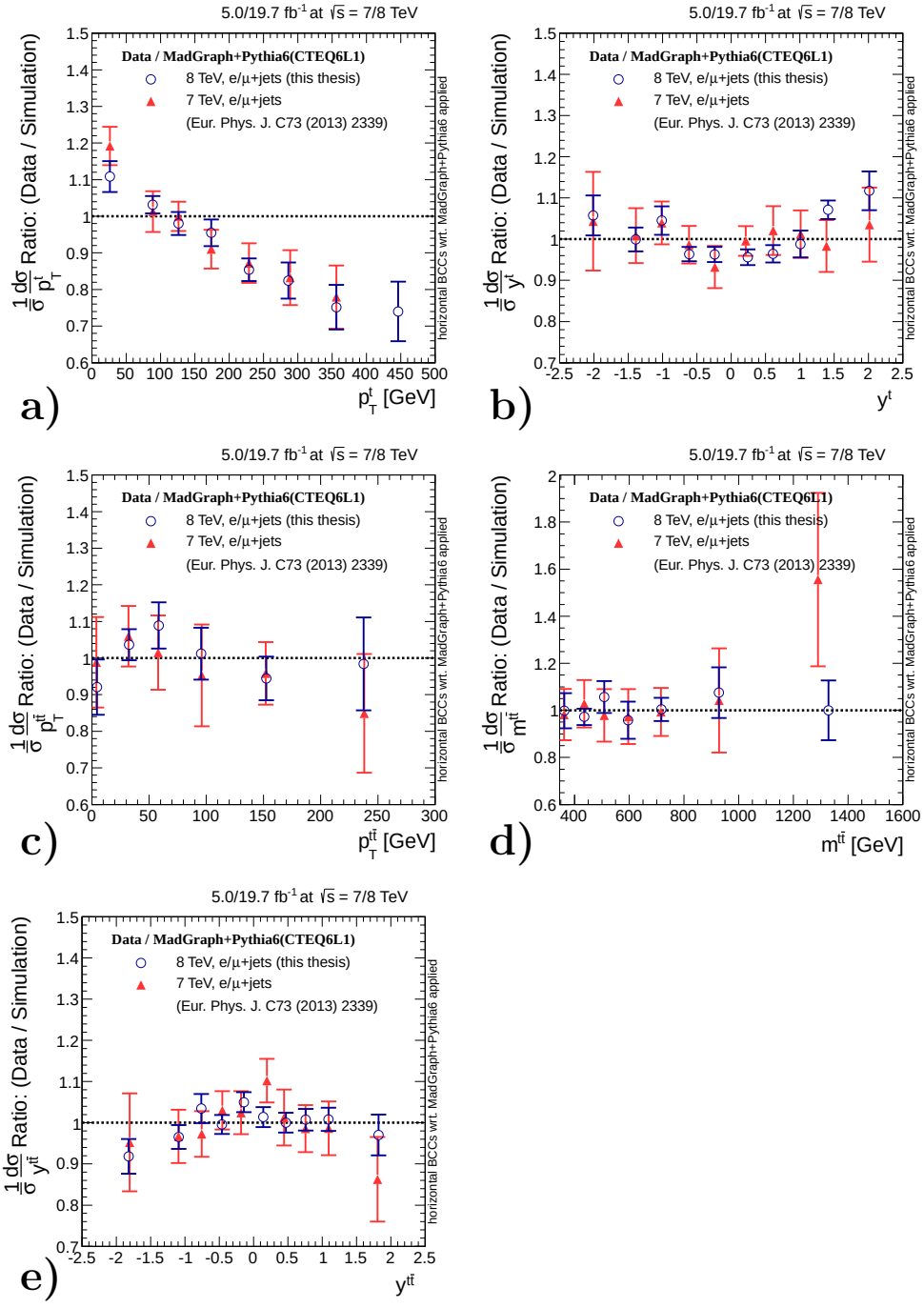

Figure 11.8: Comparison of the (CMS) data to MC ratio at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV in the ℓ+jets final state for the normalised differential distributions in the extrapolated parton level phase space. The corresponding MadGraph+pythia prediction is used to obtain the ratio and the horizontal bin centre corrections. The transverse momentum (p_T^t, a) and rapidity (y^t, b) of the top quarks as well as the transverse momentum ($p_T^{t\bar{t}}$, c), the invariant mass ($m^{t\bar{t}}$, d) and the rapidity ($y^{t\bar{t}}$, e) of the $t\bar{t}$ system are presented. The error bars indicate the combined statistical and systematic uncertainties of the measurement.

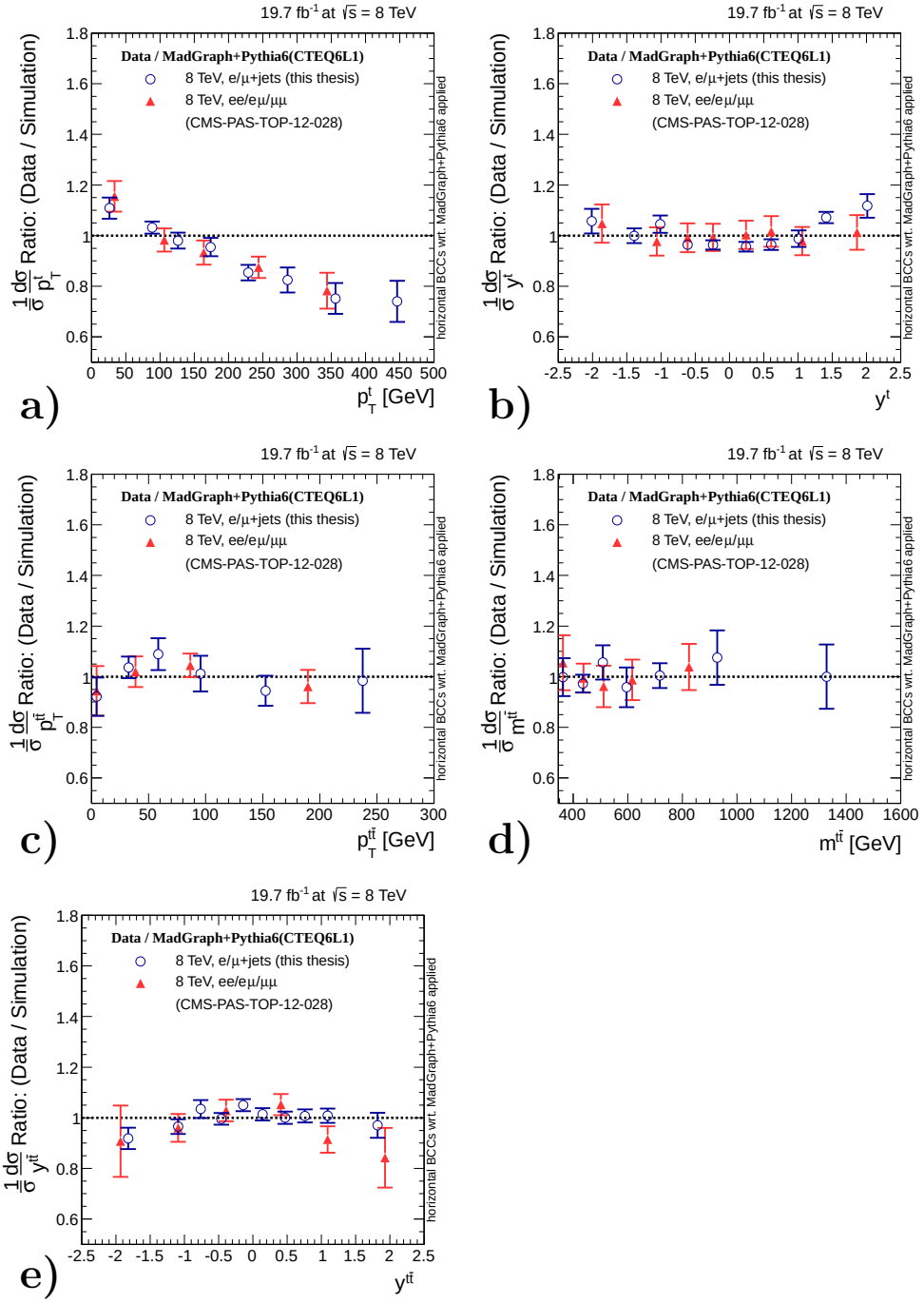

Figure 11.9: Comparison of the (CMS) data to MC ratio in the ℓ+jets and dileptonic final state at $\sqrt{s} = 8$ TeV for the normalised differential distributions in the extrapolated parton level phase space. The corresponding MadGraph+pythia prediction is used to obtain the ratio and the horizontal bin centre corrections. The transverse momentum (p_T^t, a) and rapidity (y^t, b) of the top quarks as well as the transverse momentum ($p_T^{t\bar{t}}$, c), the invariant mass ($m^{t\bar{t}}$, d) and the rapidity ($y^{t\bar{t}}$, e) of the $t\bar{t}$ system are presented. The error bars indicate the combined statistical and systematic uncertainties of the measurement.

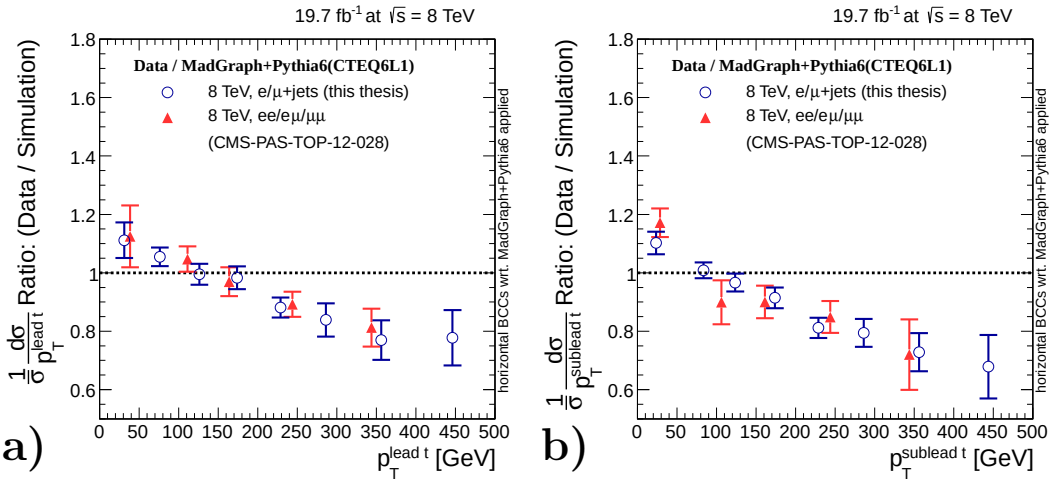

Figure 11.10: Comparison of the data to MC prediction agreement for the normalised differential distributions in the extrapolated parton level phase space, which are measured by CMS in the ℓ+jets and dileptonic final state at $\sqrt{s} = 8$ TeV. The corresponding MadGraph+pythia prediction is used to obtain the ratio and the horizontal bin centre corrections. The transverse momentum of the top quark with the higher ($p_T^{\text{lead t}}$, a) and lower ($p_T^{\text{sublead t}}$, b) transverse momentum. The error bars indicate the combined statistical and systematic uncertainties of the measurement.

Comparing Results from ATLAS and CMS at $\sqrt{s} = 7$ TeV

The CMS analysis results are compared to the ones from the ATLAS experiment. This comparison is an excellent cross check as a completely different detector, dataset and analysis strategy (selection, unfolding) is used.

No published ATLAS results are available yet for $\sqrt{s} = 8$ TeV. Therefore, the comparison is done between the $\sqrt{s} = 7$ TeV ATLAS result [210] and the $\sqrt{s} = 7$ TeV CMS result [23, 25].
The main differences between the $\sqrt{s} = 7$ TeV analysis of CMS in comparison to this thesis are the kinematic fit used for the event reconstruction, which uses an equal top-quark mass constraint instead of the Double Kinematic Fit, and that no minimum χ^2-probability criteria is imposed. As discussed previously, the same data over prediction trend, e.g. for the transverse momentum of the top quarks, is observed for the $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV CMS results. Hence, the observed qualitative agreement between ATLAS and CMS results at $\sqrt{s} = 7$ TeV is also an indicator for the agreement of the ATLAS results with the results of this thesis.

To enable a direct comparison, the CMS analysis at $\sqrt{s} = 7$ TeV is redone within the work of this thesis using the ATLAS binning. It is checked that purity and stability are still reasonably high and the regularisation parameters of the unfolding procedure are optimised for each quantity with the ATLAS binning using the minimum-global-correlation method (see Chapter 9.3.1). Repeating the CMS analysis in the ATLAS binning is possible for the following quantities: p_T^t, $m^{t\bar{t}}$ and $y^{t\bar{t}}$. Finally, the results for CMS data are compared to the ATLAS measurement and the predictions of the MadGraph+pythia and powheg+herwig MC generators in Figure 11.11. The MC predictions are taken from the official CMS "Fall11" MC production cycle for $\sqrt{s} = 7$ TeV. For a better comparison, the ratio with respect to the MadGraph+pythia prediction is added.

The achieved precision is similar for both experiments and in general good agreement between the results is found.
The ATLAS and CMS results for $m^{t\bar{t}}$ and $y^{t\bar{t}}$ coincide with each other and the predictions. Furthermore, the observed softer shape in data compared to the MadGraph+pythia prediction for the top-quark transverse momentum is also visible for $p_T^t > 200\,\text{GeV}$ in the ATLAS result. For $p_T^t < 200\,\text{GeV}$, the ATLAS measurement is in agreement with the prediction of MadGraph+pythia while the CMS result has still a softer shape. Nevertheless, ATLAS and CMS results are compatible within the uncertainties of both measurements and only the first bin shows a difference. It should be noted, that the powheg+herwig prediction is in between CMS and ATLAS data and can therefore described both measurements well.

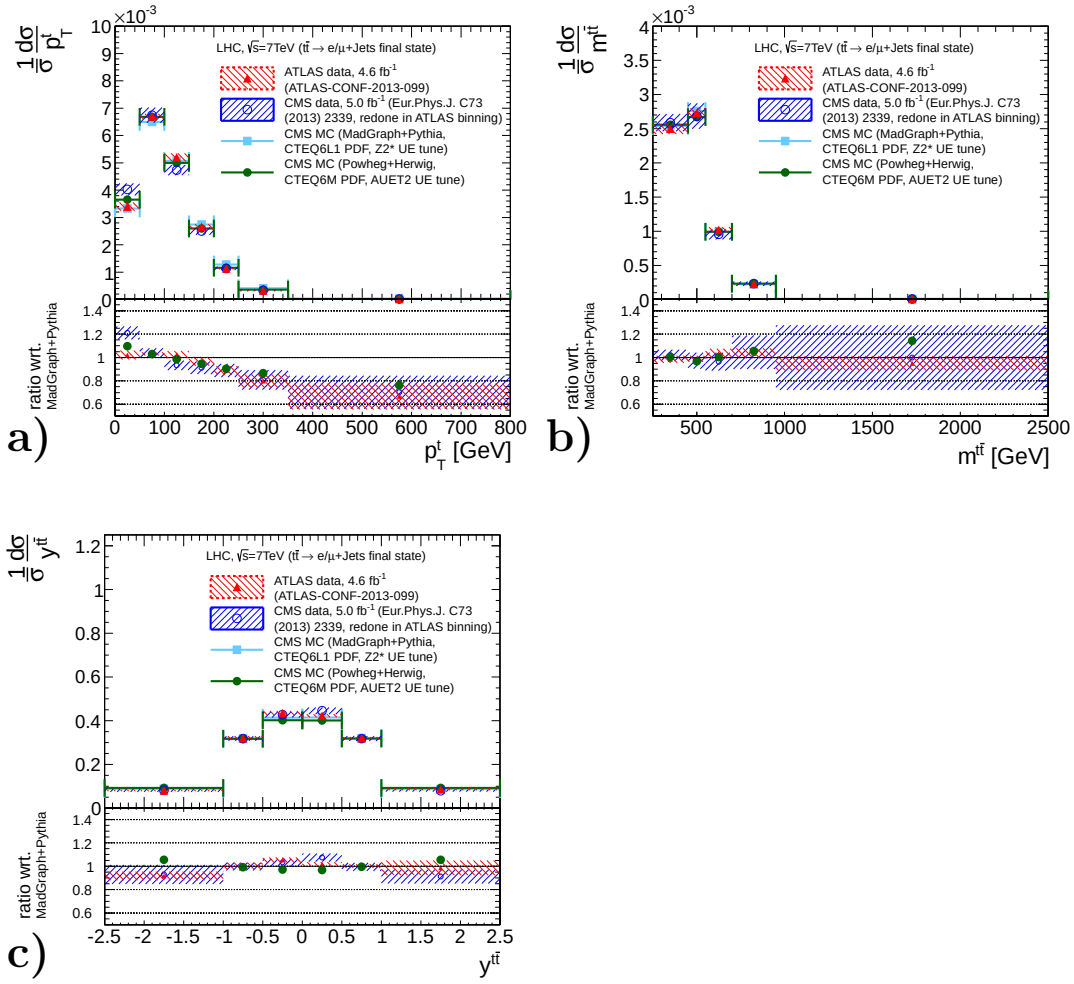

Figure 11.11: Comparison of normalised differential cross sections for the transverse momentum of the top quarks (p_T^t, a), the invariant mass ($m^{t\bar{t}}$, b) and the rapidity ($y^{t\bar{t}}$, c) of the $t\bar{t}$ system at $\sqrt{s} = 7$ TeV. The ATLAS result (blue) [210] is compared to the CMS result (red), which is redone in the same binning, and the MadGraph+pythia MC prediction (turquoise).

Comparing Different MC Predictions

To compare different simulation techniques, tools and definitions, also between ATLAS and CMS, the differential predictions for the most important cross section quantities p_T^t, y^t, $p_T^{t\bar{t}}$, $m^{t\bar{t}}$ and $y^{t\bar{t}}$ are compared within the TopLHCWG. The comparison, to which this thesis contributed, is done for the following MC predictions :

- ALPGEN+HERWIG (ATLAS only)
- MADGRAPH+PYTHIA (CMS only)
- MC@NLO+HERWIG (ATLAS and CMS)
- POWHEG+PYTHIA (ATLAS and CMS)
- POWHEG+HERWIG (ATLAS and CMS)

This study is done for $\sqrt{s} = 7$ TeV. A sufficiently large number of simulated events is incorporated to allow a detailed comparison using a binning of high granularity, i.e. beyond the wide binning, which is needed for the data measurements. All ATLAS predictions are obtained within the collaboration for the TopLHCWG [178]. As UE tunes, the Perugia *P11* tune [211], the ATLAS underlying event tune *AUET2* [117] and the *Z2* tune [212] are used. All predictions are normalised by the number of entries to compare the predicted shape of the distributions.

A perfect agreement between the ATLAS and CMS predictions is observed for the same MC generator. This gives confidence that the definition of the top quark as detailed in Chapter 8.1.1 is equal for both experiments and the previously presented comparison of CMS and ATLAS results is absolutely valid.
The predicted softer p_T^t shape of POWHEG+HERWIG in comparison to all other MC event generators studied, is confirmed by ATLAS. This ensures that this prediction is correctly simulated. The difference between the POWHEG+HERWIG and the POWHEG+PYTHIA MC prediction for p_T^t is still under investigation within the TopLHCWG [178, 213] and will be discussed in more detail in Chapter 12.1.
Furthermore, the predictions of the ALPGEN+HERWIG event generator, which is used to perform the ATLAS analysis, and the MADGRAPH+PYTHIA event generator, which is used to perform the CMS analyses, are found to be almost equal. The fine-binned CMS predictions are shown in Appendix F.

Moreover, the choice of the UE tune, which is different for several predictions between ATLAS and CMS, is observed to impact only $p_T^{t\bar{t}}$. For an improved visibility of the final results not all available predictions are shown in Chapter 11.2.1 and 11.2.2. The additional MADGRAPH+PYTHIA prediction using the Perugia P11 UE tune [211] is presented in Figure 11.12 for the $p_T^{t\bar{t}}$ and $\Delta\phi\,(t,\bar{t})$ distributions for which an improved data description can be achieved. For all other quantities (including p_T^t) the predictions are similar to the default MADGRAPH+PYTHIA prediction using the Z2* UE tune. This difference might be related to some parameters in the UE tune that effect the modelling of additional radiation.

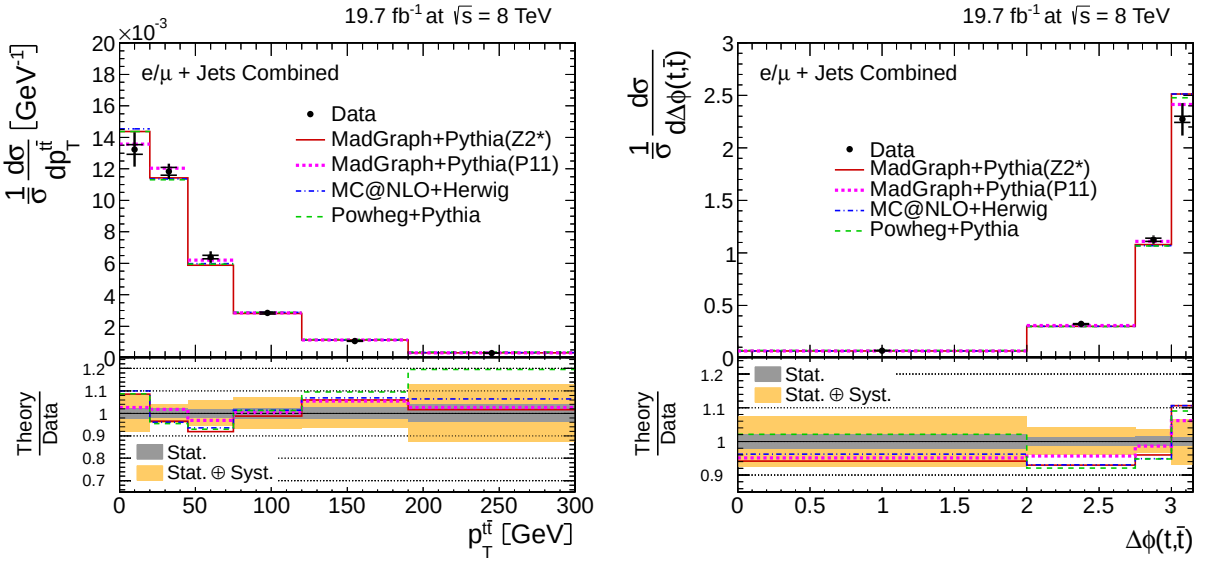

Figure 11.12: Normalised differential cross section for transverse momentum of the $t\bar{t}$ system (left) and the difference in azimuthal angle of the top quarks (right) in the extrapolated parton level phase space including the MadGraph+pythia prediction with the Perugia UE tune. The grey (orange) band in the ratio corresponds to the statistical (combined statistical and systematic) uncertainty of the data measurement.

Chapter 12

Transverse Momentum of the Top Quarks

Due to the observed differences between data and several MC predictions, the measurement of the transverse momentum of the top quarks is one of the most important results of this thesis. These differences need to be understood as a potential mismodelling of this distribution can impact other SM measurements as well as BSM searches.
All findings of this thesis concerning the top quark transverse momentum are summarised in Chapter 12.1 while additional tests are performed in Chapter 12.2. Finally, generalised correction factors for the MadGraph+pythia event generator are derived from the measurement in Chapter 12.3.

12.1 Summary of the Results

All findings regarding the measurement of the transverse momentum of the top quarks (p_T^t) that have been discussed previously are briefly listed in the following:

- softer p_T^t spectrum measured (Figure 12.1) than predicted by MadGraph/powheg+pythia and mc@nlo+herwig
- p_T^t in data better described by powheg+herwig and a perturbative QCD calculation at approx. NNLO precision [205]

} Chapter 11.2.1

- consistency among CMS results at $\sqrt{s} = 7$ TeV and 8 TeV
- consistency among ATLAS and CMS results at $\sqrt{s} = 7$ TeV, at least for $p_T^t \geq 200\,\text{GeV}$
- identical top-quark definition for ATLAS and CMS

} Chapter 11.2.5

- consistent results for the measured top-quark momentum
  - in the $t\bar{t}$ rest frame (p_T^t ($t\bar{t}$ com))
  - of the top quark with higher momentum ($p_T^{\text{lead t}}$)
  - of the top quark with lower momentum ($p_T^{\text{sublead t}}$)

} Chapter 11.2.1

- consistently softer momentum distributions of the top-quark decay products in data, even before unfolding and Double Kinematic Fit (Chapter 11.2.2 and 6.9)
- no bias of the unfolding procedure due to the assumed p_T^t prediction (Chapter 9.4)

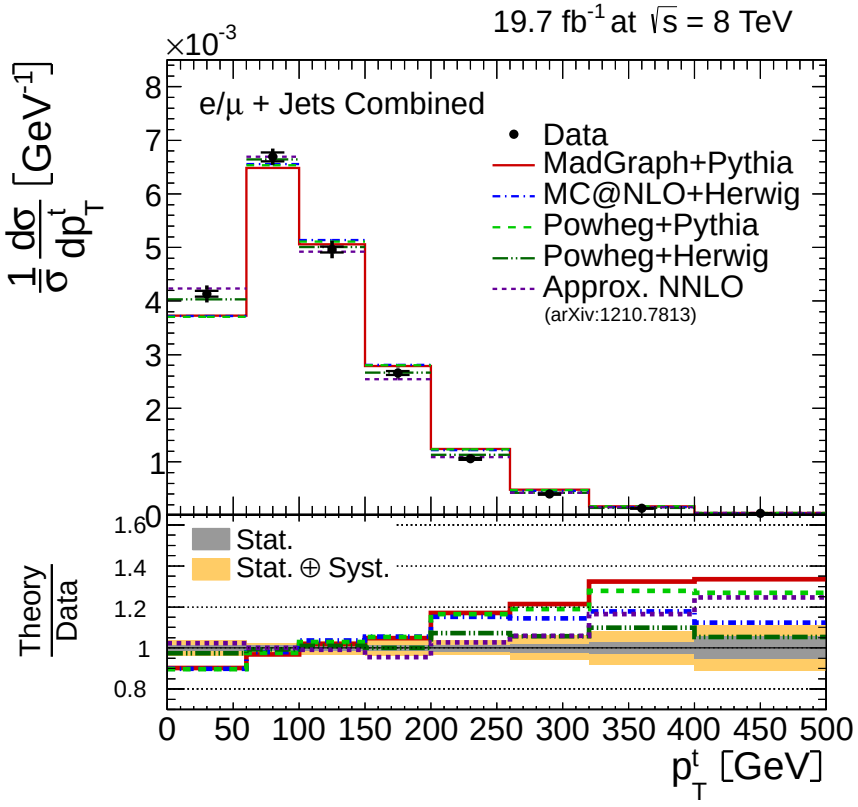

Figure 12.1: Normalised differential cross section for the transverse momentum of both top quarks in the detector rest frame in the extrapolated parton level phase space. The grey (orange) band in the ratio corresponds to the statistical (combined statistical and systematic) uncertainty of the data measurement. This Figure is identical to Figure 11.3a.

The origin of the observed difference between the POWHEG+PYTHIA and the POWHEG+HERWIG prediction for p_T^t is still under investigation. First studies indicate that it might be due to the technical implementation in HERWIG for the momentum conservation in case of additional radiation [213] if HERWIG is interfaced to a ME generator like POWHEG. The size of this effect is larger than the typical scale uncertainties and could be related to higher order effects beyond NLO or even beyond NNLO QCD precision.

One possible explanation for all aforementioned observations is a higher order effect that is not covered by the NLO uncertainties. This hypothesis can be tested further once the available perturbative QCD calculation at full NNLO+NNLL accuracy [44] is extended to differential predictions.

12.2 Additional Studies

To study the observed softer p_T^t in data in more detail, additional tests are performed separating the top quarks by different criteria (Chapter 12.2.1) or testing different kinematic ranges and experimental inputs (Chapter 12.2.2). The purpose of these tests is to demonstrate the stability of the measurement and test the dependency with respect to the modelling of some analysis inputs like JES or PU.

All tests are done at the reconstruction level by comparing the event yield of the MC prediction to data. For non-$t\bar{t}$ processes, the standard MC predictions and SM cross sections for the normalisation are used (see Chapter 5.3.2). The MadGraph+pythia event generator is consistently used for the prediction of the $t\bar{t}$ component. Using the same strategy as for all event yields, the total number of $t\bar{t}$ events is chosen such that for each distribution separately the total number of events in data and MC is equal. This corresponds to a normalisation with respect to the in situ measured inclusive cross section and allows to focus on the shape of the distribution. For an easier comparison, a linear fit is added to the data over MC ratio in most cases. Other fit functions are found to describe the data over MC ratios worse.

Moreover, the effect of different parameter choices in the MadGraph+pythia MC prediction is studied in Chapter 12.2.3. This study is performed for the normalised differential cross section as function of p_T^t to estimate the uncertainties of this prediction.

12.2.1 Separating the Two Top Quarks

The p_T^t spectrum is separately studied for both top quarks. Here, three different kinds of separations are studied: by decay mode in hadronically and leptonically decaying top quarks, by charge in top quark and antitop quark, and by p_T in the top quark with higher and lower transverse momentum. Moreover, the semileptonic electron and muon final state are investigated separately.

The results are shown in Figure 12.2 and 12.3. A softer spectrum in data is observed for all separations of the top quarks. As previously discussed, the discrepancy between data and MadGraph+pythia is larger for the top quark with lower momentum. Furthermore, the results of the separate final states are completely consistent with each other.

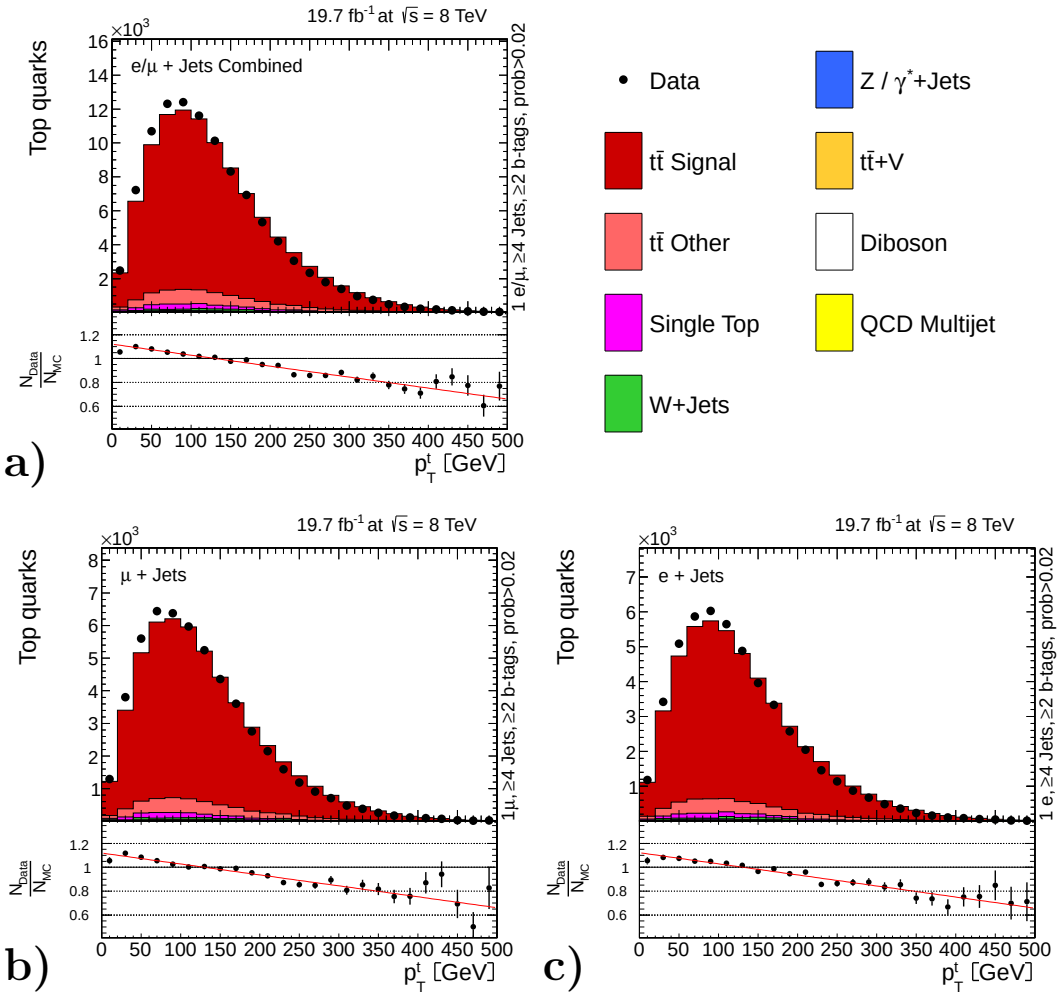

Figure 12.2: Reconstruction level event yields for the transverse momentum of the top quarks (p_T^t) in data compared to the MadGraph+pythia MC prediction for the combined ℓ+jets (a) and separately for the muon+jets (b) and electron+jets (c) final state. All plots are obtained after full event selection, kinematic reconstruction, and the minimal χ^2-probability requirement of 2%. The red line indicates a linear fit to the ratio of data and MC prediction.

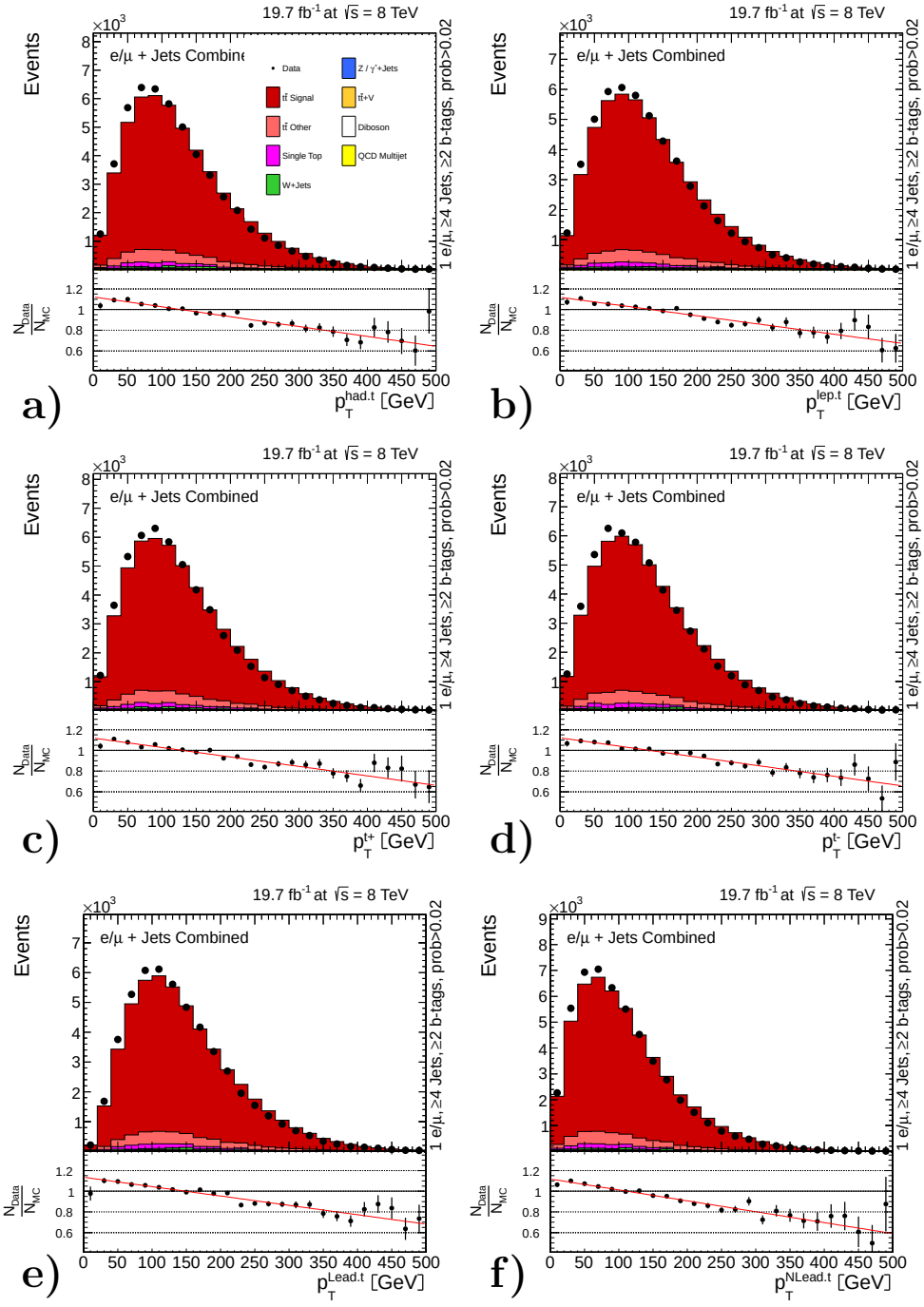

Figure 12.3: Reconstruction level event yields for the transverse momentum of the top quarks in data compared to the MadGraph+pythia MC prediction separated by the top-quark decay in hadronically and leptonically decaying top quarks (a,b), by charge in top quark and antitop quark (c and d) and by p_T into the top quark with higher and lower transverse momentum (e and f). All plots are obtained for the combined ℓ+jets final state after full event selection, kinematic reconstruction, and the minimal χ^2-probability requirement of 2%. The red line indicates a linear fit to the ratio of data and MC prediction.

12.2.2 Different Kinematic Ranges

To exclude that the difference between the p_T^t distribution predicted by most MC event generators and observed in data is due to an experimental mismodelling, the reconstructed p_T^t is studied in different kinematic regions.

p_T^t is studied for different intervals of the *number of primary vertices* (nPV), to prove that the difference is not related to a mismodelling of the underlying PU distribution. The nPV intervals are chosen to have approximately the same number of events.
As shown in Figure 12.4, a similar trend for the data over MC prediction ratio of the reconstructed p_T^t for all intervals of primary vertices is observed.

Testing a potential mismodelling of the background as well as of wrongly reconstructed events, p_T^t is studied for different intervals of χ^2 of the kinematic reconstruction. Correctly reconstructed signal events are expected to be significantly enhanced for lower values of χ^2. The chosen maximum χ^2 values of 1.4, 2.1, 3.2, 7.8 15.2 and 32.2 correspond to a minimum χ^2-probability requirement of 50, 35, 20, 2, $5 \cdot 10^{-4}$ and $1 \cdot 10^{-7}$%, respectively. The intervals are chosen to contain approximately the same number of events.
The results of this study are shown in Figure 12.5. A similar data over prediction trend is observed for all χ^2 intervals indicating that the minimal χ^2-probability requirement of 2%, which is introduced for all normalised differential cross sections (see Chapter 7.2.3 and 7.3.3), does not affect the p_T^t shape at the level of reconstructed event yields.

In order to investigate the effect of a potential mismodelling of the JES, the agreement between data and prediction for p_T^t is compared for different values of the JES. The studied up and down variations correspond to a variation of the JES, which is provided by the CMS JetMET Physics Analysis Group (see Chapter 6.4.1), of one standard deviation of the total uncertainties.
As shown in Figure 12.6, the agreement between data and MC prediction for the p_T^t distribution is slightly improved for a reduced JES. However, the observed effect of these JES variations is too small to cover the difference between data and prediction for p_T^t.
On the other hand, the agreement with data of other distributions like $m^{t\bar{t}}$ is found to degrade with variations of the JES. Furthermore, the reconstructed mass of the hadronically decaying W-boson, which is sensitive to the JES, is shown in Figure 12.7 for the same JES variations. Here, the best agreement with data is found for the JES which is increased by one sigma of the provided uncertainty, also disfavouring the approach of lowering the measured JES to describe p_T^t.
Therefore, varying only the JES such that p_T^t is properly described by the MadGraph+pythia prediction seems to be not favoured in the whole context of the performed analysis.

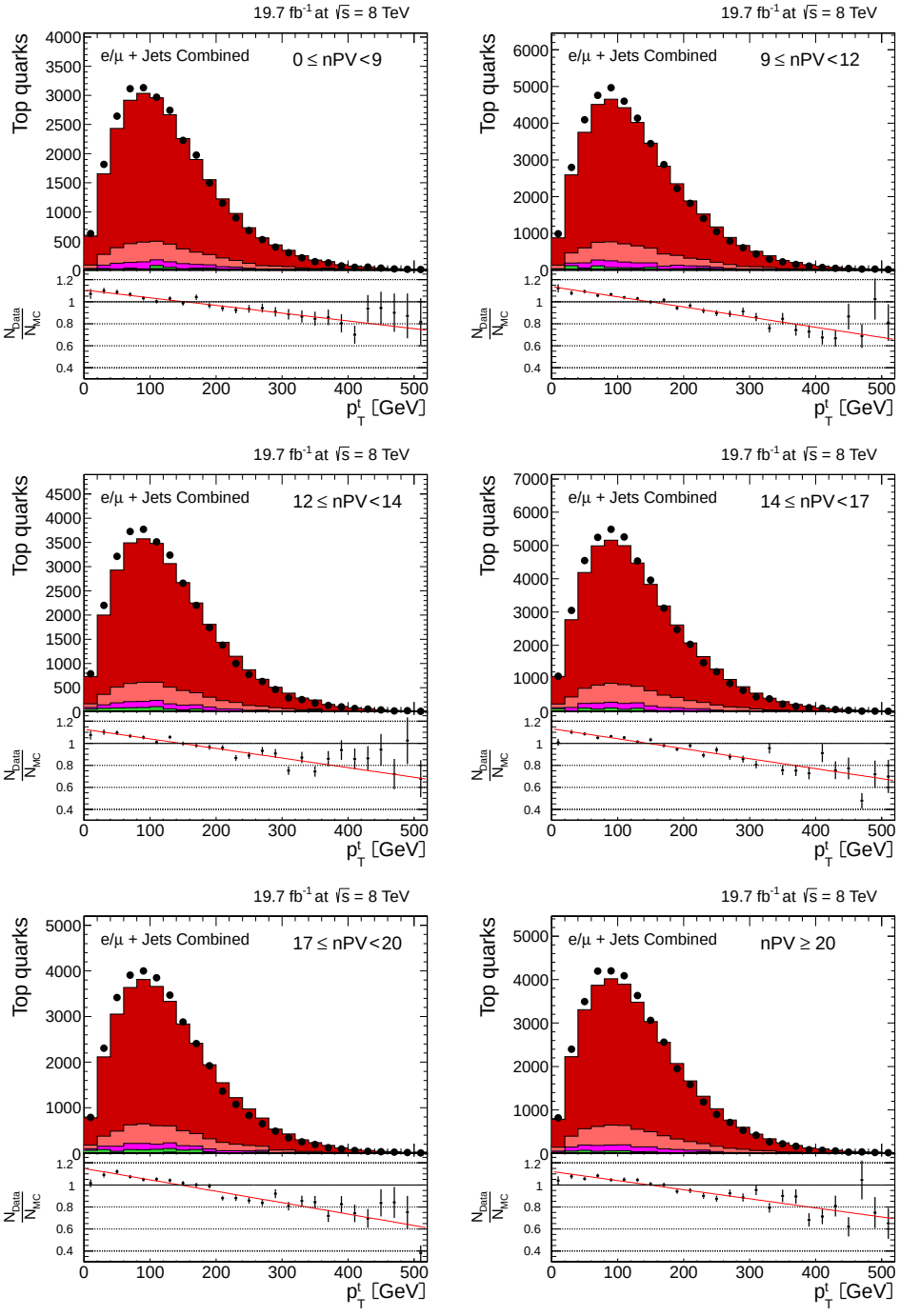

Figure 12.4: Reconstruction level event yields for the transverse momentum of the top quarks in data compared to the MadGraph+pythia MC prediction for **different numbers of primary vertices (nPV)**. All plots are obtained for the combined ℓ+jets final state after full event selection and kinematic reconstruction. The red line indicates a linear fit to the ratio of data and MC prediction.

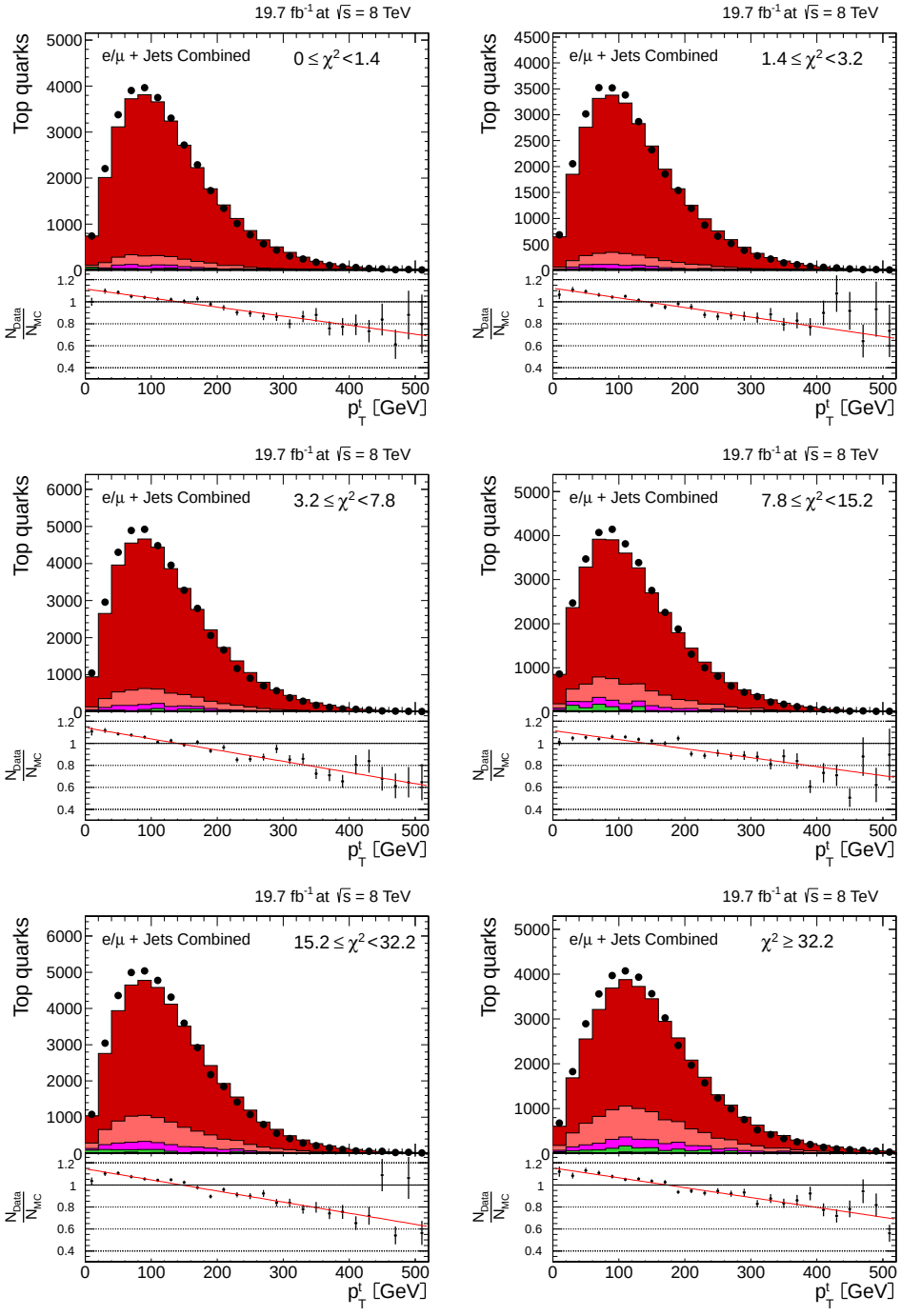

Figure 12.5: Reconstruction level event yields for the transverse momentum of the top quarks in data compared to the MadGraph+pythia MC prediction for **different values of χ^2 from the event reconstruction**. All plots are obtained for the combined ℓ+jets final state after full event selection and kinematic reconstruction. The red line indicates a linear fit to the ratio of data and MC prediction.

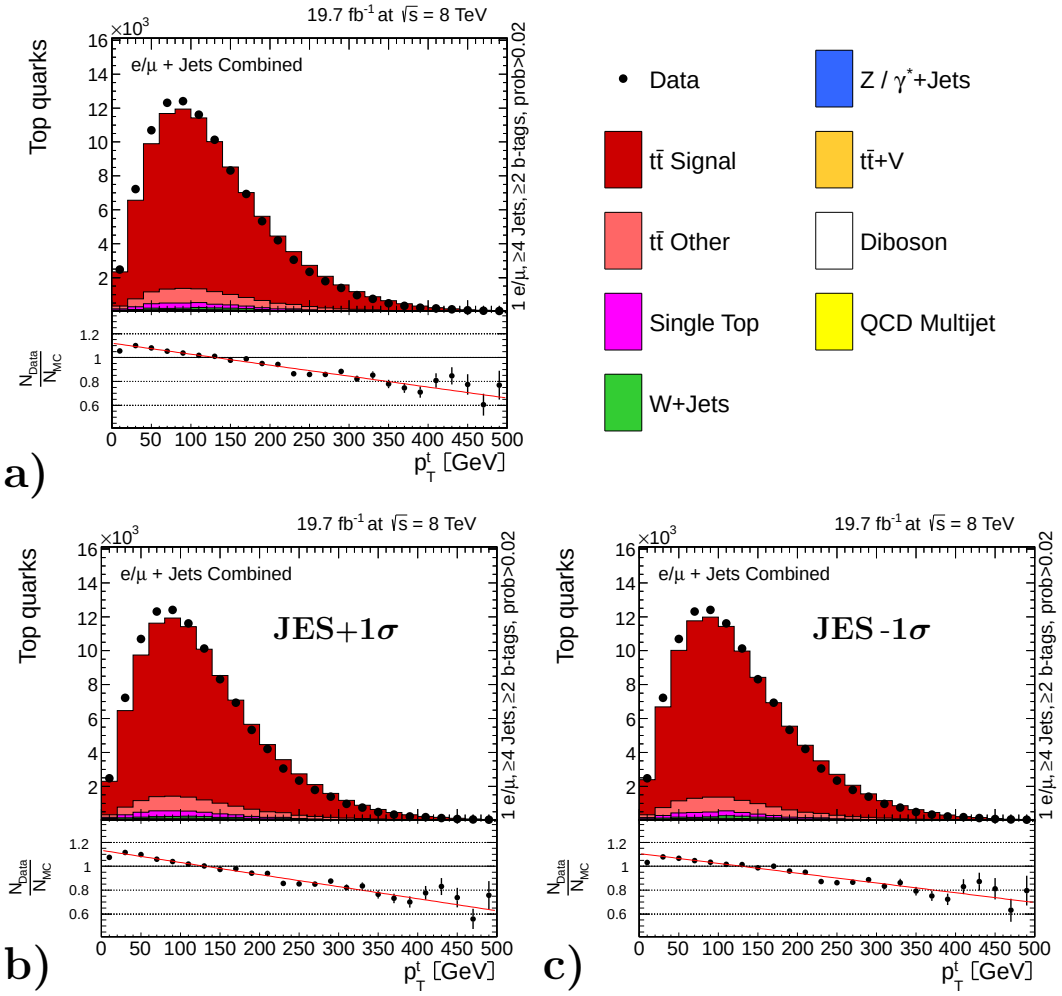

Figure 12.6: Reconstruction level event yields for the transverse momentum of the top quarks in data compared to the MadGraph+pythia MC prediction for the **default (a), increased (b) and decreased (c) Jet Energy Scale (JES)**. All plots are obtained for the combined ℓ+jets final state after full event selection, kinematic reconstruction and the minimal χ^2-probability requirement of 2%. The JES variations correspond to the one sigma uncertainties as provided by the JetMET group. The red line indicates a linear fit to the ratio of data and MC prediction.

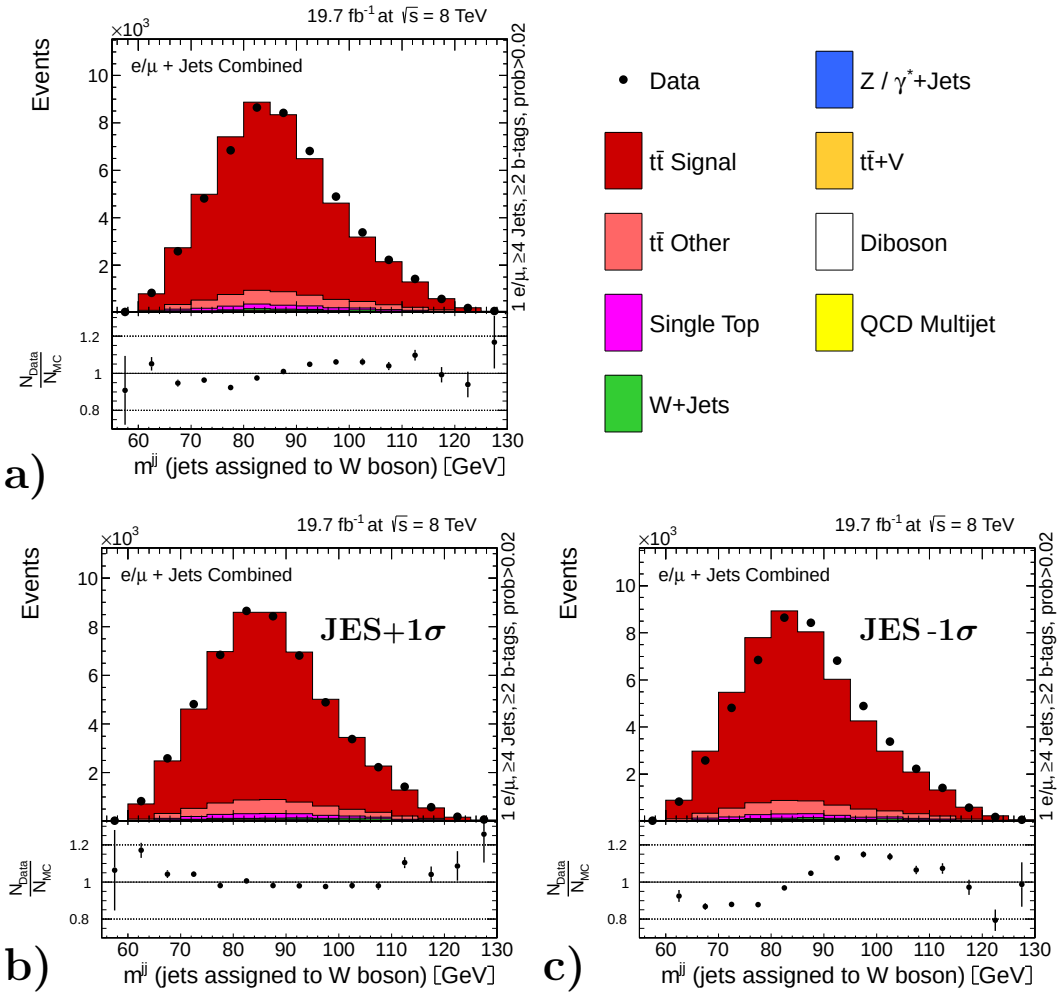

Figure 12.7: Reconstructed mass of the W boson in data compared to the MadGraph+pythia MC prediction for **default (a), increased (b) and decreased (c) Jet Energy Scale (JES)**. All plots are obtained for the combined ℓ+jets final state after full event selection, kinematic reconstruction and the minimal χ^2-probability requirement of 2%. The JES variations correspond to the one sigma uncertainties as provided by the JetMET group.

12.2.3 Different Parameters of the MadGraph+Pythia Monte-Carlo Prediction

No uncertainty bands are provided for the theory predictions in the results presented in Chapter 11.2.1 and 11.2.2. Therefore, the most important parameters of the MadGraph+pythia prediction are varied and their effect on the predicted p_T^t spectrum is studied.

Besides the variation of the hard scattering scale, the ME-PS matching threshold and the top-quark mass, also the effect of the UE tune is studied using the P11 tune [211]. Moreover, the effect of *increased multi-parton interaction* (mPiHi) and of *neglecting colour reconnection* (noCR) is tested for the P11 UE tune, using specific $t\bar{t}$ simulations which are also generated within the official CMS "Summer12" MC production cycle. In addition to these parameter studies, also the predictions of the other MC event generators are shown in Figure 12.8. All differences due to variations of the model parameters are much smaller than the difference with respect to data. The powheg+herwig prediction is a good reference as it is similar to data. The uncertainties of the measurement are 3.8, 2.3, 3.2, 3.8, 3.6, 6.0, 8.1 and 11.0% respectively for all bins (see Appendix D). Therefore, the obtained changes of less than 5% are not sufficient to describe the data measurement. Especially in the low p_T region ($p_T^t < 200\,\text{GeV}$), the model uncertainties are small (typically 1%) while the observed difference with respect to data is up to 10%.

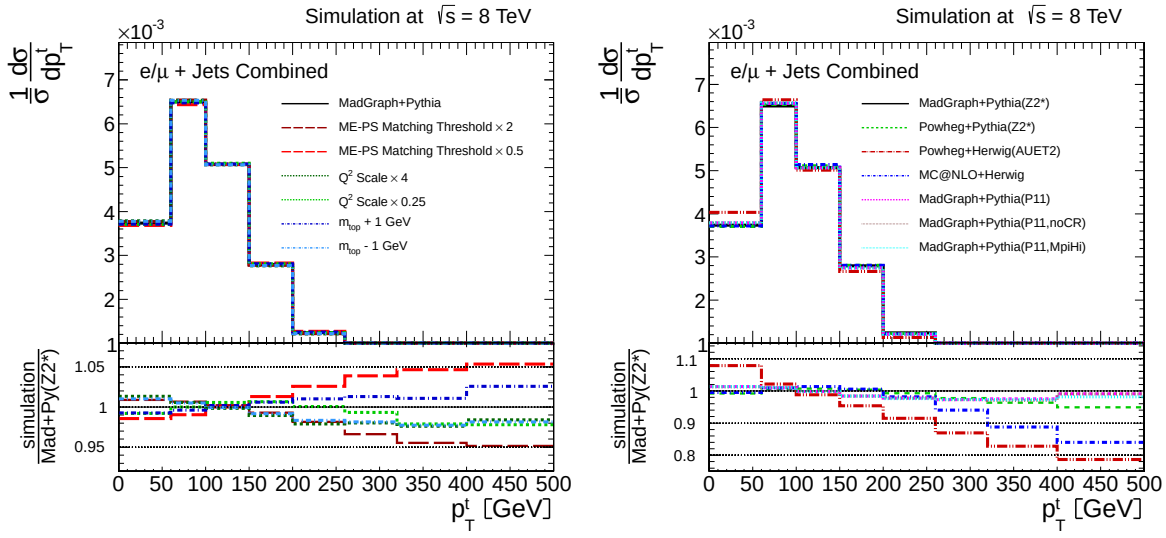

Figure 12.8: The predicted normalised differential cross section for p_T^t in the extrapolated parton level for different parameter choices of the MadGraph+pythia prediction (left) and different MC predictions (right)

12.3 Correction Factors for the MadGraph+Pythia Monte-Carlo Prediction

The origin of the difference between data and prediction for p_T^t has not been clarified so far. Possible explanations are higher order QCD corrections, EWK corrections or non-resonant production of $t\bar{t}$-like final states, e.g. $gg \to b\bar{b}\mu\nu qq'$. In order to test the global effect of the measured softer p_T^t spectrum, generalised correction factors are derived for the MadGraph+pythia prediction in Chapter 12.3.1 and their application is discussed in Chapter 12.3.2.

12.3.1 Calculation from CMS Data

The correction factors $SF(p_T^t)$ are derived from the ratio of measured and predicted normalised differential cross section in the extrapolated parton level phase space.

To obtain a steady parametrisation that is independent from the binning of the data measurement(s), horizontal BCCs are applied to all data points (see Chapter 11.2.5 for a more detailed description). This allows for adding other measurements in the same phase space like the CMS measurement in the dileptonic final state. The choice of the theory for the application of the BCCs is found to be negligible for the extracted correction factors and the approx. NNLO prediction [205] is finally chosen to derive the BCCs.

An exponential function $SF(p_T^t, a, b)$ is chosen as parametrisation for the ratio between data and prediction of $\frac{1}{\sigma}\frac{d\sigma}{dp_T^t}$:

$$SF(p_T^t) = \exp(a + b \cdot p_T^t). \tag{12.1}$$

The two parameters a and b are determined by a fit to the bin-centre-corrected normalised differential cross section ratios between data and the MadGraph+pythia MC prediction. Other parametrisations like linear functions, which are used previously to illustrate the ratio at the level of reconstructed event yields, describe the ratio of the normalised differential cross sections less well – in particular at the edges of the analysed kinematic range.

The parametrisation is applied to different CMS results at different centre-of-mass energies and $t\bar{t}$ final states. The results for $\sqrt{s} = 7$ TeV are taken from [25] and the results for $\sqrt{s} = 8$ TeV are taken from [26] and this thesis for the semileptonic final state and from [27] for the dileptonic final state. The parameters of the exponential parametrisation are listed in Table 12.1 while the data over prediction ratios are shown in Figure 12.9. Consistency is found for the results obtained from the previous measurement with $\approx 12\,\text{fb}^{-1}$ and the full $\sqrt{s} = 8$ TeV dataset presented in this thesis. Additionally, the results of the semileptonic and dileptonic $t\bar{t}$ final states are compatible and similar results are obtained for the measurements at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV.

In order to correct the top-quark kinematics for $t\bar{t}$ events, the fact that each event contains two top quarks has to be taken into account. Hence, an *event weight* (WGT)

CMS Dataset ($\sqrt{s}$)	Final States considered	a	b
7 TeV	l+jets only	0.174	-0.00137
7 TeV	dilepton only	0.222	-0.00197
7 TeV	all combined	0.199	-0.00166
8 TeV (PAS)	l+jets only	0.130	-0.00116
8 TeV (PAS)	dilepton only	0.128	-0.00121
8 TeV (PAS)	all combined	0.130	-0.00118
8 TeV	l+jets only	0.129	-0.00114
8 TeV	combined with dilepton PAS	0.133	-0.00118

Table 12.1: Parameters of the p_T^t correction factors $\exp(a+bp_T^t)$ for the MadGraph+pythia prediction based on CMS data of different centre-of-mass energies and final states.

is computed from the convolution of the two SFs:

$$\text{WGT} = \sqrt{\text{SF}(p_T^t) \cdot \text{SF}(p_T^{\bar{t}})}. \tag{12.2}$$

The final correction factors are intended to correct only the shape of the p_T^t distribution because no information about the total normalisation is included in the normalised differential cross sections. Consequently, it has to be ensured that the total number of $t\bar{t}$ events is not affected by the correction factors. This is achieved by dividing them by the average weight,

$$\langle \text{weight} \rangle = \frac{1}{N} \sum_{\text{event i}=1}^{N} \text{wgt(event i)}, \tag{12.3}$$

which is found to be very close to one.

The uncertainty of the fit is statistical only and hence, not expected to cover several aspects like the choice of the fitting function, the choice of the theory for the BCCs, the correlation of the data points, or the uncertainty of the MadGraph+pythia prediction. A conservative estimate of the uncertainties is to apply no correction and twice the correction as down and up variation:

- default: eventWeight = WGT
- down (-1σ): eventWeight = 1 = WGT(a=0,b=0)
- up ($+1\sigma$): eventWeight = WGT·WGT = WGT(2·a,2·b)

The corresponding uncertainty bands are illustrated in Figure 12.10 for the correction factors obtained from the combined (semileptonic and dileptonic) CMS dataset at $\sqrt{s} = 8$ TeV.

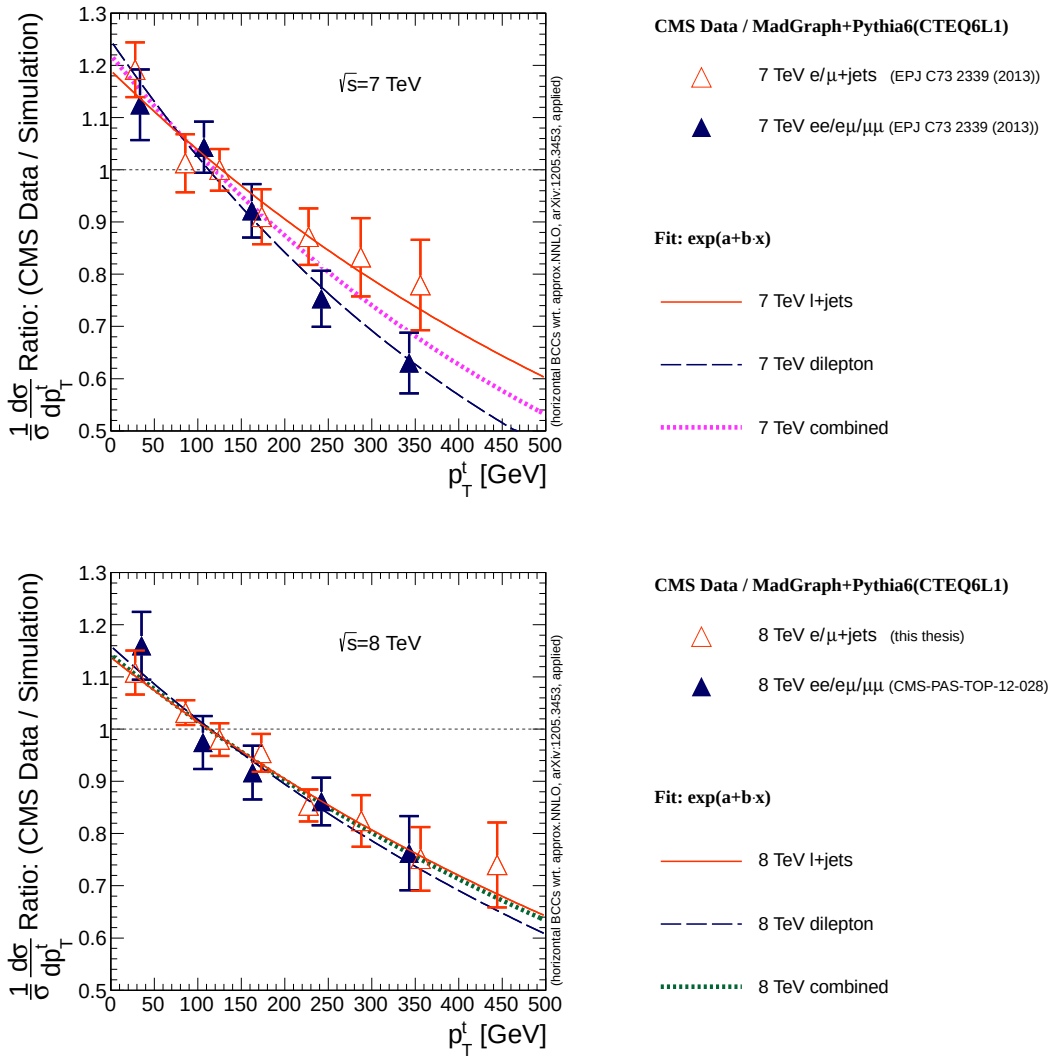

Figure 12.9: Ratio of the normalised differential cross section as function of the top-quark transverse momentum (p_T^t) as obtained from the data analysis and the MadGraph+Pythia MC prediction. Horizontal bin centre corrections are applied to all points using a perturbative QCD prediction at approx. NNLO precision. CMS results for $\sqrt{s} = 7$ TeV [25] are shown in the top row and CMS results for $\sqrt{s} = 8$ TeV ([27] and this thesis) in the bottom row.

Finally, the correction factors and uncertainties can be used to test the sensitivity of every analysis with respect to the modelling of the p_T^t spectrum.
One drawback of the approach discussed so far is that the two SFs for the top quarks must be combined first. The combination leads to an averaging of the correction factors, which is less accurate. Therefore, correction factors are derived as function of the momentum of the top-quark in the rest frame of the $t\bar{t}$ system, which is unique for every $t\bar{t}$ event. The same exponential parametrisation is used and the result is shown in Figure 12.11.

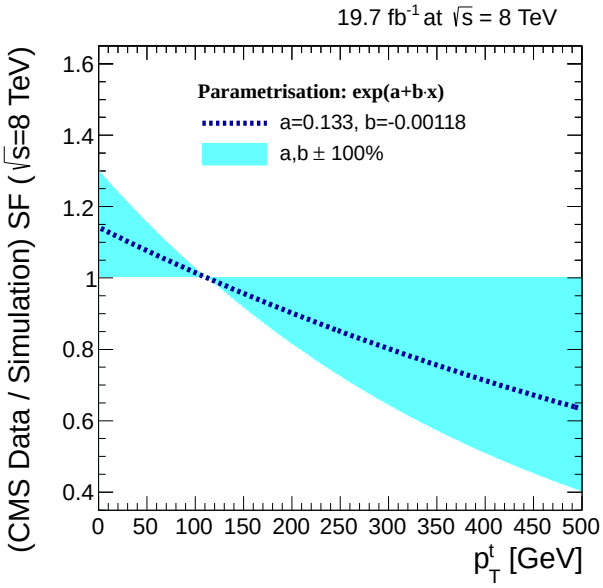

Figure 12.10: Correction factor for the shape of the p_T^t distribution with a conservative uncertainty approach for the combined semileptonic (this thesis) and dileptonic [27] CMS results at $\sqrt{s} = 8$ TeV.

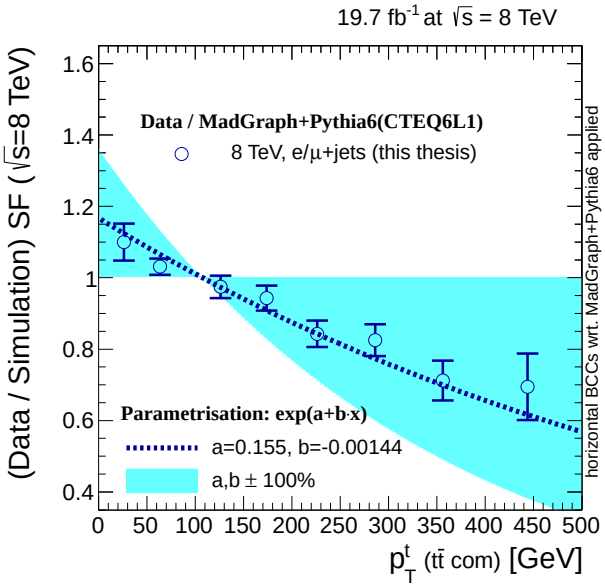

Figure 12.11: Correction factor for the shape of the p_T^t ($t\bar{t}$ com) distribution with a conservative uncertainty approach for the $\sqrt{s} = 8$ TeV CMS results of this thesis.

12.3.2 Application, Discussion and Alternatives

While the effect of a different p_T^t spectrum in the normalised differential cross section measurement is tested with pseudo data (Chapter 9.4.2), the impact on the reconstruction level event yields is shown in Figure 12.12 to 12.15 for the p_T^t ($t\bar{t}$ com) correction approach.

Using the correction factors, an improved agreement with data is observed for all momentum distributions that are directly and indirectly related to p_T^t. Among those quantities are e.g. p_T (jets), p_T (lepton) or the scalar sum of all jet momenta (H_T). Some of these quantities are also relevant for BSM searches or the search for the production of the Higgs boson in association with a top-quark pair. Furthermore, also the description of $\Delta\phi\,(t,\bar{t})$ is improved. The consistently improved agreement with data for these quantities contributes to a consistent picture and suggests that large parts of the initially observed difference between data and prediction can be explained by a mismodelling of the p_T^t ($t\bar{t}$ com) distribution.
In contrast to this, the agreement with data is worsened for other quantities like e.g. $m^{t\bar{t}}$, $m^{b\bar{b}}$ or ρ_S. It is a priori not evident if the initial better agreement was achieved accidentally or if the worsened agreement with applied correction factors is a non-physical consequence of this empirical approach.

As a matter of fact, this procedure is only an effective correction which does not really correct for the origin of the difference. This is one of the reasons why a rather conservative approach for the uncertainties is chosen. One drawback, which can result in significant uncertainties when evaluating the dependence on the modelling of p_T^t with the described procedure, is that the fractions of the $t\bar{t}$ production mechanism (gg, qq and higher order contributions like qg) are changed within the procedure. This implies a change for all properties depending on the production mode like e.g. the $t\bar{t}$ spin correlation. Moreover, the available energy of the partonic subprocess is different for $t\bar{t}$ production involving gluons and quarks in the initial state due to the different proton PDFs. Therefore, also the invariant mass of the produced $t\bar{t}$ system can differ for $t\bar{t}$ production from gluon and quark initial states. This is a possible reason for the observed change in $m^{t\bar{t}}$ when applying the correction factors.

The correction factors, which are derived from the $\sqrt{s} = 7$ TeV CMS measurements and the preliminary results of the $\sqrt{s} = 8$ TeV CMS measurements are by now widely used within the CMS collaboration to correct the MC prediction and evaluate a corresponding uncertainty. The evaluation follows the procedure developed within this thesis. In some CMS analyses, the uncertainty related to the modelling of the p_T^t distribution is even found to limit the precision of the measurement. Examples are the determination of the top-quark mass from the B-hadron lifetime [214] or the measurement of the $t\bar{t}$ spin correlation [215]. Furthermore, these correction factors are also used in analyses searching for the production of a Higgs boson in association with a $t\bar{t}$ pair [216].

To correct for the difference between MC prediction and data without implicitly changing other properties of the $t\bar{t}$ production and decay, one has to understand the origin of this difference. As a first step, it is recommended to clarify if the difference

is just a higher order QCD effect like suggested by the improved agreement with the approx. NNLO prediction or the more pronounced difference for the top quark with lower momentum. To further study this hypothesis, a differential prediction for p_T^t at full NNLO+NNLL accuracy is needed. If this is the final explanation, an NNLO $t\bar{t}$ generator would be needed to deliver an appropriate description. In this case, the above introduced p_T^t correction procedure will remain the best tool to test the impact of the difference on the performed analyses before it will become available.
Alternatively, the POWHEG+HERWIG MC prediction can be used instead of MADGRAPH+PYTHIA. This involves a consistent treatment of all uncertainties and data-driven correction factors based on the POWHEG+HERWIG prediction. However, understanding the origin of the difference between POWHEG+HERWIG and the other MC predictions in detail is an important step in interpreting the measured p_T^t distribution.

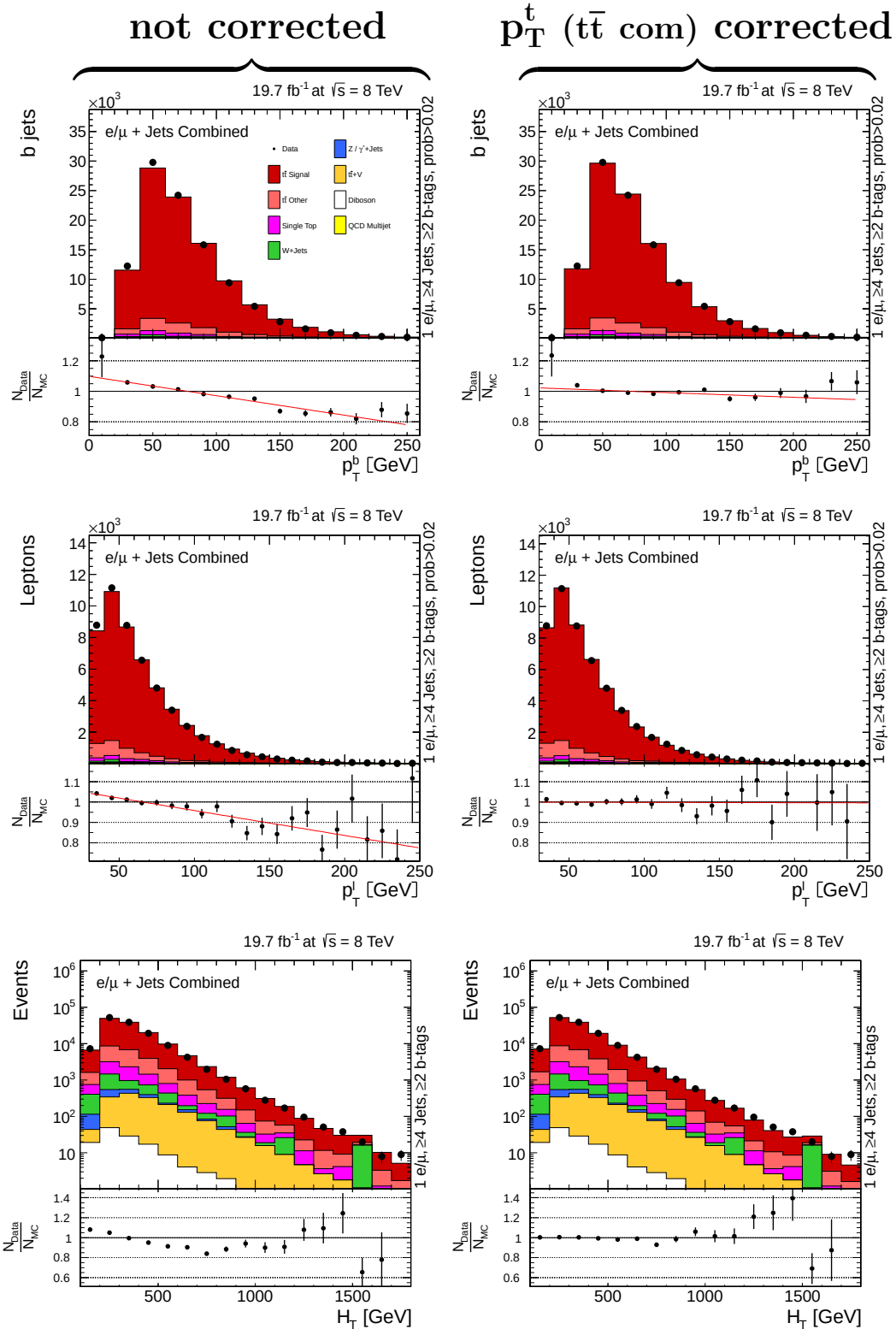

Figure 12.12: Reconstruction level event yields for p_T^b (top), p_T^l (middle) and H_T(bottom) as predicted by MadGraph+pythia **without (left) and with (right) p_T^t correction factors** applied compared to data. The red line indicates a linear fit to the ratio of data and MC prediction.

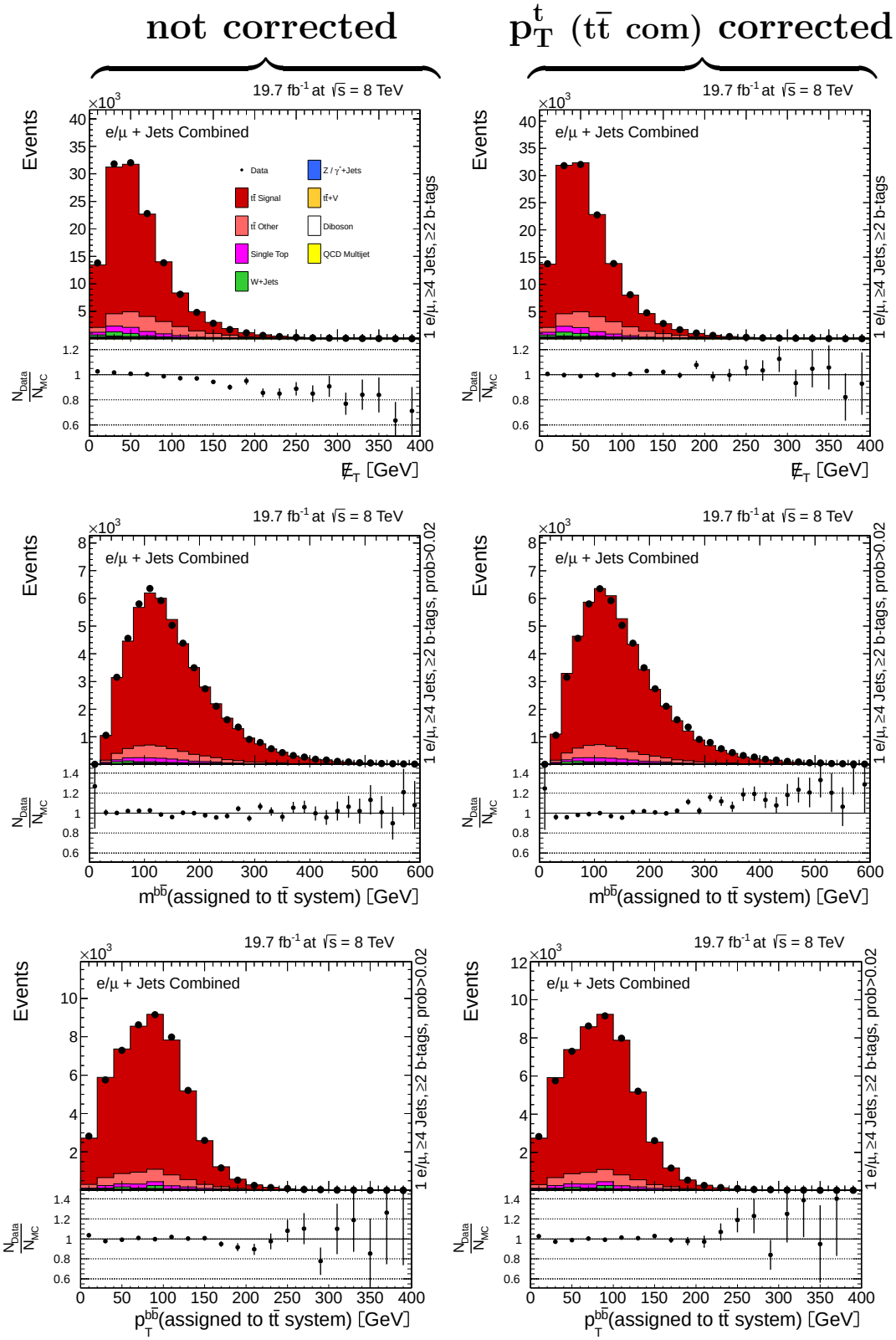

Figure 12.13: Reconstruction level event yields for E_T^{miss} (top), $m^{b\bar{b}}$ (middle) and $p_T^{b\bar{b}}$ (bottom) as predicted by MadGraph+pythia **without (left) and with (right) p_T^t correction factors** applied compared to data.

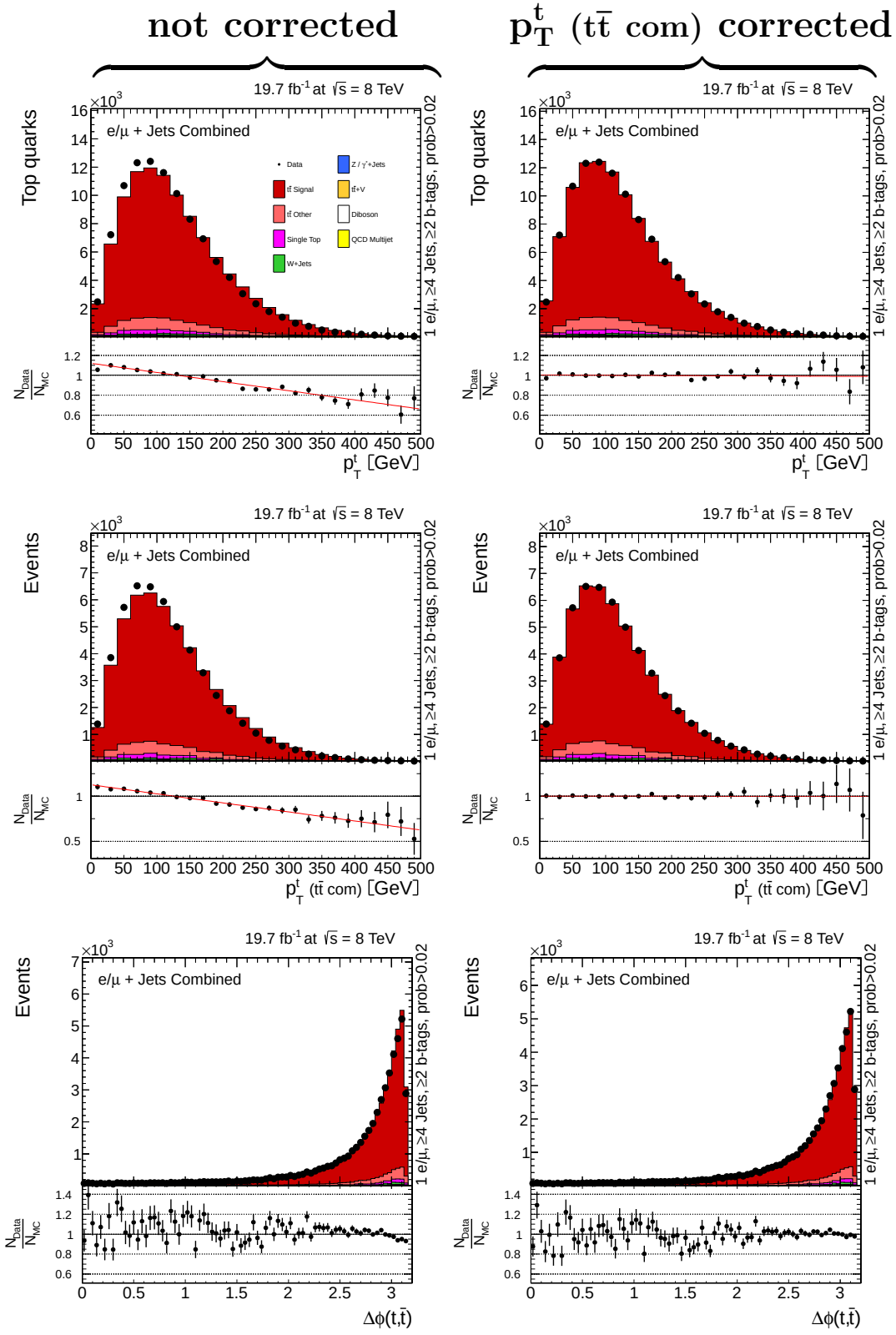

Figure 12.14: Reconstruction level event yields for p_T^t (top), p_T^t ($t\bar{t}$ com) (middle) and $\Delta\phi\,(t,\bar{t})$ (bottom) as predicted by MadGraph+pythia **without (left) and with (right) p_T^t correction factors** applied compared to data. The red line indicates a linear fit to the ratio of data and MC prediction.

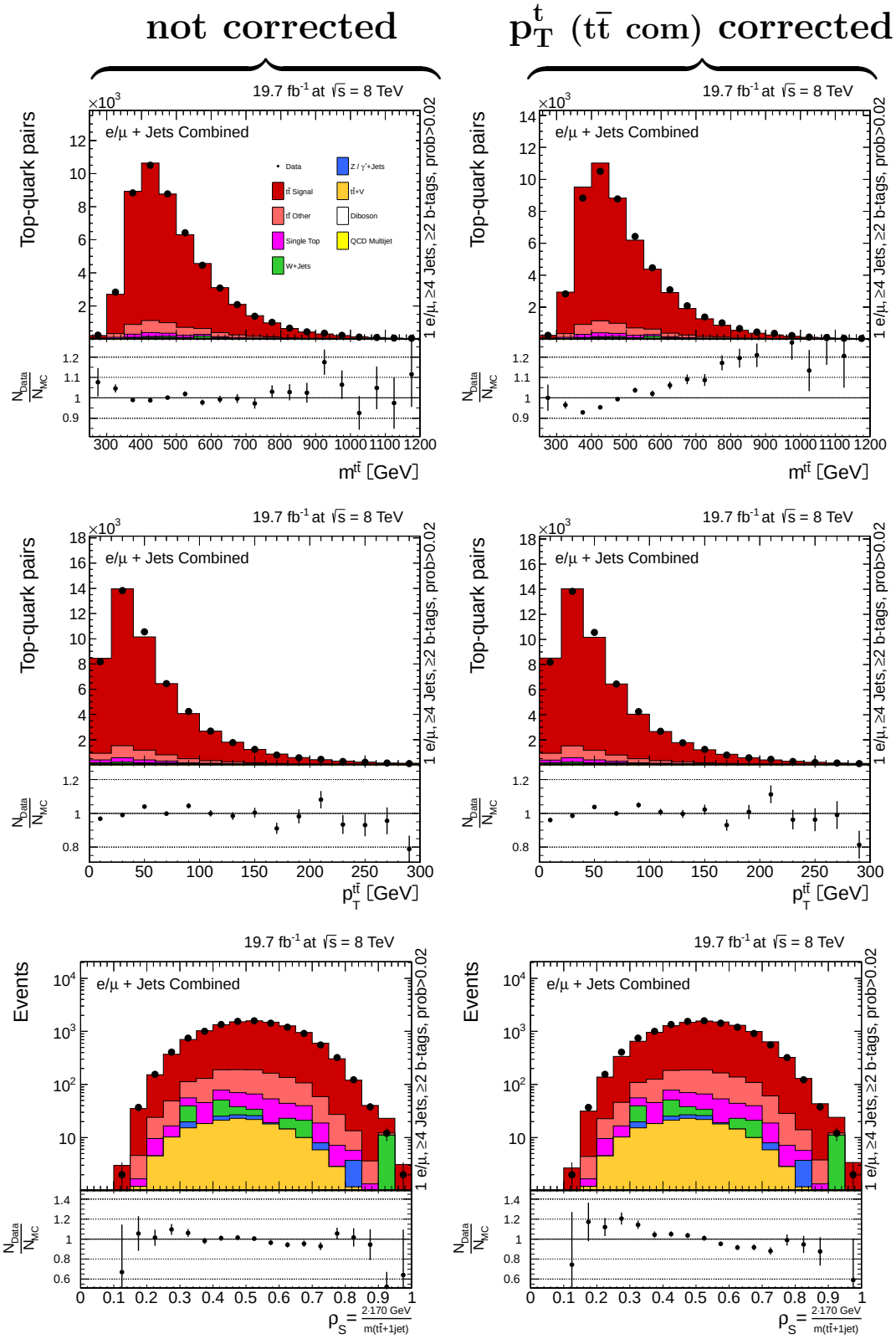

Figure 12.15: Reconstruction level event yields for $m^{t\bar{t}}$ (top), $p_T^{t\bar{t}}$ (middle) and $\rho_S = \frac{2 \cdot 170\,GeV}{m_{t\bar{t}j}}$ (bottom) as predicted by MadGraph+pythia **without (left) and with (right) p_T^t correction factors** applied compared to data.

Chapter 13

Top-Quark Mass Extraction from Differential Cross Sections

The top-quark mass (m_{top}) is a fundamental parameter of the SM and relevant in many aspects (see Chapter 3). As the top quark does not exist as a free particle, its mass cannot be directly measured. Different concepts to define the top-quark mass are detailed in Chapter 3.3. The to date most precise results for m_{top} [169–175] are obtained for the invariant mass of all decay products using an event reconstruction similar to the one used within this thesis. These measurements will be in the following referred to as direct measurements.

All direct measurements of m_{top} involve a calibration so that the finally extracted value follows the definition of the mass parameter in the MC simulation. The difference between this definition and the top-quark pole mass is still under discussion but expected to be in the order of 1 GeV or below [53, 177] (see also Chapter 3.3.1).
An alternative method is the measurement of m_{top} from the total inclusive cross section for $t\bar{t}$ production ($\sigma_{t\bar{t}}$) [217], which allows to extract directly a value that is well defined within a perturbative QCD renormalisation scheme. In comparison to direct measurements, which achieve precisions of less than 0.5%, the extraction of m_{top} from $\sigma_{t\bar{t}}$ is less accurate (about 2%).

The top-quark mass can also be extracted from (normalised) differential cross sections. This approach is only applicable with the large $t\bar{t}$ dataset recorded at the LHC which allows for precise differential measurements. No official results for m_{top} using this approach are published to date. Several of the quantities addressed in this thesis are sensitive to the choice of the top-quark mass, e.g. m^{lb} or $m^{t\bar{t}}$ (which starts at $2 \cdot m_{top}$ in the LO picture). The indirect extraction of m_{top} from a very sensitive differential distribution might be a good alternative to the existing measurements. These indirect measurements offer the possibility to extract also directly the top-quark pole mass and might turn out to be of similar precision as the direct measurements.

13.1 The Observable ρ_S

Another quantity sensitive to m_{top} that is motivated by theory studies [49] is the dimensionless observable ρ_S, which is reconstructed from the invariant mass of the $t\bar{t}$ system and an additionally produced jet as explained in 3.2.3. The extraction of m_{top} from the normalised differential cross section as function of ρ_S is discussed is the following.

This measurement can be used to validate the standard measurements of m_{top} if the sensitivity is high enough. Moreover, the extraction of a theoretically well defined top-quark mass is possible. For this, the phase space of the normalised differential cross section measurement as function of ρ_S needs to be defined such that also theory calculations are possible for exactly this definition.

The difficulty when defining a theoretically accessible phase space for the measurement of ρ_S is that the definition of ρ_S itself involves different kinds of objects. The $t\bar{t}$ pair is in this thesis defined within the parton level phase space (see Chapter 8.1.1) while jets are defined within the visible particle level phase space involving acceptance requirements (see Chapter 8.1.2). Furthermore, it is necessary to identify all jets stemming from the $t\bar{t}$ decay in order to identify the additionally produced jet. The mixture of jets and quarks in the definition of ρ_S is no problem when using MC predictions but might be a problem for direct perturbative QCD calculations.

One advantage of the definition of ρ_S is that it involves only the higher order $t\bar{t}$ +1 jet process. As mentioned in [49], this process can be directly calculated at NLO QCD precision so that remaining uncertainties are reduced with respect to LO calculations. In principle, also a higher order MC prediction can be used to perform the extraction of m_{top} from the measured normalised differential cross section.

In this thesis, a first sensitivity study for the extraction of m_{top} from $\frac{1}{\sigma_{t\bar{t}}}\frac{d\sigma}{d\rho_S}$ is presented. Not having a fully differential QCD prediction for $\rho_S(m_{top})$ available, the m_{top} extraction is studied using MC predictions. As a first step, the MadGraph+pythia MC prediction is used because here all necessary systematic variations exist.
In a first approach, m_{top} is determined from the unfolded normalised differential cross section presented in Chapter 11.2.2. Synchronously to all other results containing jets in their definition, ρ_S is determined in the visible particle level phase space. The generator truth definition of ρ_S is constructed from the top quarks at parton level and the additional gen jet with the highest momentum ($p_T > 50\,\text{GeV}, |\eta| < 2.4$) that is not identified as jet from the top-quark or W-boson decay of the $t\bar{t}$ pair. The minimum requirement of 50 GeV for the additional jet follows [49]. Details about the definition of the phase space and the definition of the additional jet can be found in Chapter 8.1.2 while the definition of the top quarks is detailed in Chapter 8.1.1. The measurement of the normalised differential cross section for ρ_S is performed using the same criteria as for all other quantities.
In principle, the unfolding is not needed when using only MC predictions for the extraction of the top-quark mass. As separating the measurement of the unfolded cross section (using MC predictions) and the extraction of the top-quark mass (using a theory

prediction like $\rho_S(m_{top}^{pole})$) is one of the key ideas of this method, the unfolding step is still done for this study.

Studying the sensitivity of this ρ_S definition with respect to the choice of m_{top}, the measured normalised differential cross section is compared to predictions of MadGraph+pythia that are generated with different m_{top} parameters. The result is shown in Figure 13.1.

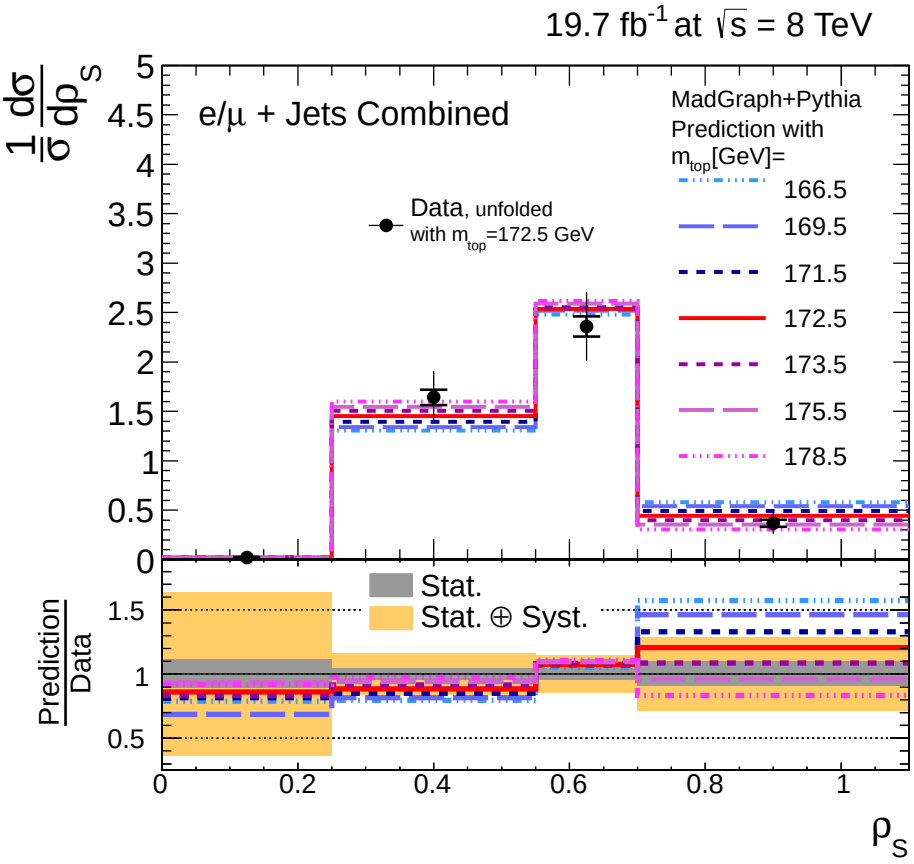

Figure 13.1: Measured normalised differential cross section as finction of ρ_S in the visible particle level phase space compared to the prediction of the MadGraph+pythia MC prediction with different top-quark mass values.

In agreement with [49], the most sensitive regions with respect to the choice of m_{top} are for low and high values of ρ_S. A lower value for m_{top} results on average in larger values for ρ_S. Also this is consistent with the studies presented in [49]. Unfortunately, the result in the low ρ_S region (first bin) in data has an uncertainty which is much larger than the sensitivity on m_{top}. The reason for this large uncertainty are pronounced migration effects, which are already discussed in Chapter 8.4. Furthermore, the third bin has only a weak sensitivity with respect to the choice of m_{top}. This is a consequence of the normalisation which leads often to one bin of the distribution, mostly the bin with the highest content, that is hardly affected by variations. The most precise results

for an extraction of m_{top} are expected from the second and last bin.

The uncertainty for the measured normalised differential cross section in these bins is about 15% and 30% respectively, predominantly limited by a large systematic uncertainty that is a factor of three larger than the statistical uncertainty. One of the main uncertainties arise from JES and JER. The uncertainties related to JES and JER are larger than for all other normalised differential measurements performed within this thesis due to the additional jet involved in the definition of ρ_S. The same is true for the $t\bar{t}$ modelling uncertainties (hadronisation, Q^2 scale and ME-PS matching threshold). These uncertainties are also larger than for other cross section quantities because the production of an extra jet involves additional uncertainties due to the involved higher order processes.

13.2 χ^2 Approach for the Extraction of the Top-Quark Mass

In Chapter 7.2.5, the impact of the choice of the top-quark mass constraint in the applied Double Kinematic Fit event reconstruction procedure is discussed. The fixed value of 172.5 GeV in the first step of the Double Kinematic Fit is not a problem for the extraction of the top-quark mass if the measurement is repeated for several predictions assuming a different value of m_{top} and the actual m_{top} dependence of the measured cross section is taken into account.

To obtain a description of the normalised differential cross section of ρ_S as function of m_{top}, all available MadGraph+pythia $t\bar{t}$ predictions with different values of the m_{top} parameter are used. Taking the m_{top} dependence of the unfolded result correctly into account, the measurement is repeated for every prediction with a different m_{top} parameter. Unlike for the other cross section quantities, the uncertainty related to the choice of m_{top} is not added to the uncertainty of the measured cross section. Instead, the predicted value as well as the measured value for $\frac{1}{\sigma}\frac{d\sigma}{d\rho_S}$ is parametrised as a function of the assumed top-quark mass. A continuous parametrisation $f(m_{top})$ is obtained from a fit to the seven mass points. The parametrisation is derived separately for data ($f_{data}(m_{top})$) and prediction ($f_{pred}(m_{top})$) for each bin of the normalised differential cross section of ρ_S:

$$f_{pred}^{\text{bin i}}(m_{top}) = \left(\frac{1}{\sigma}\frac{d\sigma}{d\rho_S}\right)_{MC}^{\text{fit(bin i)}}(m_{top}) \quad (13.1)$$

$$f_{data}^{\text{bin i}}(m_{top}) = \left(\frac{1}{\sigma}\frac{d\sigma}{d\rho_S}\right)_{data}^{\text{fit(bin i)}}(m_{top}). \quad (13.2)$$

A linear function $f(m_{top})$ is used for the parametrisation in all cases. Analogously to the extraction of m_{top} from the inclusive $t\bar{t}$ cross section [217], different m_{top} dependencies are obtained for measurement and prediction. This enables the extraction of of m_{top}. The measured result for m_{top} is more precise the larger the difference in the slopes of $f_{pred}(m_{top})$ and $f_{data}(m_{top})$ is.

Predicted and measured normalised differential cross sections for different values of m_{top} are shown in Figure 13.2 for every bin. For the measurement, the uncertainty band contains the combined statistical and total systematic uncertainty evaluated for $m_{top} = 172.5\,\text{GeV}$. For the prediction, the uncertainty band contains the major $t\bar{t}$ modelling uncertainties (Q^2 scale and ME-PS matching threshold). As systematically varied MC predictions are only present for the central mass point, all systematic uncertainties are derived for $m_{top} = 172.5\,\text{GeV}$. Furthermore, these uncertainties are assumed to be constant within the relevant range of m_{top}. Assuming a constant absolute uncertainty σ_{tot}^{sys} avoids the d'Agostini bias [218]. Assuming a constant relative uncertainty, values of m_{top} with smaller cross section contributions would be unnaturally preferred for the last bin of ρ_S.

To extract the most probable top-quark mass, a χ^2 expression ($\chi^2_{\text{bin i}}$) is build for

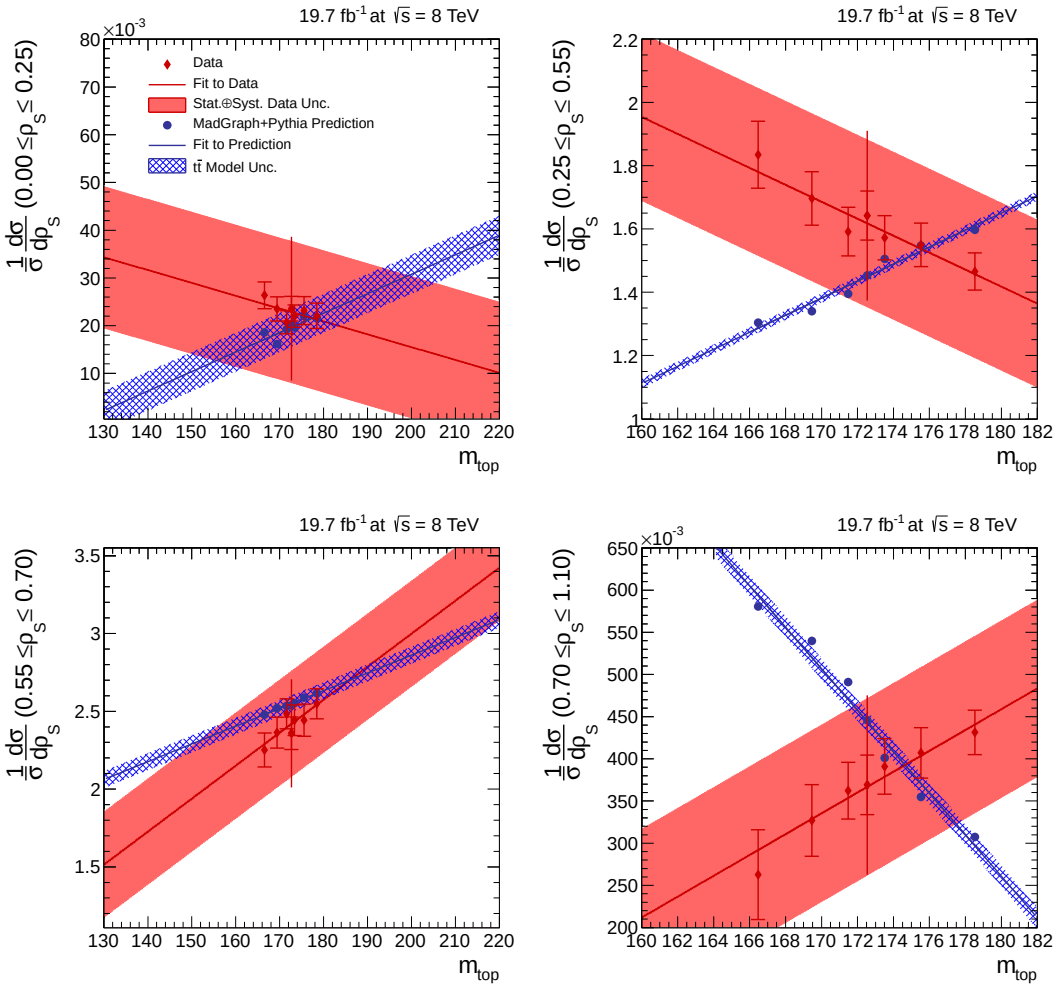

Figure 13.2: Predicted and measured normalised differential cross section $\left(\frac{1}{\sigma}\frac{\mathrm{d}\sigma}{\mathrm{d}\rho_\mathrm{S}}\right)$ for different assumed top-quark masses using the MadGraph+pythia MC prediction in all four bins of the ρ_S measurement. The unfolded data is shown as red diamonds and the prediction as blue points. The fitted functions are shown as lines in the respective colour. In addition, uncertainty bands are shown. The data point for $m_\mathrm{top} = 172.5\,\mathrm{GeV}$ contains statistical and systematic uncertainties while the other data points contain only statistical uncertainties.

every bin i from the parametrised normalised differential cross sections of the MC prediction and the measurement using the total uncertainties of the measurement

$$\delta_\mathrm{data}(m_\mathrm{top}) = \sqrt{\left(\delta^\mathrm{sys}_\mathrm{data}|_{m_\mathrm{top}=172.5\,\mathrm{GeV}}\right)^2 + \left(\delta^\mathrm{stat}_\mathrm{data}(m_\mathrm{top})\right)^2}$$

and the uncertainty of the prediction $(\delta_\mathrm{pred}|_{m_\mathrm{top}=172.5\ \mathrm{GeV}})$:

$$\chi^2_{\mathrm{bin\ i}}(m_\mathrm{top}) = \frac{\left(f_\mathrm{pred}(m_\mathrm{top}) - f_\mathrm{data}(m_\mathrm{top})\right)^2}{\left(\delta_\mathrm{data}(m_\mathrm{top})\right)^2 + \left(\delta_\mathrm{pred}|_{m_\mathrm{top}=172.5\mathrm{GeV}}\right)^2}. \tag{13.3}$$

m_{top} is extracted from the minimum of the χ^2 distribution and the one sigma uncertainties are derived from $\Delta\chi^2 = 1$ assuming normal distributed errors. The corresponding χ^2 distribution and the extracted top-quark masses for bin two and four of ρ_S are shown in Figure 13.3. Due to construction, the minimum of the χ^2 distributions is zero. The other bins do not allow for a measurement with comparable precision.

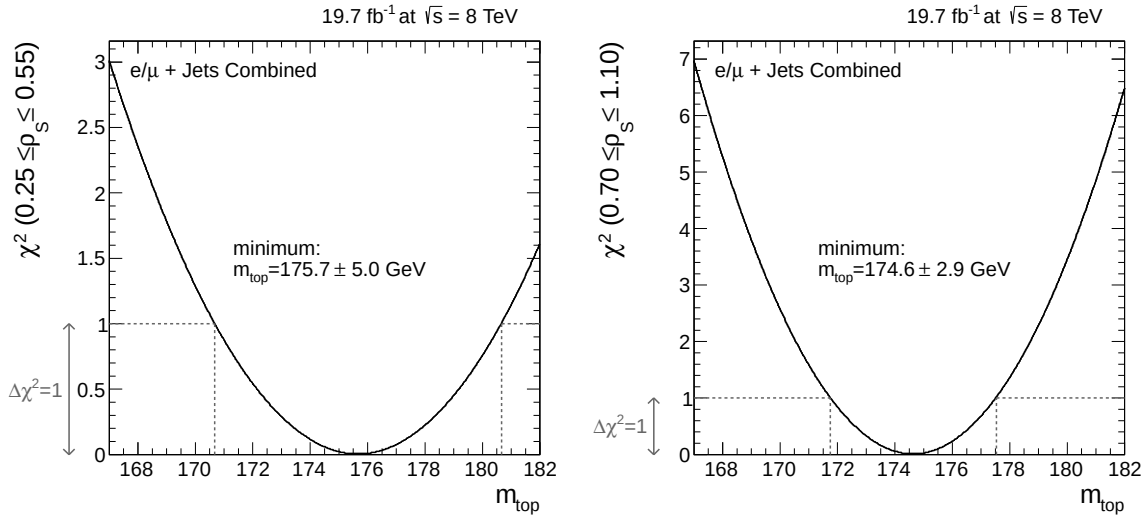

Figure 13.3: χ^2 distributions for the separate extraction of the top-quark mass from the most significant bins of the normalised differential cross section as function of ρ_S.

A symmetric χ^2 distribution is obtained for all bins due to the linear parametrisations. For the most significant last bin, a value of $m_{top} = 174.6\,\text{GeV}$ is obtained with a precision of 1.7%.

Combining the information of the most significant bins, a global χ^2 distribution (χ^2_{tot}) is constructed. In order to account for correlations between the different bins, the total statistical and systematic covariance matrix of the measurement (COV_{data}) as determined with the method discussed in Chapter 10.3 is considered:

$$\chi^2_{Tot} = \sum_{\text{bin i,j}}^{\text{i,j}\in(2,4)} \Delta x_i \cdot \left(COV^{-1}_{data}\right)_{ij} \cdot \Delta x_j \tag{13.4}$$

$$\Delta x_i = f^i_{data}(m_{top}) - f^i_{pred}(m_{top}). \tag{13.5}$$

The uncertainty of the prediction is neglected because it is small compared to the uncertainty of the measurement in the considered bins. The extracted result from the global χ^2 approach is shown in Figure 13.4. The obtained result of 174.5 GeV and the achieved precision of 1.6% are almost identical to the result of the most significant single bin.

Using the full covariance matrix assumes that the correlation between the bins is precisely known. Therefore, the impact of the assumed correlation on the extracted

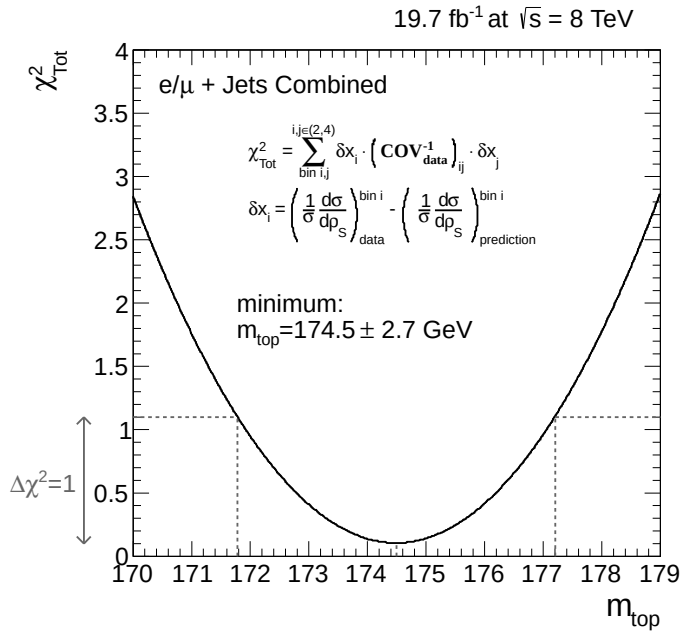

Figure 13.4: χ^2 distribution for simultaneous extraction of the top-quark mass from the most significant bins of normalised differential cross section as function of ρ_S.

result for m_{top} is studied in detail.

At first, m_{top} is extracted using a simple approach ($\chi^2_{Tot,simple}$) without correlations for all bins. In this case, Equation 13.4 is simplified:

$$\chi^2_{Tot,simple}(m_{top}) = \chi^2_{tot}|_{\rho_{ij}=0} = \sum_{\text{bin } i=1}^{4} \chi^2_i(m_{top}). \tag{13.6}$$

The obtained result is shown in Figure 13.5. It is comparable to the default approach with reduced uncertainties. This already suggests that the impact of the correlations on the measured value of m_{top} value and the achieved precision in not negligible.

Additionally, the result is extracted only from the two most significant bins assuming different correlations. In this case, the covariance matrix (COV_{24}) can be written as function of the absolute total uncertainties (σ_2 and σ_4) and the correlation ρ_{24} between the bins:

$$COV_{24} = \begin{pmatrix} \sigma_2^2 & \rho_{24}\sigma_2\sigma_4 \\ \rho_{24}\sigma_2\sigma_4 & \sigma_4^2 \end{pmatrix}$$

The inverse covariance matrix is then:

$$COV_{24}^{-1} = \begin{pmatrix} \frac{1}{\sigma_2^2\cdot\left(1-\rho_{24}^2\right)} & \frac{1}{\sigma_2\sigma_4\cdot\left(\rho_{24}-\frac{1}{\rho_{24}}\right)} \\ \frac{1}{\sigma_2\sigma_4\cdot\left(\rho_{24}-\frac{1}{\rho_{24}}\right)} & \frac{1}{\sigma_4^4\cdot\left(1-\rho_{24}^2\right)} \end{pmatrix}$$

The nominal correlation determined with the methods explained in Chapter 9.3.1 and 10.3 and listed in Appendix E.3 is $\rho_{24} = -0.72$. Probing the impact of the assumed

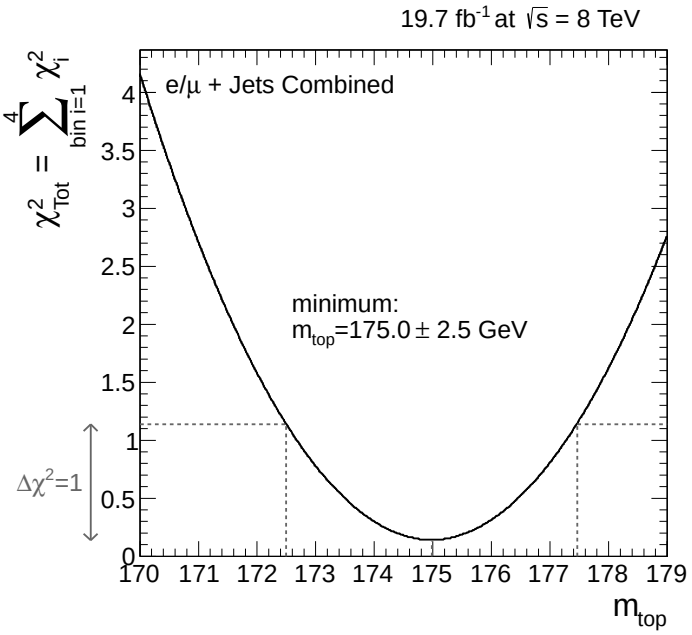

Figure 13.5: χ^2 distribution for simultaneous extraction of the top-quark mass from all bins of the normalised differential cross section as function of ρ_S when neglecting all correlations.

correlation ρ_{24} it is varied from -1 to +1 in steps of 0.2 with additional steps around the nominal value. The results for the extracted value of m_{top} for the minimum of Equation 13.4 and the uncertainty derived from $\Delta\chi^2 = 1$ are shown in Figure 13.6.

For the entire range of $-1.0 \leq \rho_{24} \leq 1.0$, the extracted value of m_{top} is stable at the level of ± 1 GeV. Higher values of m_{top} are obtained for large positive correlations and lower values for large negative correlations. The precision of the extracted m_{top} is stable at the level of 0.1% for $-0.8 \leq \rho_{24} \leq -0.4$. For assumed correlations that are very different to the nominal value, the uncertainty of the extracted result for m_{top} is significantly reduced. This means, that the uncertainty calculated assuming $\rho_{24} = -0.72$ is not underestimated. The result for this assumed correlation is

$$m_{top} = 174.5 \pm 2.7\,\text{GeV},$$

including statistical and systematic uncertainties.

The extracted result for m_{top} for each bin and for the simultaneous extraction of the two most significant bins is compared to the world average value [172] in Figure 13.7. The measured top-quark mass for the single bins of ρ_S and the combined value are in good agreement with the world average value.

In summary, a first extraction of the top-quark mass from the differential cross section of ρ_S is performed to study the expected sensitivity of the method. The obtained result is in agreement with other measurements and the achieved precision is of the order 2.5 GeV ($\approx 1.5\%$). This is approximately as precise as the top-quark mass extracted

from the inclusive $t\bar{t}$ cross section but still a factor of three less precise than the most precise direct measurements of m_{top} [169, 173, 174]. In comparison to the investigations in [49], higher experimental uncertainties are observed. This might be due to the number of considered uncertainty sources, the exact definition of ρ_S, the difference in the experimental cuts applied and methods used as well as due the choice of the MC event generator used for the estimate. MadGraph+pythia is used in this thesis while the more precise prediction of powheg for $t\bar{t}$+1 jet@NLO ME precision matched to pythia for the PS is used for the investigations in [49].

A calibration of the obtained value and uncertainties is needed to turn this study into a complete measurement. This would involve creating pseudo data sets with different top-quark masses and comparing the measured result with the true value.

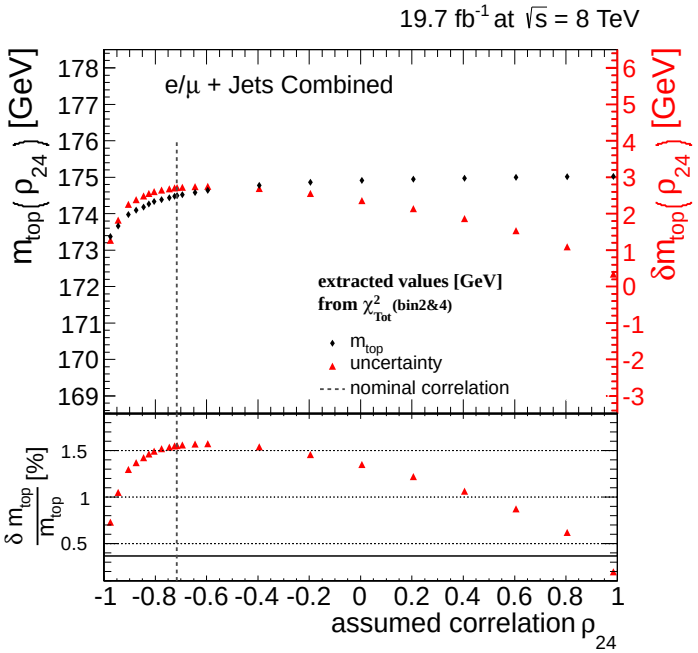

Figure 13.6: Extracted value and uncertainty of m_{top} from the most significant bins of the normalised differential cross section as function of ρ_S assuming different values for the correlation ρ_{24} between these bins.

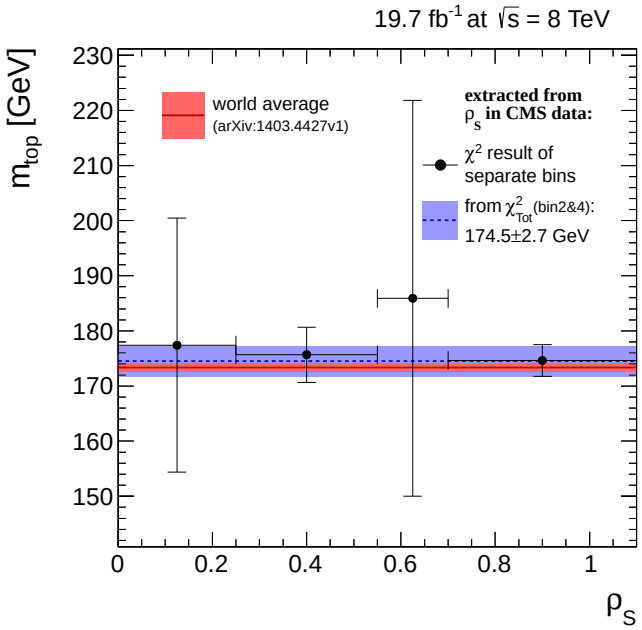

Figure 13.7: Extracted value and uncertainty of m_{top} for each bin of the normalised differential cross section as function of ρ_S and for the simultaneous extraction from the two most significant bins.

13.3 Possible Further Improvements

A first extraction of the top-quark mass from the normalised differential cross section as function of ρ_S $\left(\frac{1}{\sigma}\frac{d\sigma}{d\rho_S}\right)$ was presented to study the expected sensitivity of the method. The analysis presented here is optimised for the normalised differential cross section measurements. Hence, the precision of the determined top-quark mass can be improved by further optimising the analysis specifically for this measurement.

With the method introduced in Chapter 10.3, the correlation between the bins of $\frac{1}{\sigma}\frac{d\sigma}{d\rho_S}$ is calculated for every single uncertainty source. This information can be used to perform the χ^2-based extraction of the most probable top-quark mass separately for every MC prediction that is used to evaluate a single source of systematic uncertainty. It might turn out that this approach leads to more accurate results than the combined χ^2 method. In any case, this approach allows to quantify the separate systematic uncertainty contributions, which is not possible with the combined method.

The large uncertainty of $\frac{1}{\sigma}\frac{d\sigma}{d\rho_S}$ also induced by a large JES uncertainty is limiting the precision of the extracted result for m_{top}. The JES uncertainty can be reduced e.g. by a two-dimensional approach to determine the JES in-situ from the reconstructed W-boson mass peak (see e.g. [169, 219]). Within this approach, the JES calibration is done for light jets and only uncertainties due to the difference in JES between gluon and light jets and between b jets and light jets remain.
For a significant improvement in precision, the main issue is to reduce the model uncertainty for the ρ_S measurement. This can be possibly achieved by using higher order MC predictions (e.g. $t\bar{t}$+1 jet at NLO QCD precision), which are already available. Once a reduction of these uncertainties is achieved, one has to additionally check the uncertainties stemming from other systematic variations like colour reconnection, which are usually relevant in top-quark mass measurements. They are considered to be negligible for the moment but might become important when reaching a precision in the order of 1 GeV.

In the presented measurement, the top-quark mass is determined using MC predictions. The main challenge to extract a top-quark mass which is "theoretically well defined" within a renormalisation scheme, e.g. the $\overline{MS}$ top-quark mass, is to construct a phase space definition for ρ_S that can be accessed from both – theory and experiment. This phase space has to be on the one hand close enough to the reconstruction level $t\bar{t}$ +1 jet definition to limit uncertainties from migration effects and extrapolation. On the other hand, a prediction that is parametrising $\frac{1}{\sigma}\frac{d\sigma}{d\rho_S}$ as function of a "theoretically well defined" top-quark mass is needed. One possibility for the phase space definition is to use a visible parton level definition, involving the top quarks and the highest p_T parton above a certain threshold produced through ISR.

Another possible extension for the determination of the top-quark mass from differential cross sections is a simultaneous extraction from more than one sensitive distribution involving e.g. m^{lb} [43]. This allows maximising the input information and the reduction of uncertainties being different for the different cross section quantities.

Chapter 14

Summary, Conclusions and Outlook

Analysis Resumé

Understanding top-quark production and properties, in particular $t\bar{t}$ differential cross sections, is crucial as test of QCD predictions and searches for new phenomena beyond the SM. In this thesis, top-quark pair production at the LHC in proton-proton collisions at $\sqrt{s} = 8$ TeV is studied. The analysed dataset is recorded by the CMS experiment and corresponds to an integrated luminosity of 19.7 fb^{-1}. The analysis documented here is an extension of a previous work, which is also performed within this thesis project, and contributed largely to CMS publications at $\sqrt{s} = 7$ TeV [25, 28] and 8 TeV [26].
In the analysis presented in this thesis, the $t\bar{t}$ production cross section is measured as a function of the kinematic properties of the $t\bar{t}$ system, the top quarks and their final-state products (leptons and jets). Eighteen normalised differential $t\bar{t}$ cross sections are measured using the semileptonic ($t\bar{t} \to WWb\bar{b} \to q\bar{q}'\ell\nu_\ell b\bar{b}$) final states with one prompt muon or electron. This analysis is the most comprehensive to date for this process. In addition, the normalised differential cross sections are used to determine the top-quark mass.

An event sample with an expected purity of 95% $t\bar{t}$ events is selected from data. The four momenta of the top quarks are obtained from the measured lepton, jets and neutrino with a constrained kinematic fitting procedure. Here, the neutrino momentum is determined from the momentum imbalance in the transverse plane. An improved event reconstruction method with respect to previous CMS analyses [23, 25, 26] is introduced, involving a fitting procedure with two sequential steps. The first step is optimised to find the correct assignment of the measured final state objects to the underlying top and antitop quark while the second step is optimised to obtain the kinematic properties of the top quarks with relaxed top-quark mass constraints to avoid additional uncertainties.

The requirement of a minimal fit probability for the final measurement suppresses $t\bar{t}$ events from decay modes other than the semileptonic signal and ensures a high fraction of well-reconstructed signal events. In comparison to previous analyses in the same final state [23, 25, 26], the expected fraction of correctly reconstructed signal events is increased from 25% to over 50%. Consequently, the improved event reconstruction

and selection results in reduced migration effects and a smaller total uncertainty of the results.

The reconstructed event yields are corrected for the background contribution. Detector and migration effects are corrected using a regularised unfolding technique. The normalised differential cross sections for quantities related to the final-state lepton or jets are determined in a visible particle level phase space defined by the kinematical and geometrical acceptance of the final-state particles. These cross sections do not suffer from uncertainties due to extrapolation outside the experimentally accessible region. In contrast, cross sections as a function of top-quark and top-quark pair quantities are extrapolated to the full parton level phase space to allow for a comparison with recent QCD calculations.
Finally, the differential cross sections are normalised to the in-situ measured inclusive cross section in the corresponding phase space to reduce correlated systematic uncertainties. The results are consistent with CMS results in the dileptonic $t\bar{t}$ final state at $\sqrt{s} = 8$ TeV [27] and results of ATLAS and CMS at $\sqrt{s} = 7$ TeV [22, 25, 210]. ATLAS and CMS measurements of the *transverse momentum of the top quarks* (p_T^t) at $\sqrt{s} = 7$ TeV agree for $p_T^t > 200\,\text{GeV}$ while slight differences are observed for $p_T^t < 200\,\text{GeV}$. All measurements are compared with predictions of the MadGraph+pythia, powheg+pythia, powheg+herwig and mc@nlo+herwig MC event generators and if available also with perturbative QCD calculations up to approx. NNLO precision.

The typical precision of the differential measurements is at the order of $3 - 7\%$, which is better than the typical precision of $5 - 10\%$ achieved for the corresponding CMS analyses at $\sqrt{s} = 7$ TeV [23, 25]. While most cross sections are described by QCD predictions within uncertainties, several quantities favour or disfavour some of these predictions.

As already observed in the published CMS analysis for $\sqrt{s} = 7$ TeV data [25], the measured p_T^t distribution is softer in data than predicted by the MadGraph+pythia, powheg+pythia and mc@nlo+herwig MC event generators while a perturbative QCD prediction at approx. NNLO precision [205] yields a better description. For $p_T^t < 200\,\text{GeV}$, the observed difference between data and MC predictions is up to 10% and not fully covered by the uncertainties of the measurement.
Moreover, the powheg+herwig MC prediction provides a good prescription of p_T^t and all other differential measurements presented in this thesis. Within the *Top-Quark LHC working group* (TopLHCWG), the discussion about the difference between powheg+herwig and the other MC predictions started [213]. Understanding this difference in detail will be an important step in interpreting the measured p_T^t distribution.

A wide range of tests ensures that the measured difference between the p_T^t distribution in data and MadGraph+pythia is not artificially introduced by the unfolding method or a mismodelling, e.g. of the pile-up distribution or the *Jet Energy Scale* (JES). Consistent with p_T^t, the transverse momentum distributions of the directly measured final-state lepton and b jets are observed to be softer than predicted by MadGraph+pythia. This difference is already observed before full event reconstruction.

In addition, the transverse momentum of the top quarks in the $t\bar{t}$ rest frame and the transverse momentum distributions of the top quark with higher and lower transverse momentum are found to be softer in data than predicted by MadGraph+pythia. The difference is observed to be slightly larger for the top quark with lower transverse momentum.

The improved description of the measured p_T^t distribution by the approx. NNLO prediction, the observed larger discrepancy between data and predictions for the top quark with lower transverse momentum and first studies regarding the difference between the powheg+pythia and powheg+herwig predictions indicate that the difference for p_T^t in data and the MadGraph+pythia, powheg+pythia and mc@nlo+herwig predictions might be due to higher order QCD effects.

The correct modelling of the p_T^t spectrum is crucial for many Higgs measurements and analyses searching for new physics phenomena. The results for p_T^t from this work are used to derive generalised correction factors for the MadGraph+pythia $t\bar{t}$ prediction, which are used in current CMS analyses [214–216]. Until the origin of the difference in p_T^t is fully understood or the powheg+herwig MC event generator becomes the new simulation standard, the introduced p_T^t correction procedure will remain the best tool to correct the simulation of the p_T^t distribution.

The dependence of the $t\bar{t}$ production cross section on the *top-quark mass* (m_{top}) allows determining this parameter from the measured normalised differential cross sections. Motivated by theoretical studies [49] the extraction of m_{top} from the dimensionless quantity $\rho_S = \frac{2 \cdot 170\,\text{GeV}}{m_{t\bar{t}j}}$ is studied. Here, $m_{t\bar{t}j}$ is the invariant mass of the $t\bar{t}$ pair and an additionally produced jet. The available MadGraph+pythia MC prediction is used for the m_{top} extraction. The obtained result of 174.5 GeV is within the achieved precision of about 1.5% compatible with other CMS measurements and the world average value. However, the uncertainty is still a factor of three higher in comparison to the most precise single m_{top} measurement [173] and a calibration of the obtained value and uncertainties is needed to turn this study into a complete measurement.

Analysis Perspectives

This work covers the most comprehensive differential $t\bar{t}$ cross section analysis to date. While several improvements with respect to previous CMS measurements have been implemented, further improvements and extensions of the analysis are possible and will be briefly discussed in the following.

The most relevant experimental uncertainty source is the JES. CMS publishes uncertainties for all jet flavours [153, 220]. Therefore, it is possible to calculate an individual JES uncertainty for the analysis-specific jet-flavour composition.
Another aspect that will increase the precision of the differential measurements, is an improved understanding of the modelling differences between herwig and pythia, if these are used for the parton showering of MC predictions. These differences are addressed within the hadronisation uncertainty and are currently one of the major uncertainty sources. This uncertainty includes not only a different fragmentation model,

but e.g. also different flavour abundances which might affect the b-jet identification.
In order to improve the precision for all differential cross section measurements significantly, it can be useful to constrain the systematic uncertainty for the modelling of the $t\bar{t}$ signal events with data. One possible approach is to use combined likelihood methods that fit theory model parameters (such as e.g. Q^2) as nuisance parameters within the unfolding procedure. Alternatively, an iterative unfolding procedure can be used, which corrects the assumed shape of the prediction in the unfolding step by the result of the previous iteration separately for each systematic uncertainty.

The precision of the performed cross section measurements in the visible particle level phase space can be improved by optimising the analysis for those quantities for example by using resolutions for particle level objects instead of parton level objects in the event reconstruction.
Using dedicated reconstruction techniques for boosted topologies [221, 222] can help to improve the efficiency for these events and extend e.g. the range of the measurement in $m^{t\bar{t}}$.
The precision of the top-quark mass measurement can be improved by further optimising the analysis specifically for this measurement. For example, the large JES uncertainties can be reduced by an in-situ calibration using the hadronically decaying W boson or the modelling uncertainties can be reduced using higher order MC predictions.

The measured normalised differential cross sections of this thesis are a valuable input for many other analyses and can be used further to improve our understanding of nature. Moreover, the provided full (statistical and systematic) covariance matrices are crucial for several of these follow-up analyses.
The sensitivity of several cross section quantities like e.g. the invariant mass of the top-quark pair [9] or the difference in azimuthal angle of the two top quarks [62] regarding new physics signatures can be used to derive limits from the presented measurements for different BSM scenarios.
The precision of the differential cross section measurements might be further improved by combining them with those from ATLAS. A combination is possible as equal top-quark definitions are used as also checked within this thesis and presented in the TopLHCWG [178].

The large size of the LHC dataset can be used to extend the analysis to double- or (for the expected next data-taking period) even multi-dimensional differential measurements. A two-dimensional measurement for $y^t$and $m^{t\bar{t}}$ is for example beneficial as input for PDF constraints [191, 192]. Other multidimensional measurements can be helpful to investigate p_T^t further.
A possible extension of the differential $t\bar{t}$ cross section measurements is using the concept of so-called "pseudo tops" as discussed in [223, 224].

Altogether, this thesis proves that normalised differential $t\bar{t}$ cross section analyses are an important piece in the puzzle of high energy physics. Their precise knowledge and correct modelling has an impact even beyond top-quark physics. Therefore, they will also remain relevant for the LHC-restart at a higher centre-of-mass energy.

Appendix A

Details on Trigger- and Lepton-Selection Efficiencies

One-dimensional efficiencies and data to MC prediction SFs obtained from the T&P study as explained in Chapter 6.7 for the corresponding trigger and lepton selection are shown for the electron (Figure A.1 and A.2) and the muon (Figure A.3). All statistical uncertainties shown refer to Clopper-Pearson confidence intervals [163].

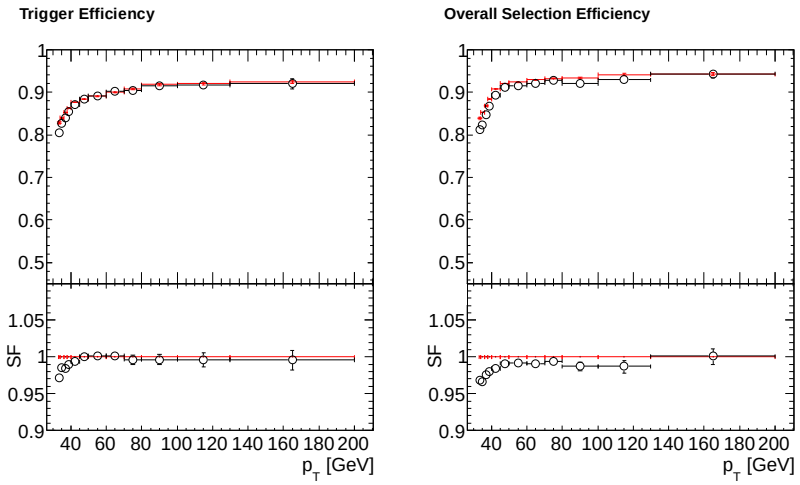

Figure A.1: The single-electron tag-and-probe trigger efficiencies (left) and selection efficiencies (right) for MC and data (upper part) and the corresponding scale factor (lower part) as a function of p_T of the electron. Red lines denote MC while black markers denote data. The shown uncertainty bars refer to the statistical uncertainties only.

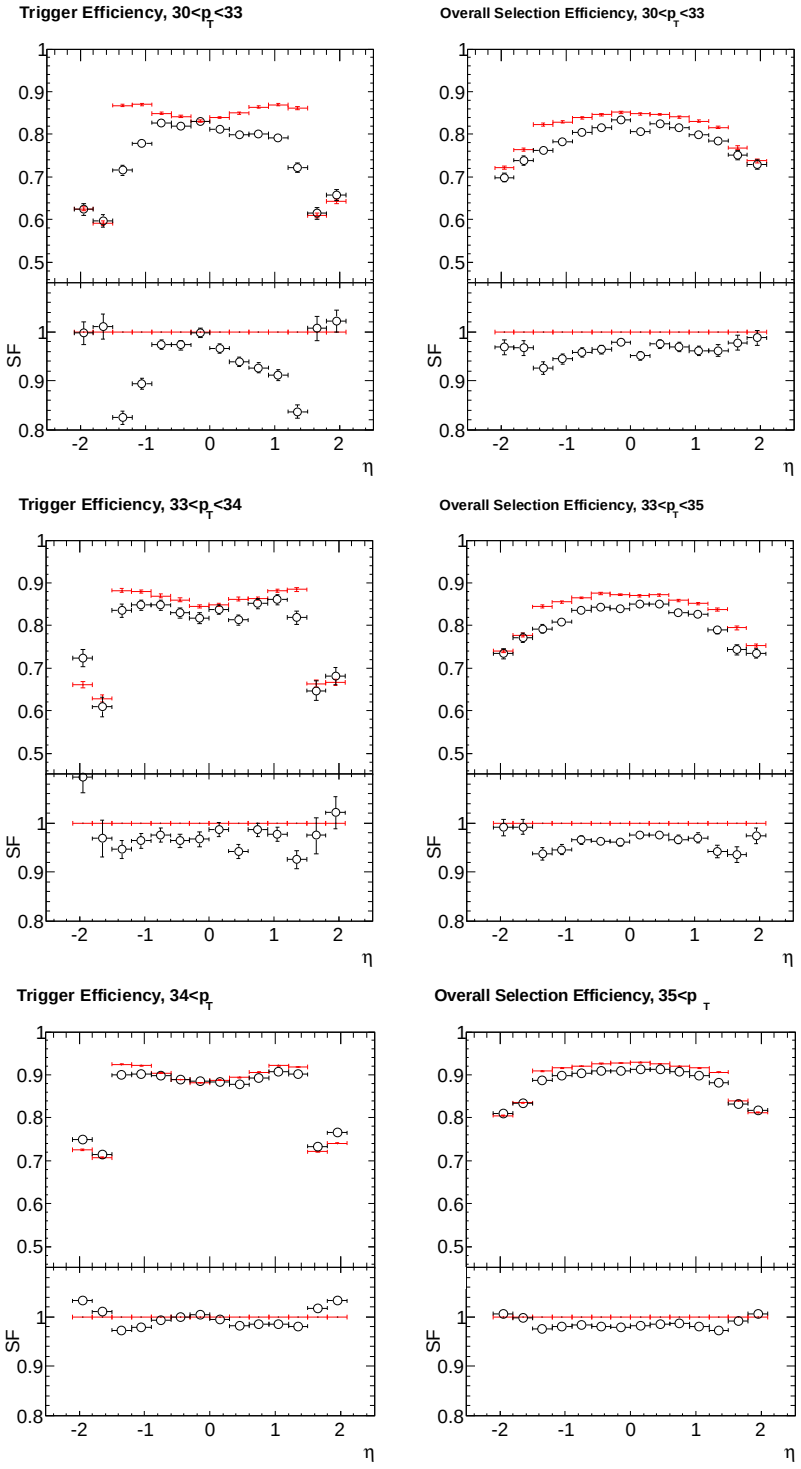

Figure A.2: The single-electron tag-and-probe trigger efficiencies (left) and selection efficiencies (right) for MC and data (upper part) and the corresponding scale factor (lower part) as a function of η of the electron in different ranges of p_T. Red lines denote MC while black markers denote data. The shown uncertainty bars refer to the statistical uncertainties only.

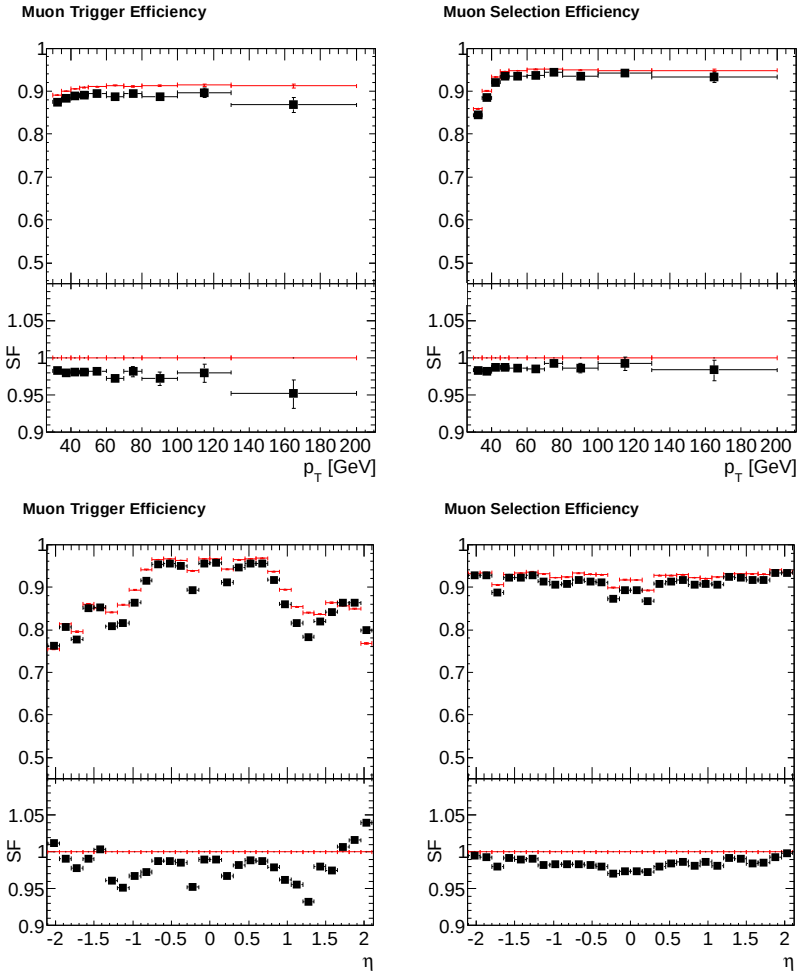

Figure A.3: The single-muon tag-and-probe trigger efficiencies (left) and selection efficiencies (right) for MC and data (upper part) and the corresponding scale factor (lower part) as a function of p_T (top row) and η (bottom row) of the muon. Red lines denote MC while black markers denote data. The shown uncertainty bars refer to the statistical uncertainties only.

Appendix B

Efficiencies of the Minimal χ^2-Probability Requirement

The minimal χ^2-probability requirement introduced in Chapter 7.2.3 is an essential part of the analysis strategy. To ensure a good understanding of the applied selection requirement of 2%, the corresponding efficiency is compared between data and MC prediction as a function of the several differential cross section quantities for the lepton, b jets, top quarks and $t\bar{t}$ system. The obtained results are shown Figure B.1 to B.3. The uncertainties shown correspond only to statistical uncertainties on the data and the MC prediction.

The differential shapes of the efficiency is well described and observed small differences are expected to be fully covered by the systematic uncertainties. A difference for the absolute efficiency is discussed in Chapter 7.3.3 and does not show up here because the efficiencies are derived from the event yields where the $t\bar{t}$ component is individually normalised to the in-situ measured inclusive cross section to focus on shape effects.

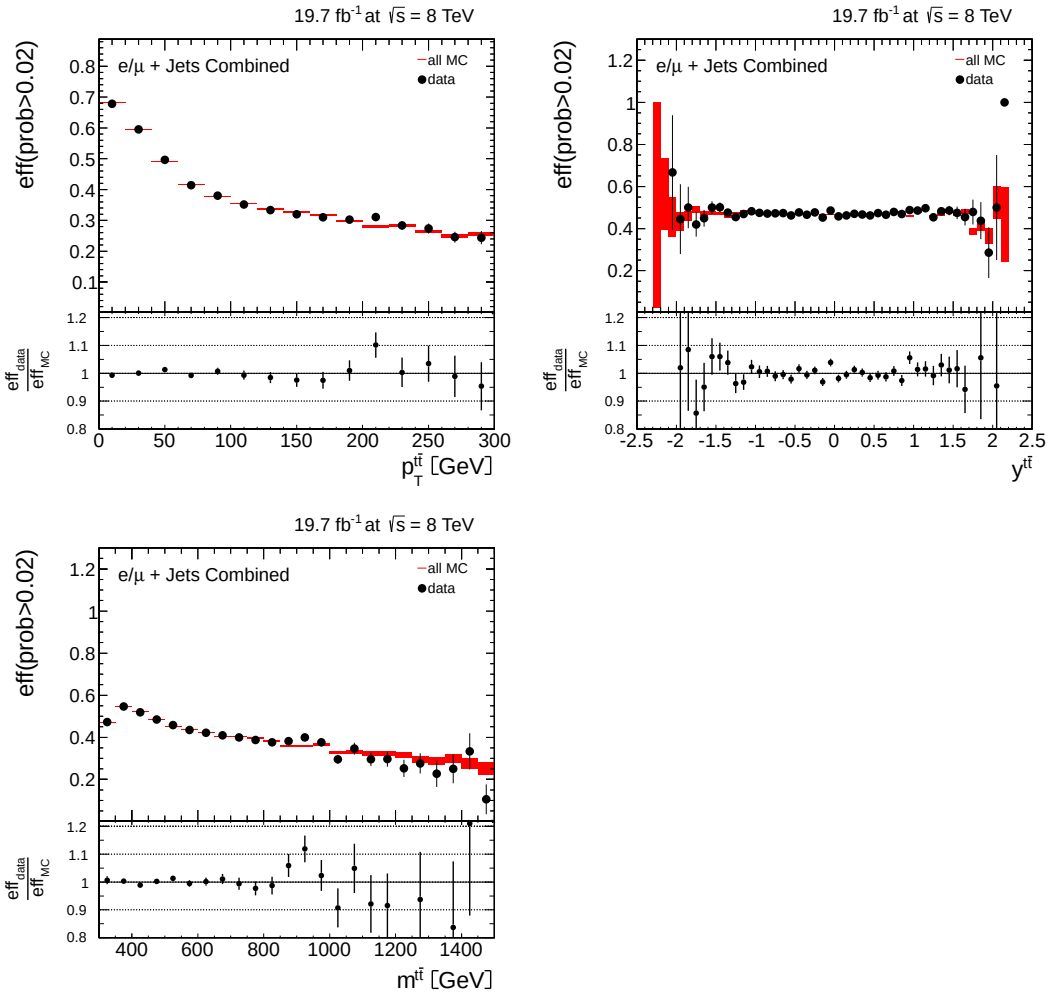

Figure B.1: Efficiency for the minimal χ^2-probability requirement in data and MadGraph+pythia MC prediction as function of the $t\bar{t}$ cross section quantities. The shown uncertainties are statistical uncertainties only.

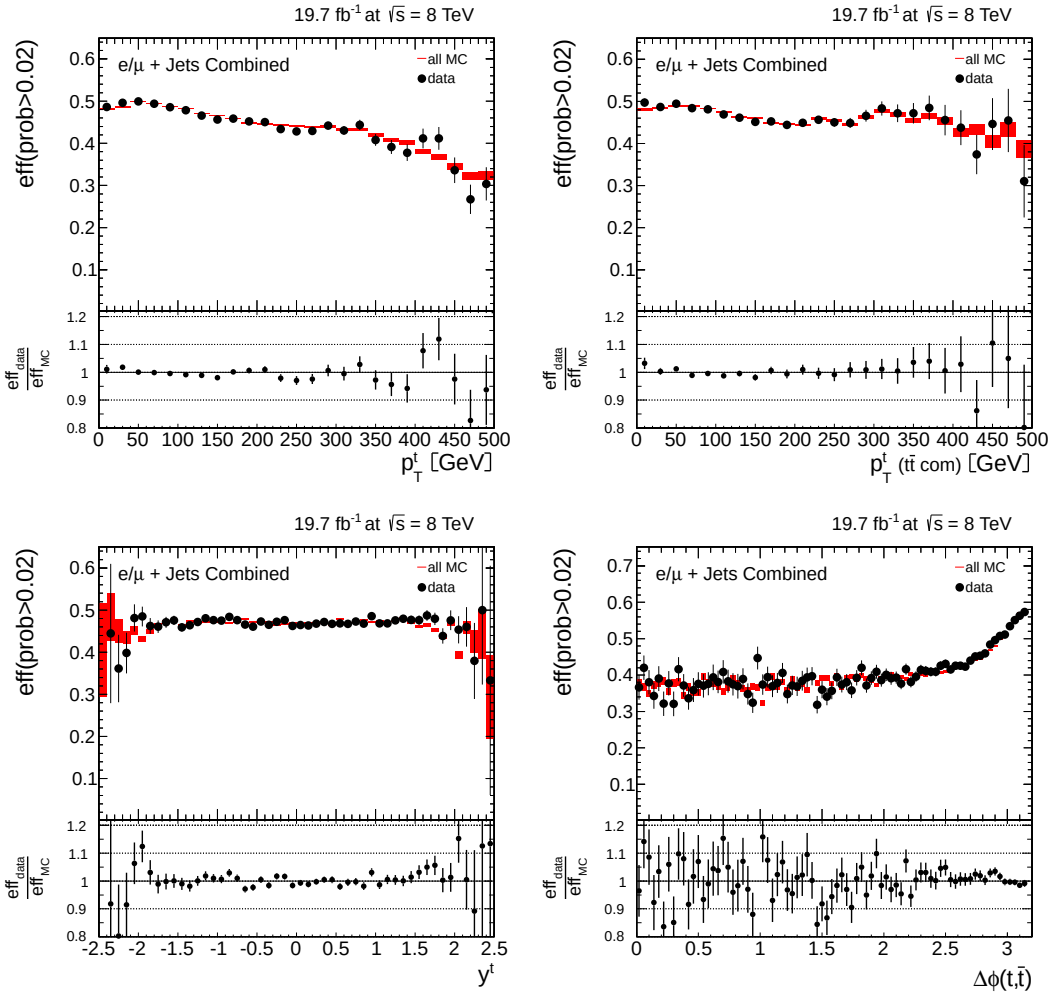

Figure B.2: Efficiency for the minimal χ^2-probability requirement in data and MC prediction as function of the top-quark cross section quantities. The shown uncertainties are statistical uncertainties only.

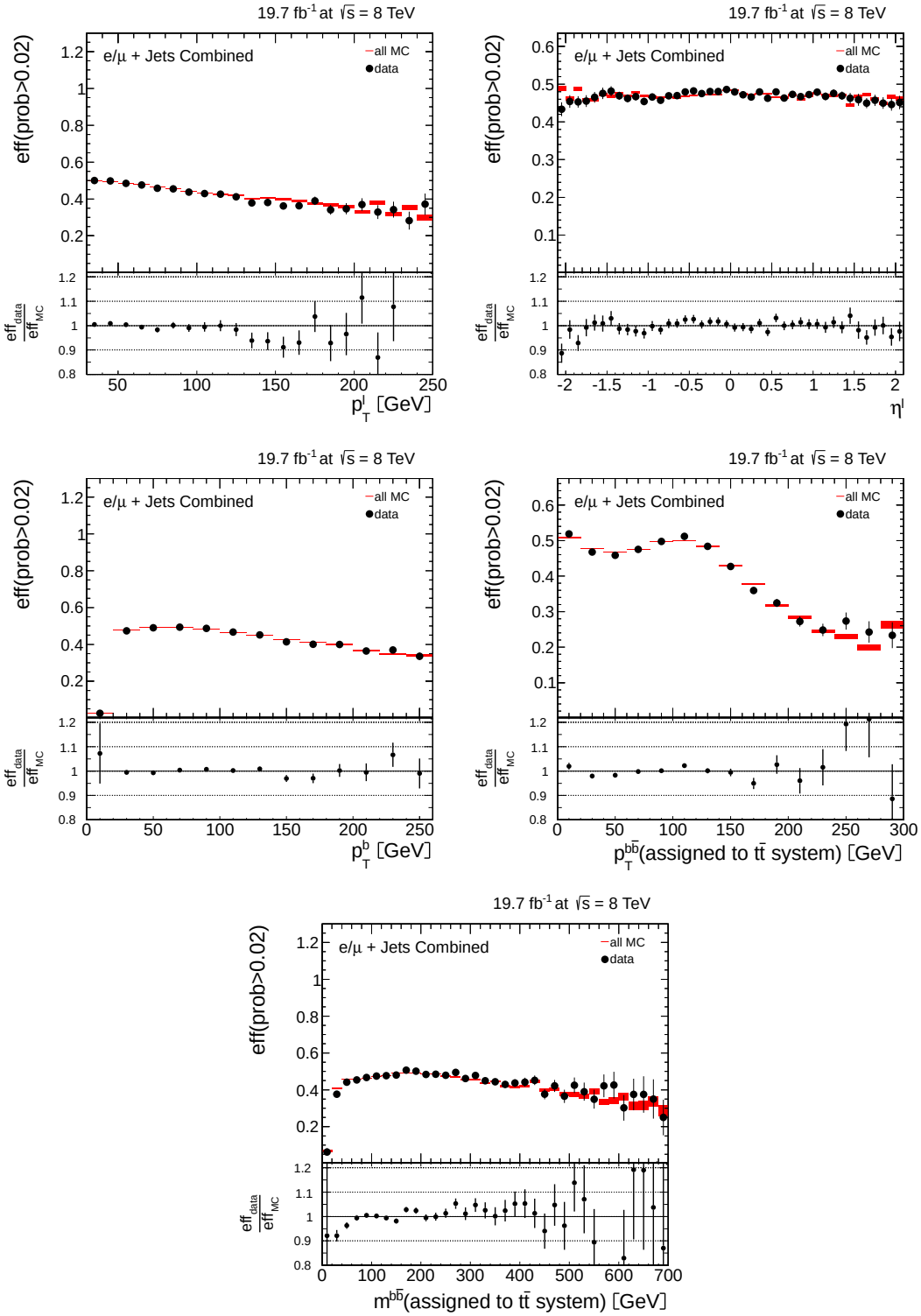

Figure B.3: Efficiency for the minimal χ^2-probability requirement in data and MadGraph+pythia MC prediction as function of the lepton and b-jet quantities. The shown uncertainties are statistical uncertainties only.

Appendix C

Migration Matrices for all Cross Section Quantities

The migration matrix is defined as:

$$A_{ij}^{migr} = \frac{N_{\text{reco in bin j}}^{\text{gen in bin i, }\in\text{phase space}}}{\sum_{\text{bin i=1}}^{N^{bins}} N_{\text{reco in bin j}}^{\text{gen in bin i, }\in\text{phase space}}}$$

For all events generated in the corresponding bin shown on the x axis, the entry in each row represents the the fraction of those events which are reconstructed in the bin shown on the y axis. Therefore, the diagonal elements are identical to the stability defined in Chapter 8.4 while far off-diagonal entries indicate long range migrations. Furthermore, the migrations from events which are not part of the defined phase space but are reconstructed and selected are not shown.

Finally, the obtained migration matrices for the other visible particle level phase are shown in for the lepton and b jet cross section quantities in Figure C.1 and for the other final state quantities in Figure C.2. Moreover, the obtained migration matrices for the extrapolated parton level phase space are shown in C.3 for cross section quantities related to transverse momentum of the top quark and in Figure C.4 for the other cross section quantities related to top quarks and the $t\bar{t}$ system.

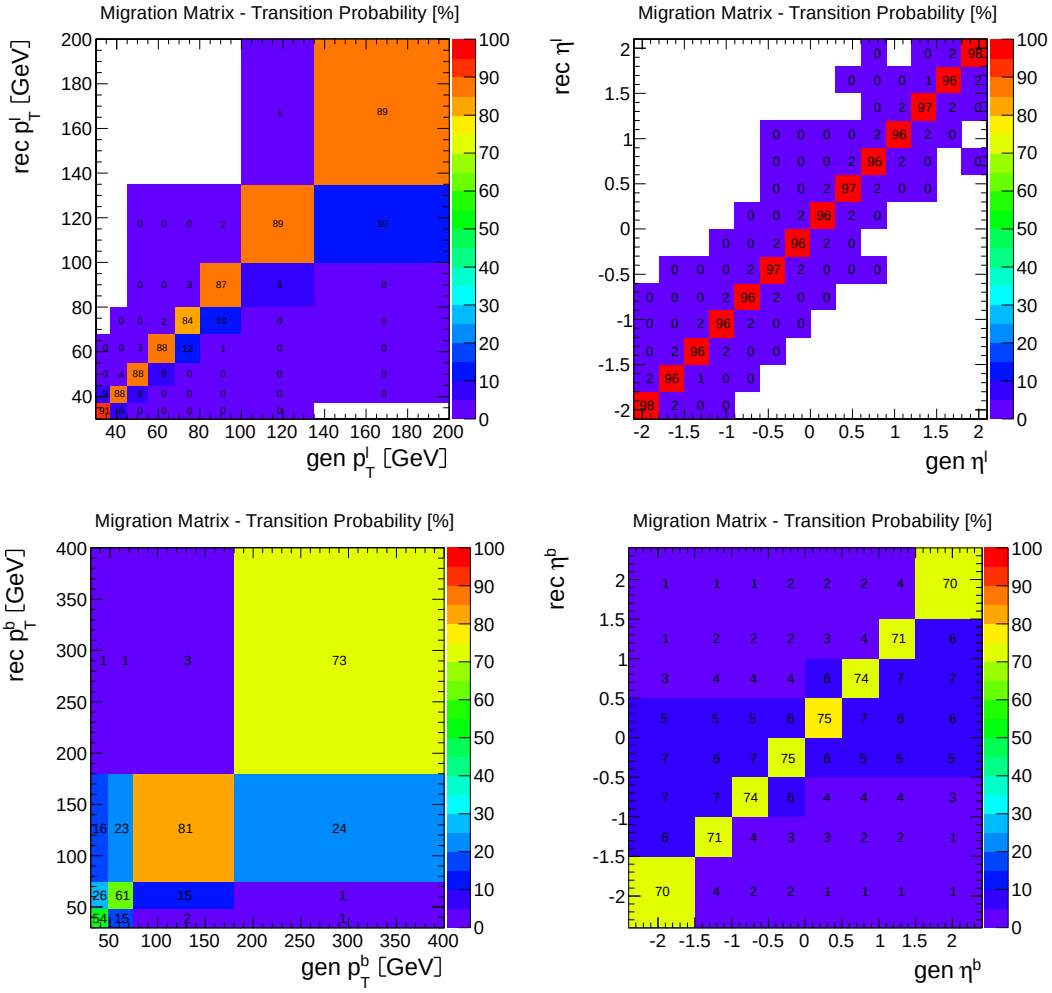

Figure C.1: Migration matrix for the cross section quantities related to leptons and single b jets in the visible particle level phase space for the combined ℓ+jets channel.

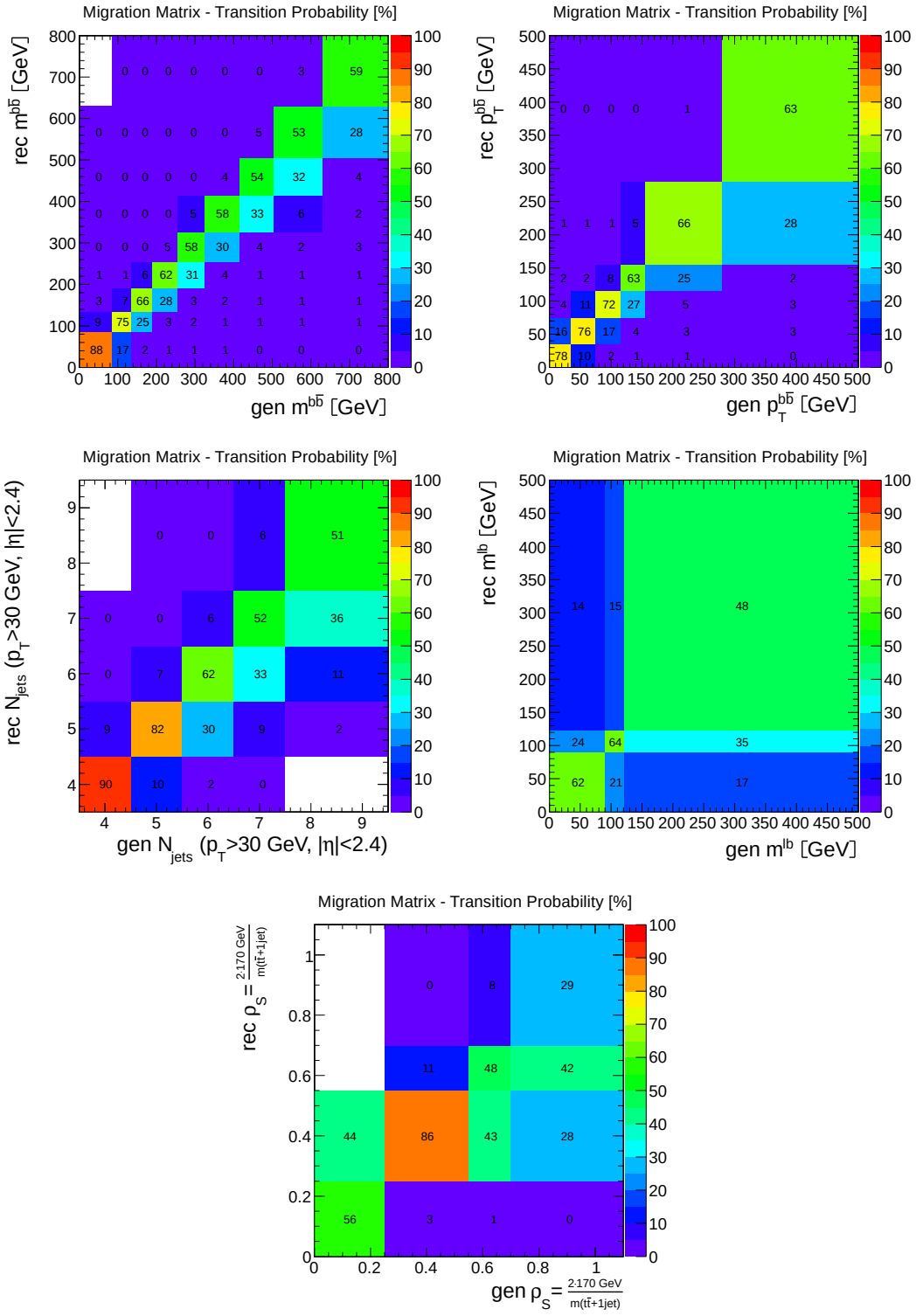

Figure C.2: Migration matrix for the cross section quantities related to the $b\bar{b}$ system, the invariant mas of the lepton and the leptonic b jet and ρ_S in the visible particle level phase space for the combined ℓ+jets channel.

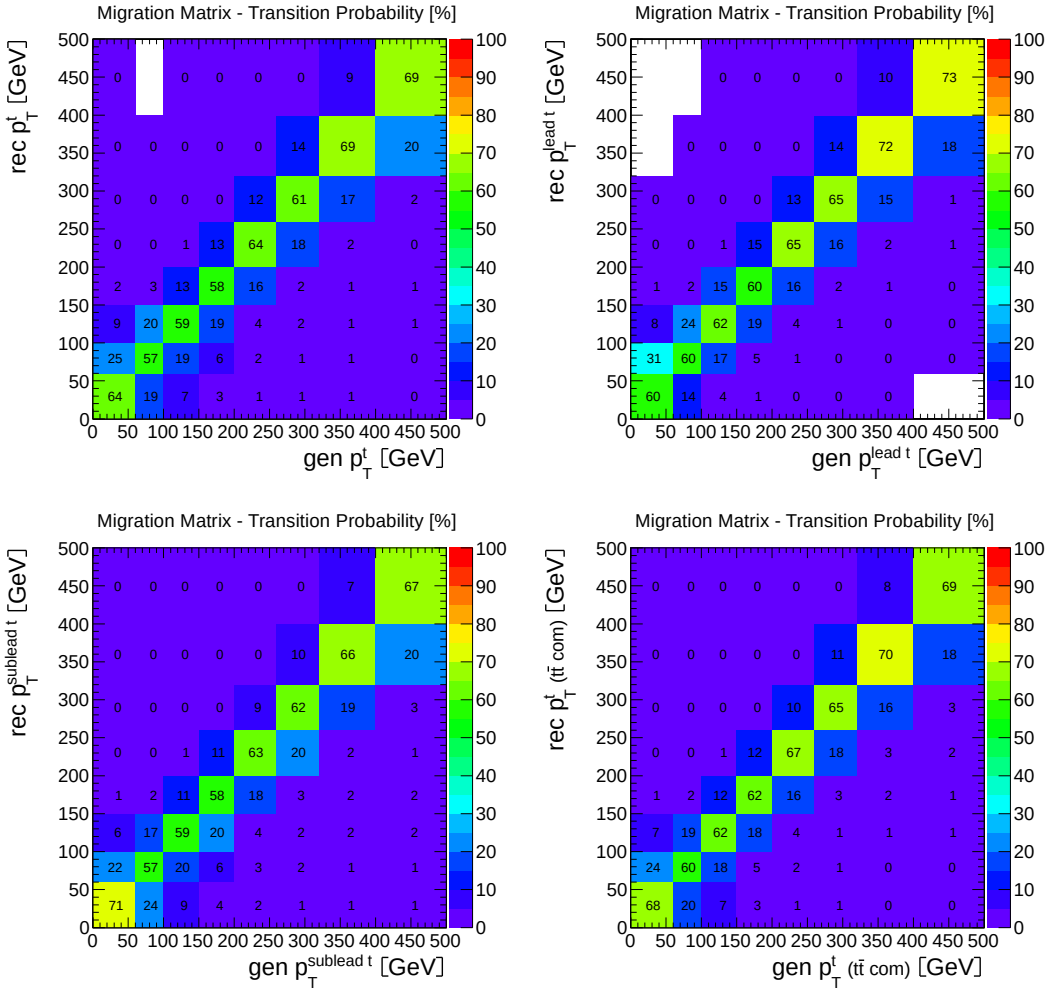

Figure C.3: Migration matrix for the cross section quantities related to single top quarks and the $t\bar{t}$ system in the extrapolated parton level phase space for the combined ℓ+jets channel.

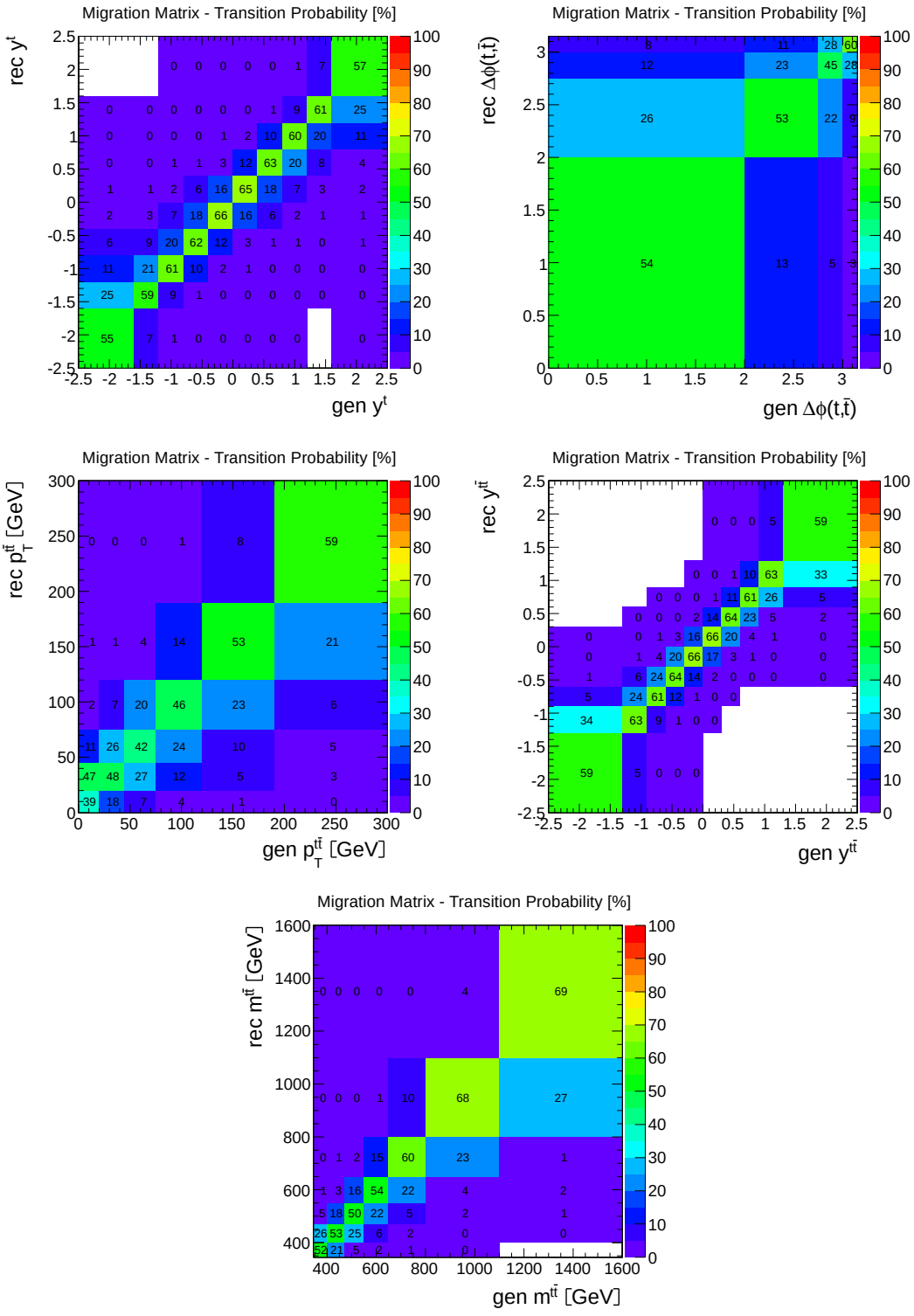

Figure C.4: Migration matrix for the cross section quantities related to single top quarks and the $t\bar{t}$ system in the extrapolated parton level phase space for the combined ℓ+jets channel.

Appendix D

Summary Tables of the Normalised Differential Cross Sections

To provide the measured normalised differential cross sections in detail as input for potential follow-up analyses, the measured value, statistical and systematic uncertainties are listed for each bin and every cross section quantity in Table D.2 to D.5. In addition to the experimental result (sub-script exp) the MadGraph+pythia MC prediction (sub-script MC), which is used to perform the measurement, is quoted too. In general, systematic uncertainties dominate over statistical ones. Only in a few cases they are approximately equal.

$\mathbf{p_T^{t\bar{t}}}$ [GeV]	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
0 to 20	0.014369	0.013230	2.3	7.9	8.2
20 to 45	0.011422	0.011840	2.1	3.6	4.1
45 to 75	0.005876	0.006398	1.8	5.5	5.8
75 to 120	0.002810	0.002844	2.3	6.6	7.0
120 to 190	0.001129	0.001066	2.6	5.8	6.3
190 to 300	0.000311	0.000306	3.8	12.3	12.9
$\mathbf{y^{t\bar{t}}}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
-2.5 to -1.3	0.066111	0.060722	3.5	3.0	4.6
-1.3 to -0.9	0.228084	0.220167	1.8	2.4	3.0
-0.9 to -0.6	0.315649	0.326531	1.8	2.9	3.4
-0.6 to -0.3	0.374468	0.372938	1.7	1.6	2.3
-0.3 to 0.0	0.406563	0.426863	1.6	1.6	2.3
0.0 to 0.3	0.407308	0.412947	1.6	1.9	2.4
0.3 to 0.6	0.374207	0.374318	1.7	1.8	2.4
0.6 to 0.9	0.315055	0.317369	1.8	1.9	2.6
0.9 to 1.3	0.227914	0.229797	1.8	2.2	2.8
1.3 to 2.5	0.066069	0.064105	3.3	3.9	5.1
$\mathbf{m^{t\bar{t}}}$ [GeV]	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
345 to 400	0.004701	0.004694	2.1	7.2	7.5
400 to 470	0.004422	0.004300	2.1	2.9	3.6
470 to 550	0.002530	0.002673	2.0	6.1	6.4
550 to 650	0.001216	0.001165	2.3	7.9	8.2
650 to 800	0.000464	0.000466	2.5	4.2	4.9
800 to 1100	0.000106	0.000114	3.0	9.6	10.0
1100 to 1600	0.000011	0.000011	8.0	9.8	12.7

Table D.1: Normalised differential cross section for the combined ℓ+jets channel in the extrapolated parton level phase space as a function of the top-quark pair observables: transverse momentum $p_T^{t\bar{t}}$, rapidity $y^{t\bar{t}}$ and invariant mass $m^{t\bar{t}}$ of the top quark pair system. The quoted uncertainties in the last three columns refer to the data result (sub-script exp). The quoted Monte-Carlo prediction (MC) is obtained from the MadGraph+pythia event generator.

$\mathbf{p_T^t}$ [GeV]	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
0 to 60	0.003732	0.004137	1.2	3.6	3.8
60 to 100	0.006487	0.006692	1.3	1.9	2.3
100 to 150	0.005063	0.004962	1.1	3.0	3.2
150 to 200	0.002786	0.002660	1.3	3.5	3.8
200 to 260	0.001238	0.001057	1.6	3.2	3.6
260 to 320	0.000484	0.000399	2.2	5.6	6.0
320 to 400	0.000173	0.000130	2.8	7.7	8.1
400 to 500	0.000050	0.000037	5.5	9.5	11.0
$\mathbf{p_T^t}$ ($\mathbf{t\bar{t}}$ **com**) [GeV]	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
0 to 60	0.004035	0.004438	1.6	4.4	4.7
60 to 100	0.006820	0.007032	1.7	1.4	2.2
100 to 150	0.005054	0.004925	1.5	2.9	3.2
150 to 200	0.002588	0.002440	1.9	3.2	3.7
200 to 260	0.001068	0.000900	2.4	3.8	4.4
260 to 320	0.000389	0.000321	3.3	4.2	5.4
320 to 400	0.000132	0.000094	4.5	6.3	7.8
400 to 500	0.000036	0.000025	9.6	9.3	13.4
$\mathbf{y^t}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
-2.5 to -1.6	0.069567	0.073563	2.5	3.9	4.6
-1.6 to -1.2	0.175211	0.175063	1.5	2.5	2.9
-1.2 to -0.8	0.249948	0.261230	1.4	3.0	3.3
-0.8 to -0.4	0.311458	0.300112	1.3	1.3	1.8
-0.4 to 0.0	0.346099	0.333246	1.2	1.5	1.9
0.0 to 0.4	0.346227	0.331009	1.2	1.6	2.0
0.4 to 0.8	0.311138	0.300043	1.3	1.8	2.2
0.8 to 1.2	0.249590	0.246602	1.4	3.0	3.3
1.2 to 1.6	0.175218	0.187734	1.4	1.6	2.1
1.6 to 2.5	0.069545	0.077701	2.4	3.4	4.2

Table D.2: Normalised differential cross section for the combined ℓ+jets channel in the extrapolated parton level phase space as a function of the following top-quark quantities: transverse momentum (p_T^t), transverse momentum in the $t\bar{t}$ rest frame (p_T^t ($t\bar{t}$ com)) and rapidity (y^t). The quoted uncertainties in the last three columns refer to the data result (sub-script exp). The quoted Monte-Carlo prediction (MC) is obtained from the MadGraph+pythia event generator.

$\mathbf{p_T^{lead\ t}}$ [GeV]	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
0 to 60	0.002349	0.002611	2.2	5.0	5.5
60 to 100	0.006059	0.006391	1.5	2.5	3.0
100 to 150	0.005587	0.005560	1.3	3.3	3.6
150 to 200	0.003416	0.003358	1.6	3.6	4.0
200 to 260	0.001621	0.001428	1.9	3.4	3.9
260 to 320	0.000663	0.000556	2.5	6.4	6.8
320 to 400	0.000243	0.000187	3.1	8.2	8.8
400 to 500	0.000072	0.000056	6.0	10.6	12.2
$\mathbf{p_T^{sublead\ t}}$ [GeV]	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
0 to 60	0.005115	0.005638	1.4	3.2	3.5
60 to 100	0.006916	0.006975	1.9	2.0	2.7
100 to 150	0.004538	0.004387	1.7	2.7	3.2
150 to 200	0.002157	0.001972	2.2	3.2	3.9
200 to 260	0.000854	0.000693	2.8	3.3	4.3
260 to 320	0.000306	0.000243	4.0	4.5	6.0
320 to 400	0.000103	0.000075	5.3	7.3	9.0
400 to 500	0.000028	0.000019	11.4	11.2	16.0
$\boldsymbol{\Delta\phi\,(t,\bar{t})}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
0.00 to 2.00	0.064357	0.068326	2.2	7.2	7.5
2.00 to 2.75	0.299154	0.321726	1.2	3.9	4.0
2.75 to 3.00	1.079666	1.125005	1.4	3.3	3.5
3.00 to 3.15	2.513355	2.272023	1.3	6.6	6.8

Table D.3: Normalised differential cross section for the combined ℓ+jets channel in the extrapolated parton level phase space as a function of the transverse momentum of the highest p_T top quark ($p_T^{lead\ t}$), the transverse momentum of the lowest p_T top quark ($p_T^{lead\ t}$) and the difference in azimuthal angle of the two top quarks ($\Delta\phi\,(t,\bar{t})$). The quoted uncertainties in the last three columns refer to the data result (sub-script exp). The quoted Monte-Carlo prediction (MC) is obtained from the MadGraph+pythia event generator.

$\mathbf{p_T^l}$ [GeV]	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
30 to 37	0.013093	0.014263	2.2	2.4	3.3
37 to 45	0.021267	0.022109	1.4	2.8	3.1
45 to 55	0.017823	0.018294	1.4	1.8	2.3
55 to 68	0.013507	0.013605	1.5	1.4	2.1
68 to 80	0.009542	0.009420	2.0	1.6	2.5
80 to 100	0.006013	0.005893	1.8	1.6	2.4
100 to 135	0.002718	0.002511	2.1	6.3	6.7
135 to 200	0.000692	0.000573	3.3	9.9	10.4
$\boldsymbol{\eta^l}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
-2.1 to -1.8	0.083295	0.097643	4.3	4.3	6.0
-1.8 to -1.5	0.129895	0.137417	3.3	3.1	4.5
-1.5 to -1.2	0.187700	0.194319	2.7	3.1	4.1
-1.2 to -0.9	0.248929	0.238727	2.5	2.8	3.7
-0.9 to -0.6	0.302304	0.303410	2.0	2.3	3.0
-0.6 to -0.3	0.345832	0.341437	1.9	1.4	2.3
-0.3 to 0.0	0.367847	0.360210	1.9	1.6	2.5
0.0 to 0.3	0.368524	0.338984	2.0	2.0	2.8
0.3 to 0.6	0.345572	0.334740	1.9	3.1	3.6
0.6 to 0.9	0.303200	0.306142	1.9	4.4	4.8
0.9 to 1.2	0.249268	0.250379	2.2	2.2	3.2
1.2 to 1.5	0.188158	0.201151	2.6	2.3	3.5
1.5 to 1.8	0.130012	0.135696	3.3	2.7	4.3
1.8 to 2.1	0.082800	0.093079	4.2	5.6	7.0
$\mathbf{p_T^b}$ [GeV]	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
30 to 48	0.012795	0.013966	1.2	7.0	7.1
48 to 75	0.011889	0.012348	0.9	1.9	2.1
75 to 180	0.003909	0.003669	0.7	4.9	5.0
180 to 400	0.000170	0.000134	3.3	12.8	13.2
$\boldsymbol{\eta^b}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
-2.4 to -1.5	0.091601	0.100452	1.7	2.9	3.3
-1.5 to -1.0	0.204017	0.208162	1.3	1.5	2.0
-1.0 to -0.5	0.288985	0.276691	1.1	1.3	1.7
-0.5 to 0.0	0.341621	0.322657	1.1	2.1	2.3
0.0 to 0.5	0.341659	0.328323	1.1	2.0	2.2
0.5 to 1.0	0.289444	0.289402	1.1	1.1	1.6
1.0 to 1.5	0.204200	0.208720	1.4	2.2	2.6
1.5 to 2.4	0.091773	0.102906	1.6	2.9	3.3

Table D.4: Normalised differential cross section for the combined ℓ+jets channel in the visible particle level phase space as a function of the transverse momentum p_T pseudo-rapidity η of the lepton (p_T^l,η^l) and the b-jet (p_T^b,η^b). The quoted uncertainties in the last three columns refer to the data result (sub-script exp). The quoted Monte-Carlo prediction (MC) is obtained from the MadGraph+pythia event generator.

$\mathbf{m^{lb}}$ [GeV]	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
0 to 90	0.002764	0.002592	2.3	10.3	10.5
90 to 122	0.009492	0.009348	1.3	4.5	4.7
122 to 500	0.001178	0.001230	1.4	7.8	7.9
$\mathbf{m^{b\bar{b}}}$ [GeV]	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
0 to 85	0.001955	0.002004	1.9	7.8	8.1
85 to 135	0.005637	0.005787	1.7	4.0	4.3
135 to 190	0.004547	0.004436	2.0	3.2	3.8
190 to 255	0.002494	0.002368	2.5	4.7	5.3
255 to 325	0.001121	0.001132	3.7	7.7	8.6
325 to 415	0.000432	0.000450	4.9	8.9	10.2
415 to 505	0.000150	0.000157	8.8	13.1	15.8
505 to 630	0.000049	0.000047	12.6	28.4	31.1
630 to 800	0.000012	0.000012	21.9	15.4	26.8
$\mathbf{p_T^{b\bar{b}}}$ [GeV]	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
0 to 35	0.003293	0.003412	2.4	4.1	4.7
35 to 75	0.006784	0.006553	1.7	3.0	3.5
75 to 115	0.008290	0.008514	1.5	3.0	3.4
115 to 155	0.004954	0.005040	2.0	6.9	7.1
155 to 280	0.000632	0.000579	3.8	12.1	12.7
280 to 500	0.000020	0.000017	14.1	31.5	34.6
$\mathbf{N_{jets}}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
3.5 to 4.5	0.513525	0.516308	0.7	14.5	14.5
4.5 to 5.5	0.321758	0.325448	1.3	13.2	13.3
5.5 to 6.5	0.121527	0.115853	3.3	11.9	12.4
6.5 to 7.5	0.034157	0.034852	7.0	13.4	15.1
7.5 to 9.5	0.004416	0.003695	19.3	27.6	33.6
$\boldsymbol{\rho_S}= \frac{\mathbf{2\cdot 170\,GeV}}{\mathbf{m_{t\bar{t}j}}}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{MC}$	$\left(\frac{1}{\sigma}\frac{d\sigma}{dX}\right)_{exp}$	stat. [%]	syst. [%]	tot. [%]
0.00 to 0.25	0.018865	0.023553	10.9	62.9	63.9
0.25 to 0.55	1.444428	1.642313	4.7	15.6	16.3
0.55 to 0.70	2.563340	2.358678	4.4	14.1	14.7
0.70 to 1.10	0.443636	0.369040	9.6	27.1	28.8

Table D.5: Normalised differential cross section for the combined ℓ+jets channel in the visible particle level phase space as a function of the invariant mass of the lepton and the leptonic b jet (m^{lb}), the invariant mass ($m^{b\bar{b}}$) and the transverse momentum ($p_T^{b\bar{b}}$) of the $b\bar{b}$ system, the jet multiplicity (N_{jets}) and ρ_S. The quoted uncertainties in the last three columns refer to the data result (sub-script exp). The quoted Monte-Carlo prediction (MC) is obtained from the MadGraph+pythia event generator.

Appendix E

Covariance and Correlation Matrices of the Differential Measurements

The covariance matrices for all normalised differential cross section measurements are derived following the method explained in Chapter 9.3.1 for the statistical uncertainty and the method explained in Chapter 10.3 for all systematic uncertainties. The corresponding results can be found in Appendix E.1 for the statistical and Appendix E.2 for the systematic uncertainties. Furthermore, the total (stat.$\oplus$syst.) correlation of the normalised differential cross sections is derived from the covariance matrices using Equation 10.16 in Chapter 10.3. The obtained results are shown in Appendix E.3.

E.1 Statistical Covariance Matrices

The statistical covariance matrices for all normalised differential cross sections are shown in Figure E.1 and E.2 for the quantities measured in the visible particle level phase space and in Figure E.3 and E.4 for the quantities measured in the extrapolated parton level phase space.

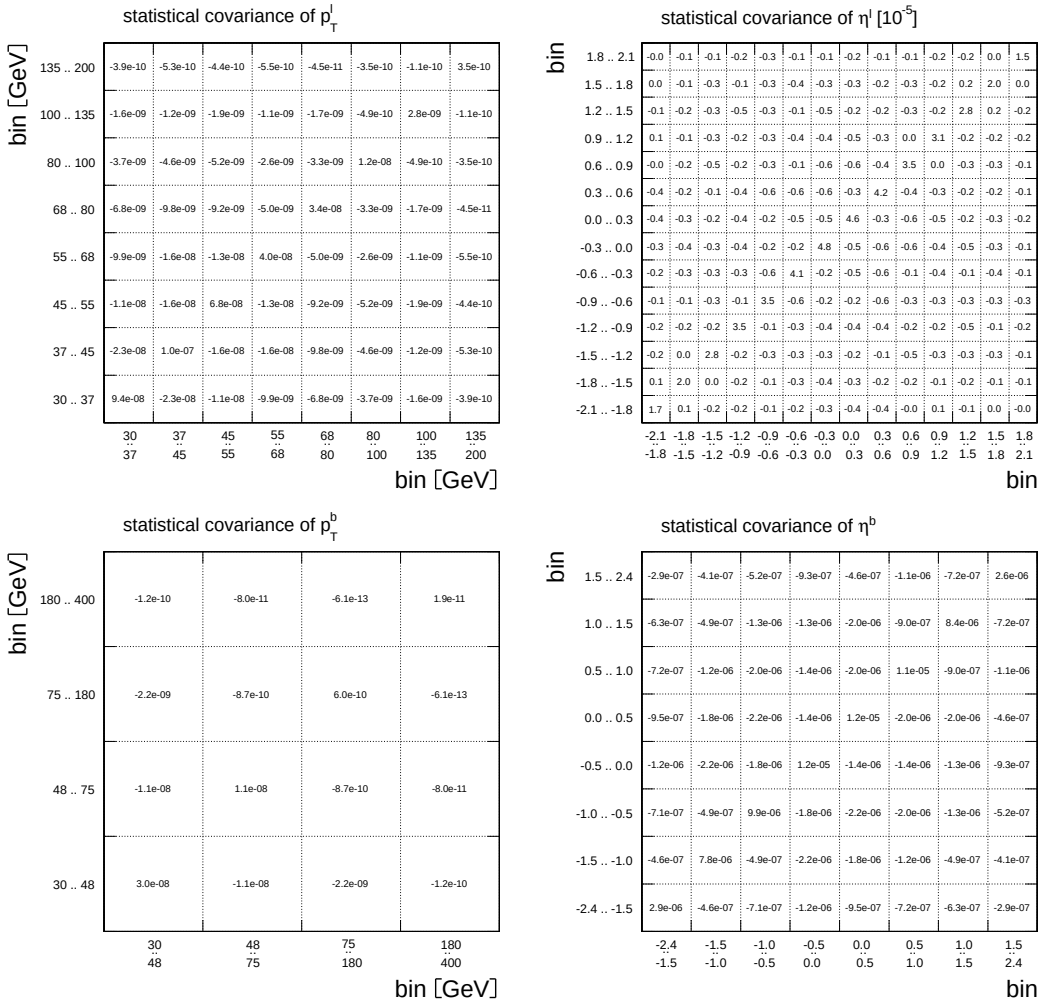

Figure E.1: Statistical covariance of the measured normalised differential cross section for the lepton and b-jet quantities in the visible particle level phase space.

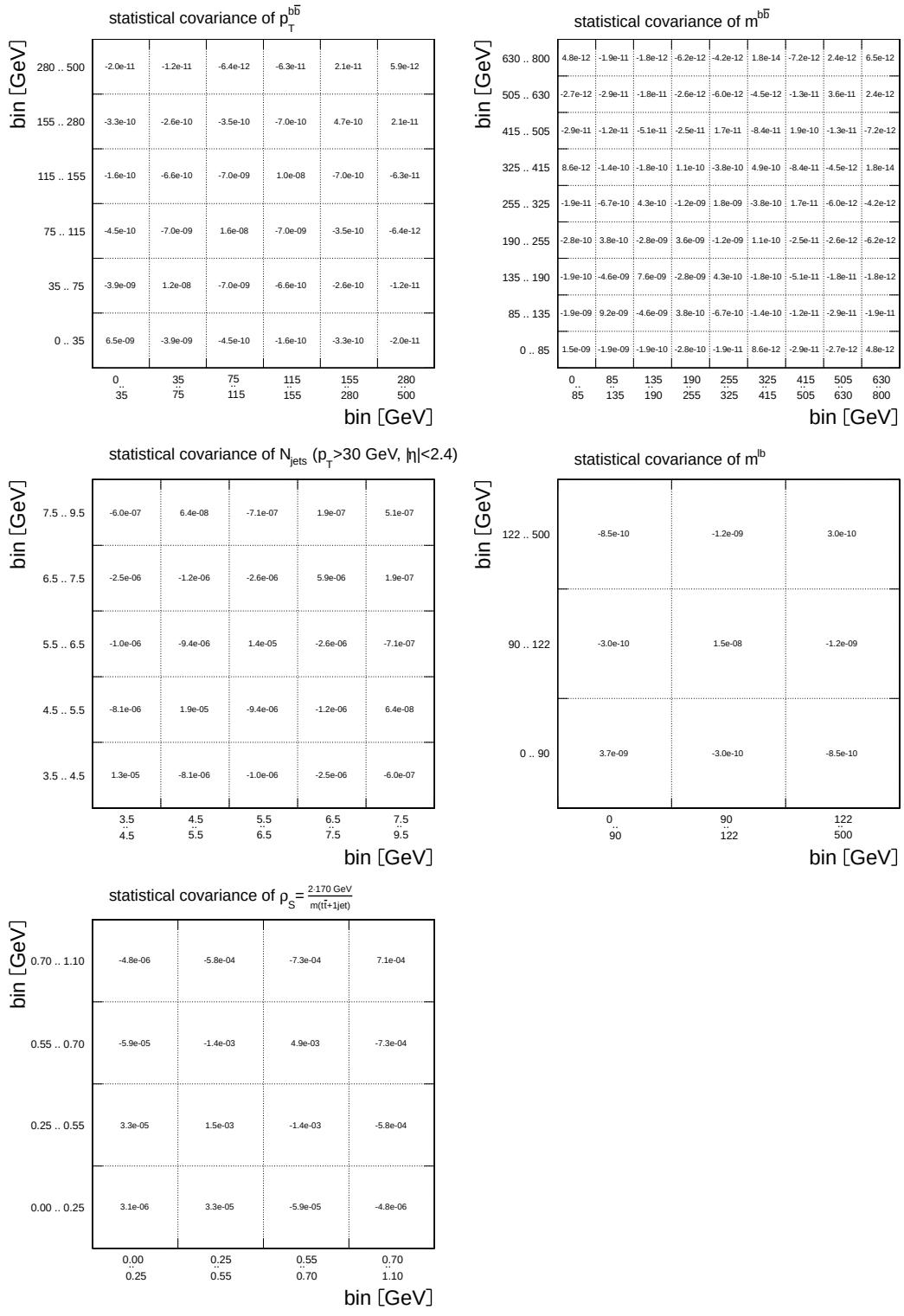

Figure E.2: Statistical covariance of the measured normalised differential cross section for the $b\bar{b}$ system quantities, the jet multiplicity, the invariant mass of the lepton and the leptonic b bjet and ρ_S in the visible particle level phase space.

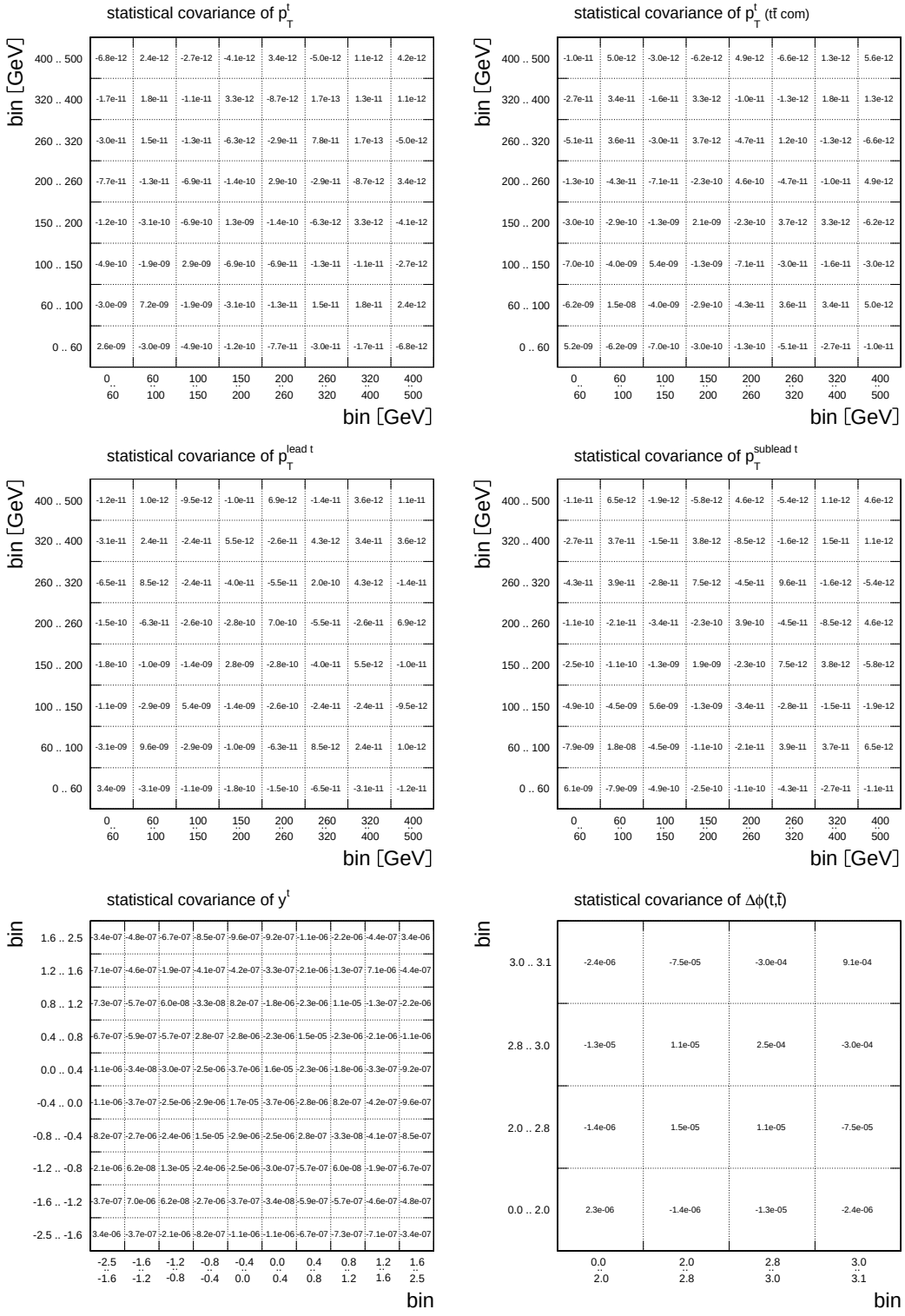

Figure E.3: Statistical covariance of the measured normalised differential cross section for the top-quark quantities in the extrapolated parton level phase space.

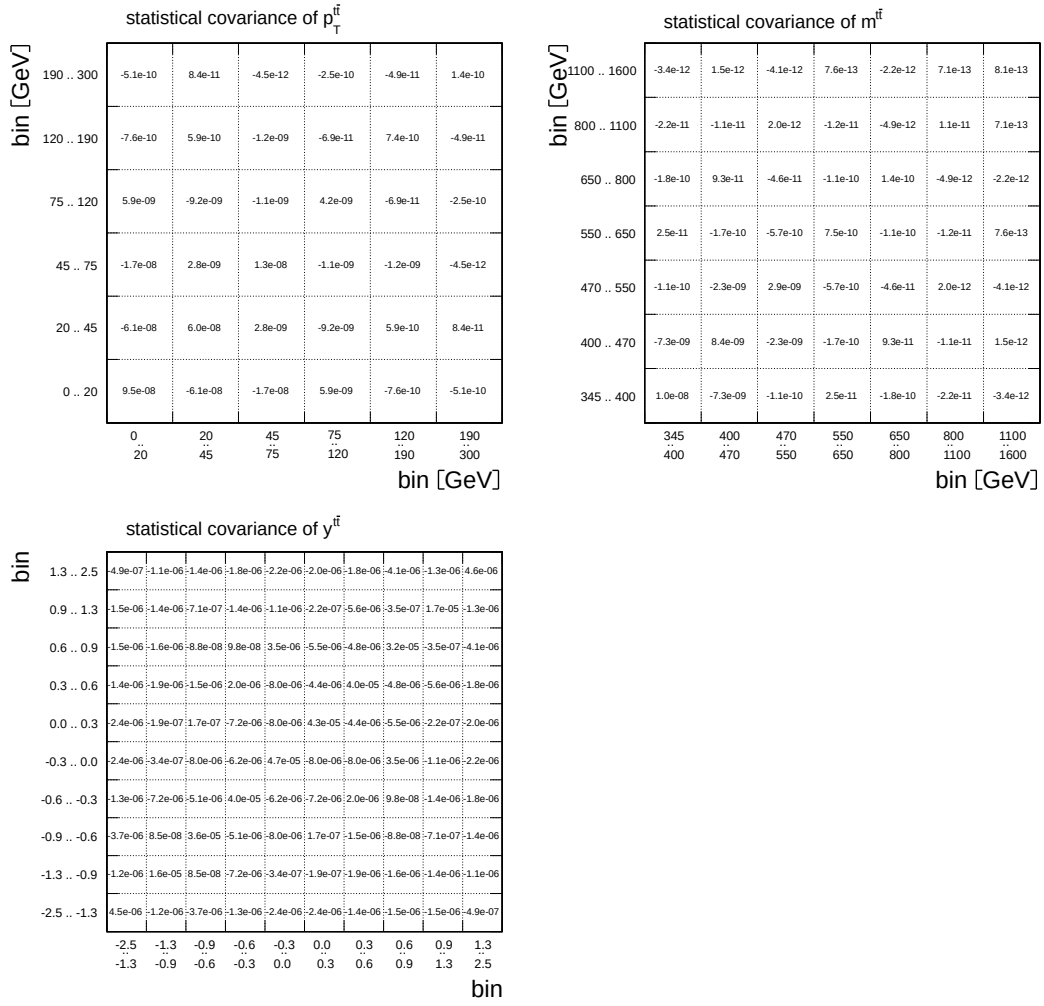

Figure E.4: Statistical covariance of the measured normalised differential cross section for the $t\bar{t}$ system quantities in the extrapolated parton level phase space.

E.2 Systematic Covariance Matrices

The systematic covariance matrices for all normalised differential cross sections are shown in Figure E.5 and E.6 for the quantities measured in the visible particle level phase space and in Figure E.7 and E.8 for the quantities measured in the extrapolated parton level phase space.

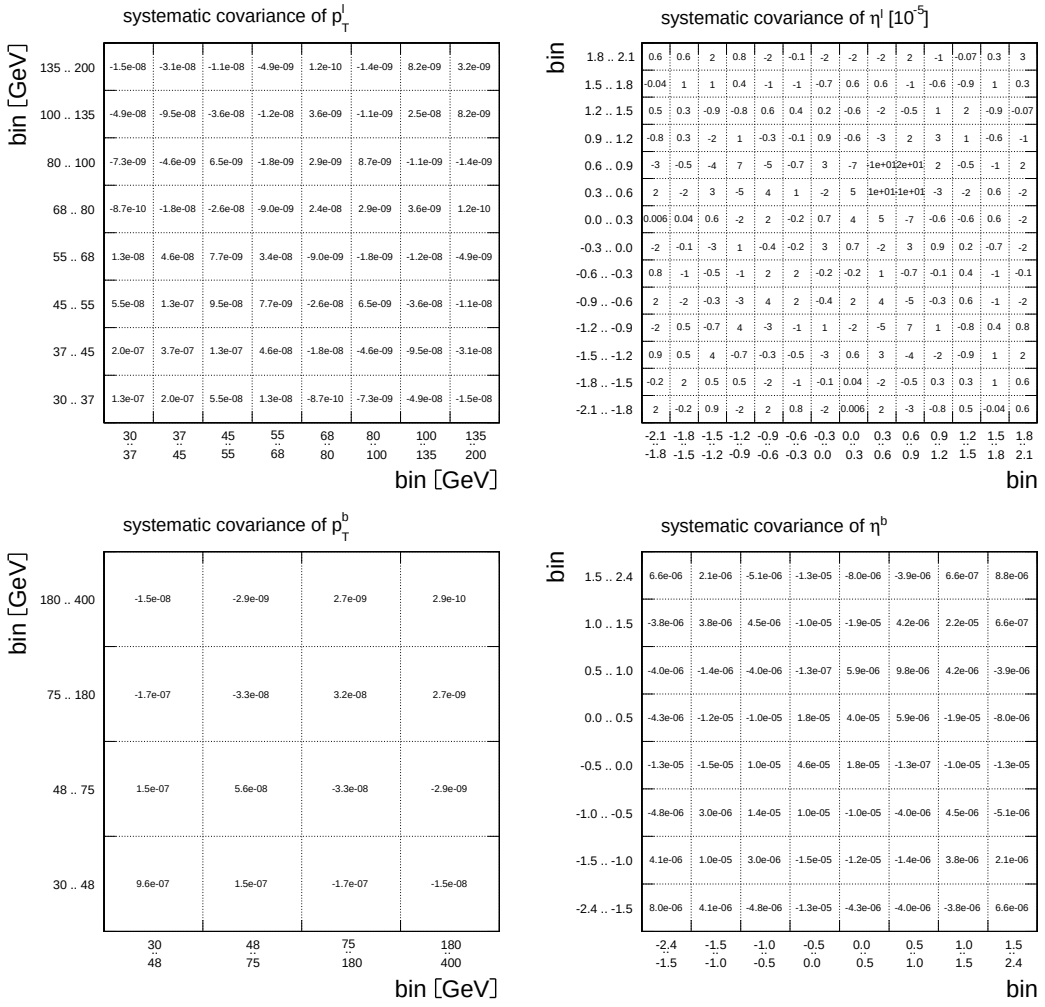

Figure E.5: Systematic covariance of the measured normalised differential cross section for the lepton and b-jet quantities in the visible particle level phase space.

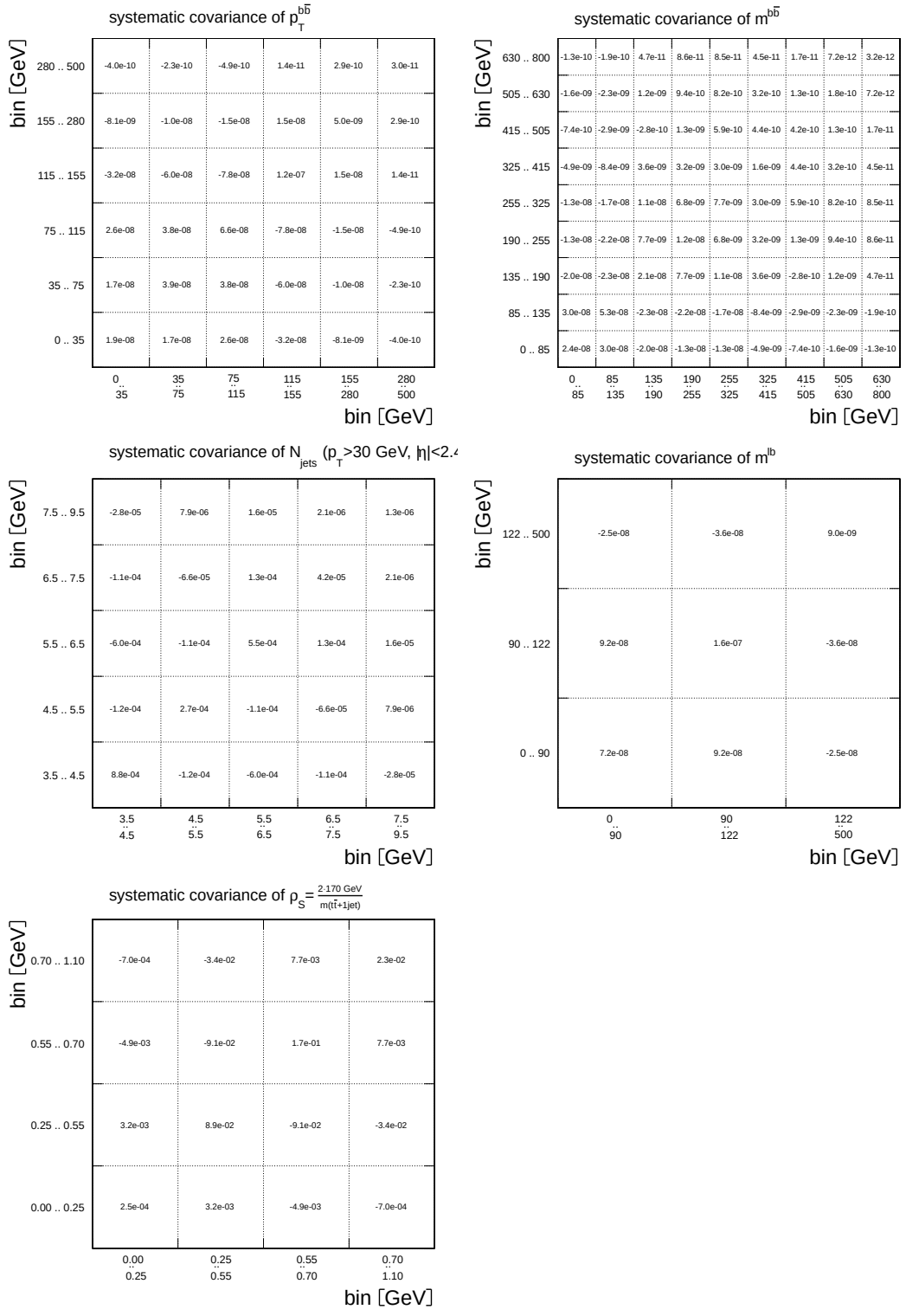

Figure E.6: Systematic covariance of the measured normalised differential cross section for the $b\bar{b}$ system quantities, the jet multiplicity, the invariant mass of the lepton and the leptonic b bjet and ρ_S in the visible particle level phase space.

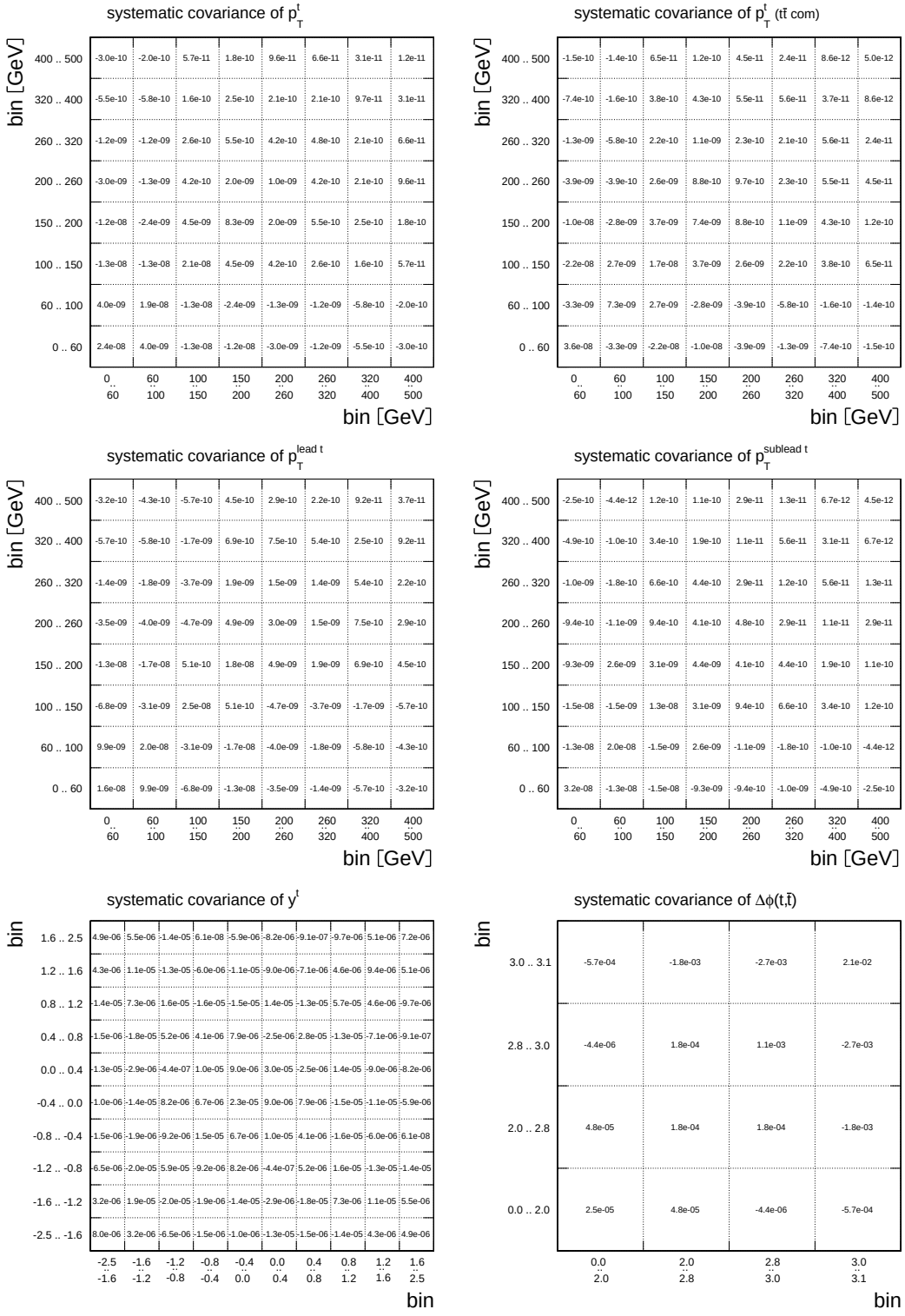

Figure E.7: Systematic covariance of the measured normalised differential cross section for the top-quark quantities in the extrapolated parton level phase space.

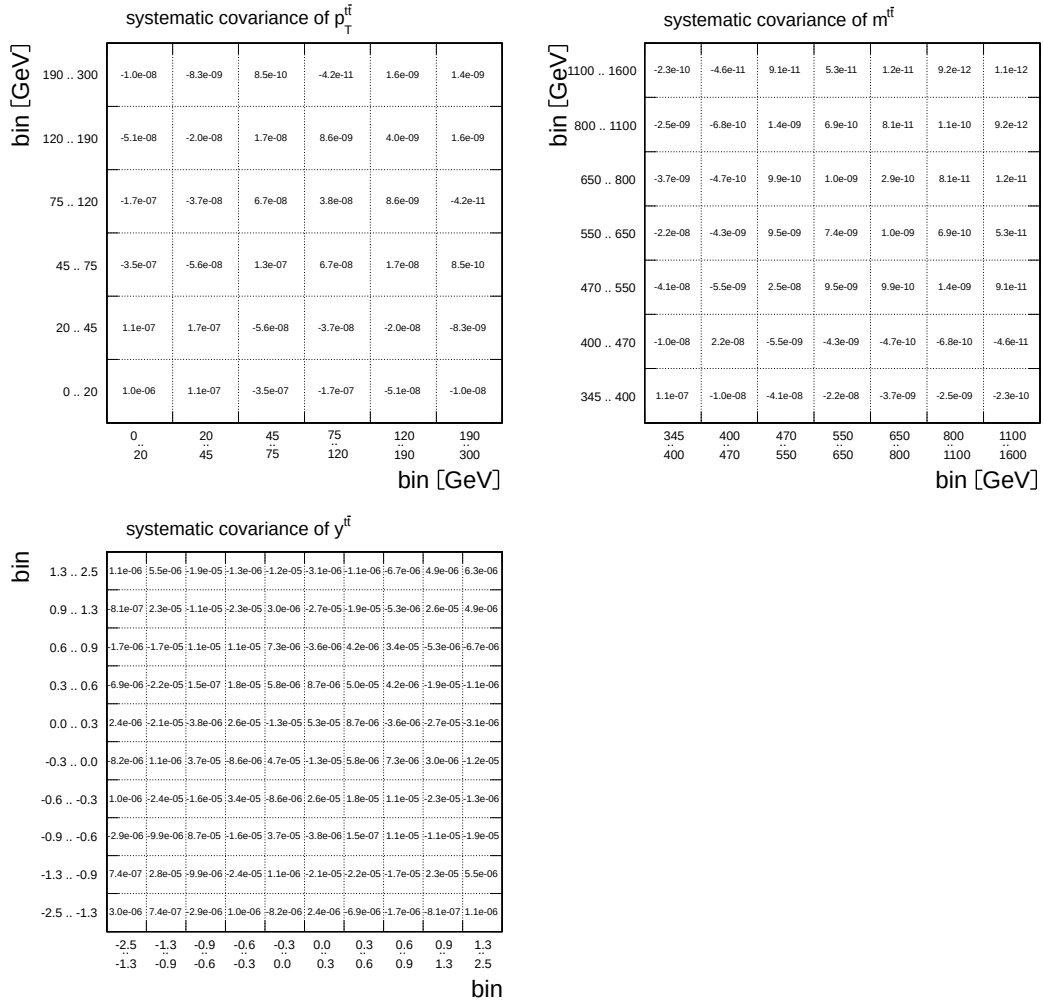

Figure E.8: Systematic covariance of the measured normalised differential cross section for the $t\bar{t}$ system quantities in the extrapolated parton level phase space.

E.3 Correlations of the Final Result

The combined (statistical and systematic) correlation matrices for all normalised differential cross sections are shown in Figure E.9 and E.10 for the quantities measured in the visible particle level phase space and in Figure E.11 and E.12 for the quantities measured in the extrapolated parton level phase space. As expected, predominantly negative correlations are observed due to the normalisation of the results. Moreover, migration effects often result in a positive correlation for neighbouring bins with approximately equal content.

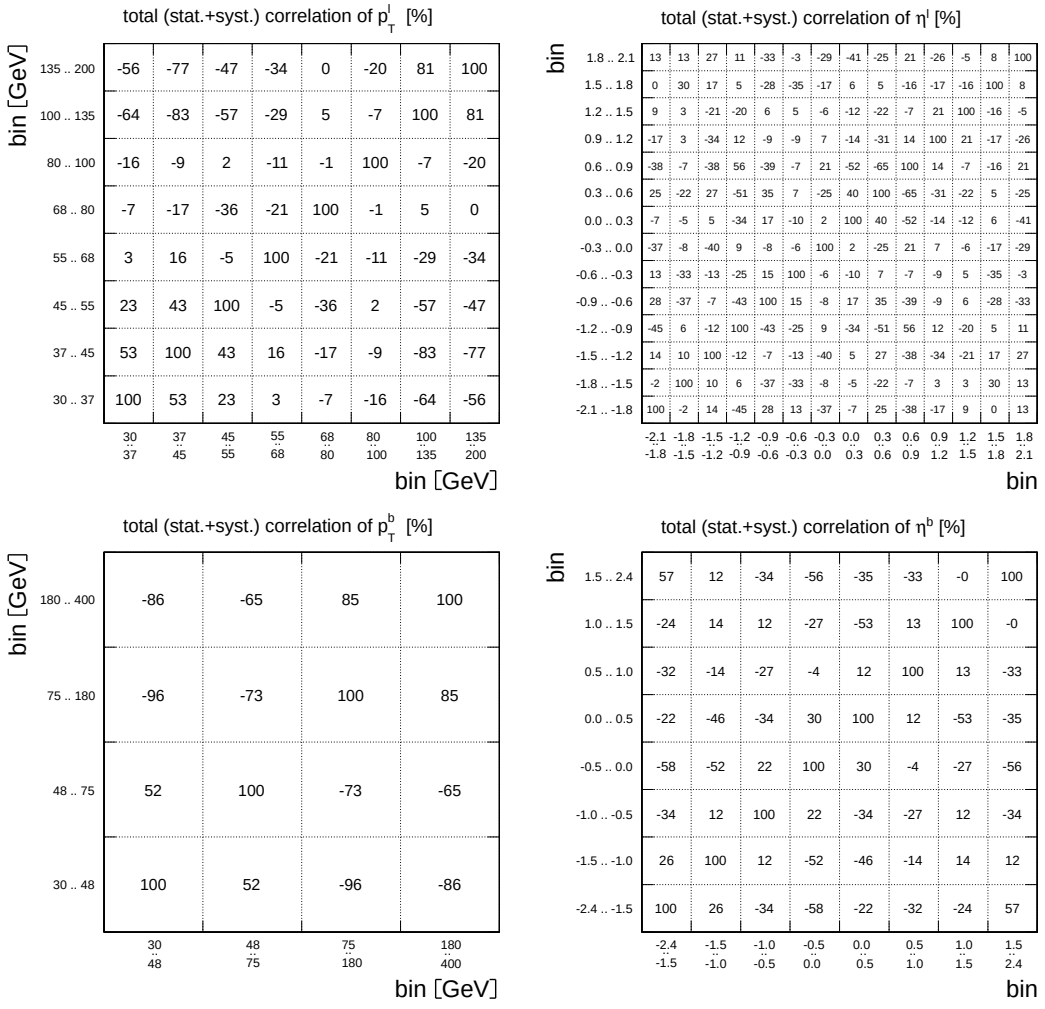

Figure E.9: Total (statistical and systematic) correlation of the measured normalised differential cross section for the lepton and b-jet quantities in the visible particle level phase space.

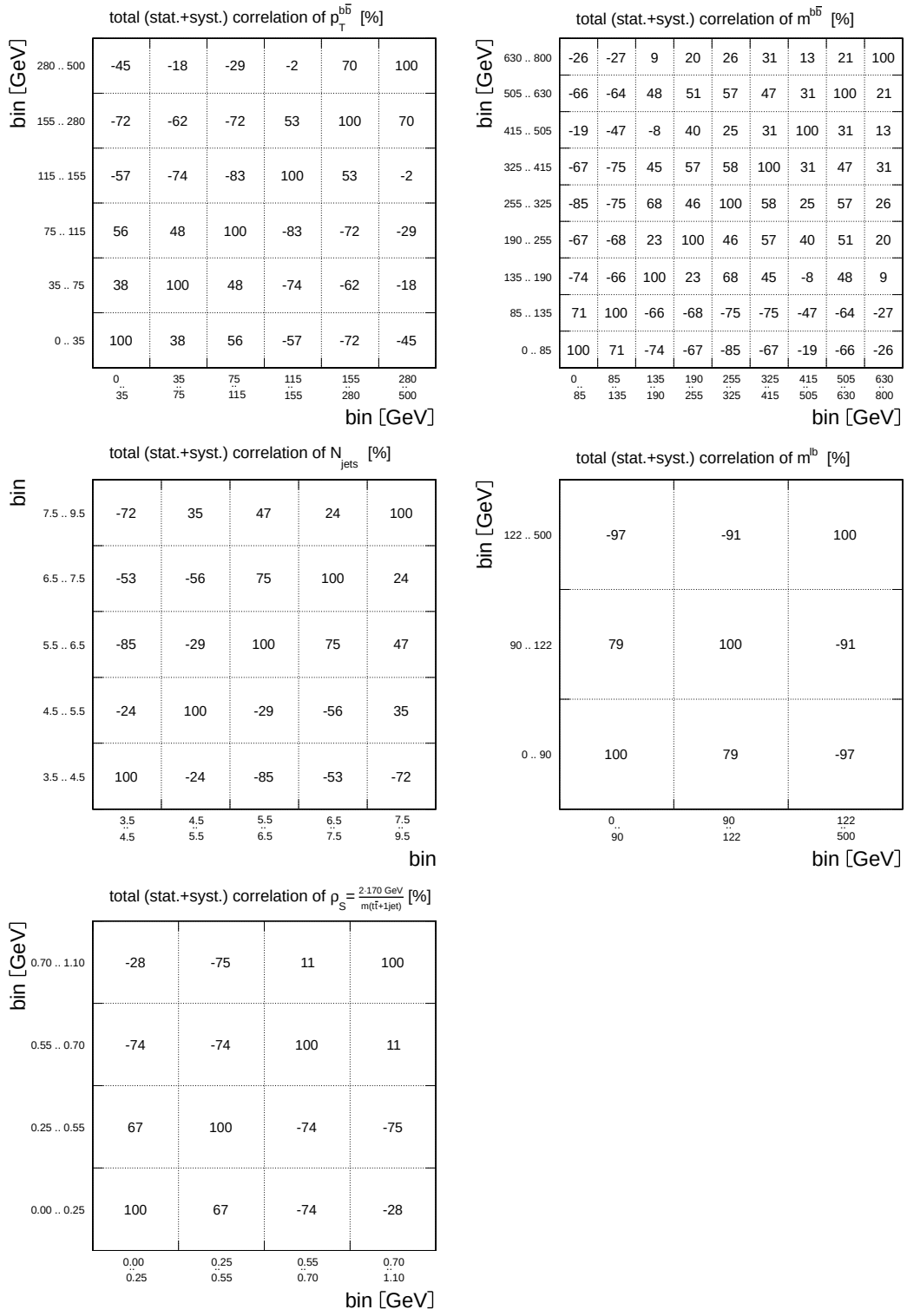

Figure E.10: Total (statistical and systematic) correlation of the measured normalised differential cross section for the $b\bar{b}$ system quantities, the jet multiplicity, the invariant mass of the lepton and the leptonic b bjet and ρ_S in the visible particle level phase space.

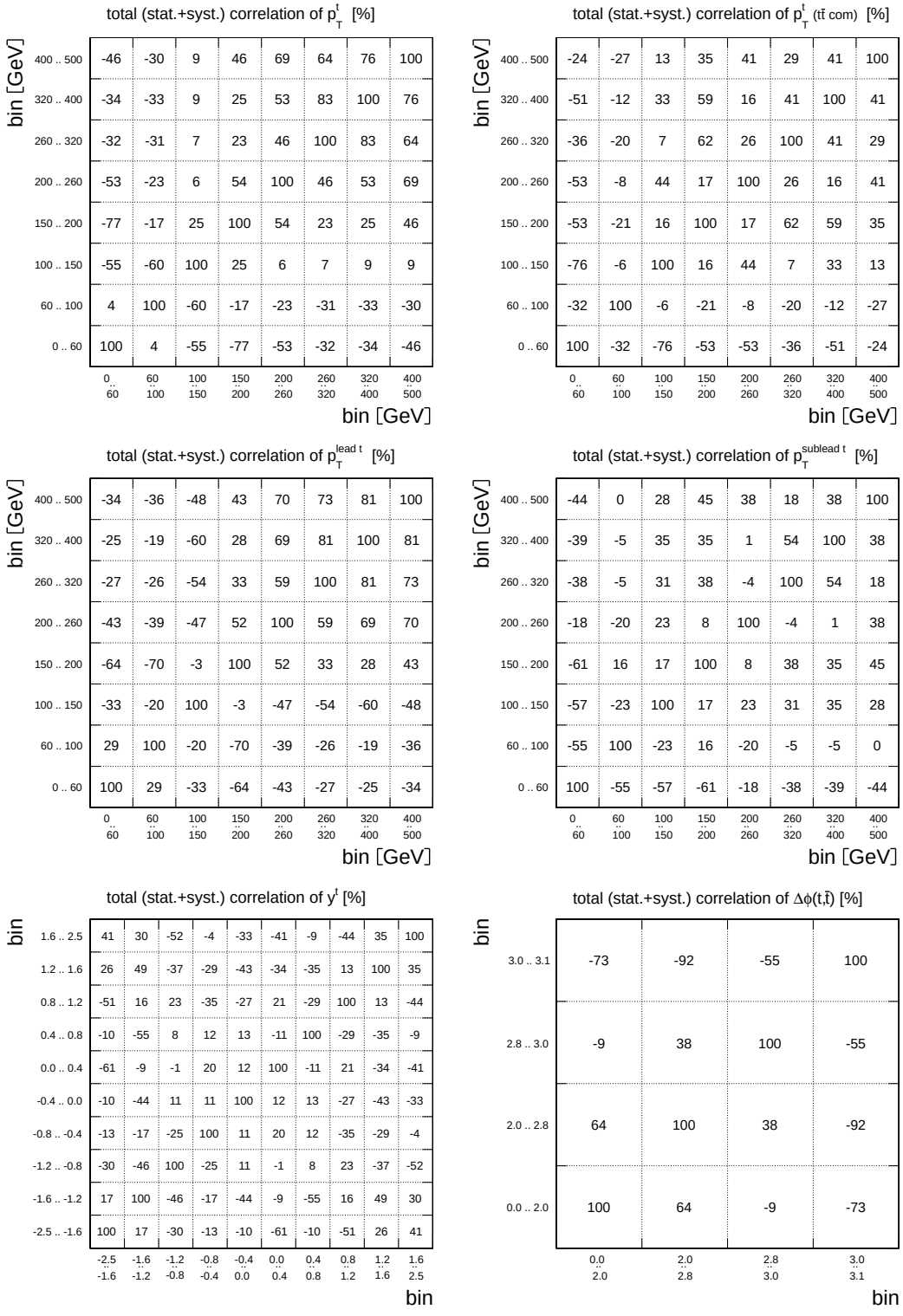

Figure E.11: Total (statistical and systematic) correlation of the measured normalised differential cross section for the top-quark quantities in the extrapolated parton level phase space.

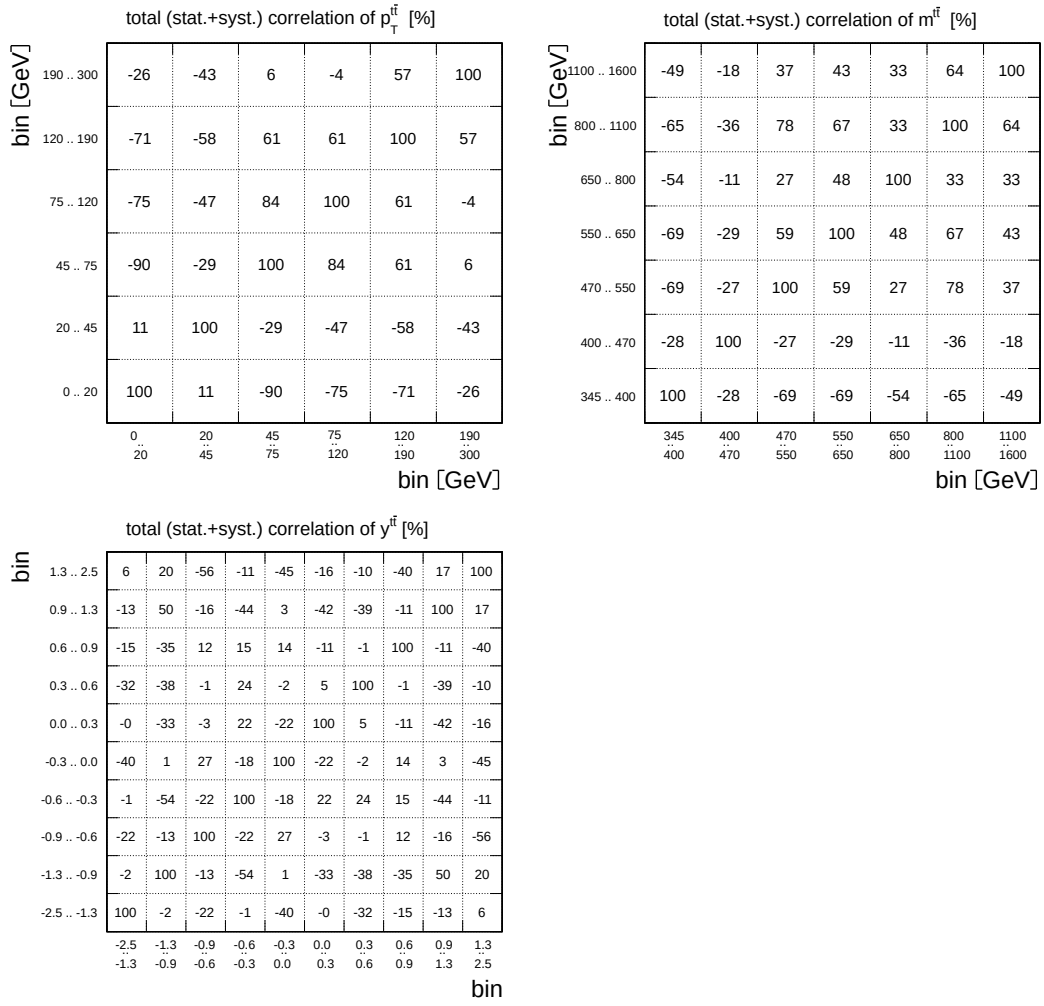

Figure E.12: Total (statistical and systematic) correlation of the measured normalised differential cross section for the $t\bar{t}$ system quantities in the extrapolated parton level phase space.

Appendix F

Fine-Binned Monte-Carlo Predictions at $\sqrt{s} = 7$ TeV

Within the TopLHCWG, the predicted shapes of the most important cross section quantities have been studied for various MC event generators at $\sqrt{s} = 7$ TeV (see discussion in Chapter 11.2.5).
One million simulated events are used for all MC event generators. All predictions are normalised by the number of entries to compare the predicted shape of the distributions. The corresponding CMS MC predictions for the extrapolated parton level phase space can be found in Figure F.1. The statistical uncertainty for a single MC prediction is indicated as light grey band.

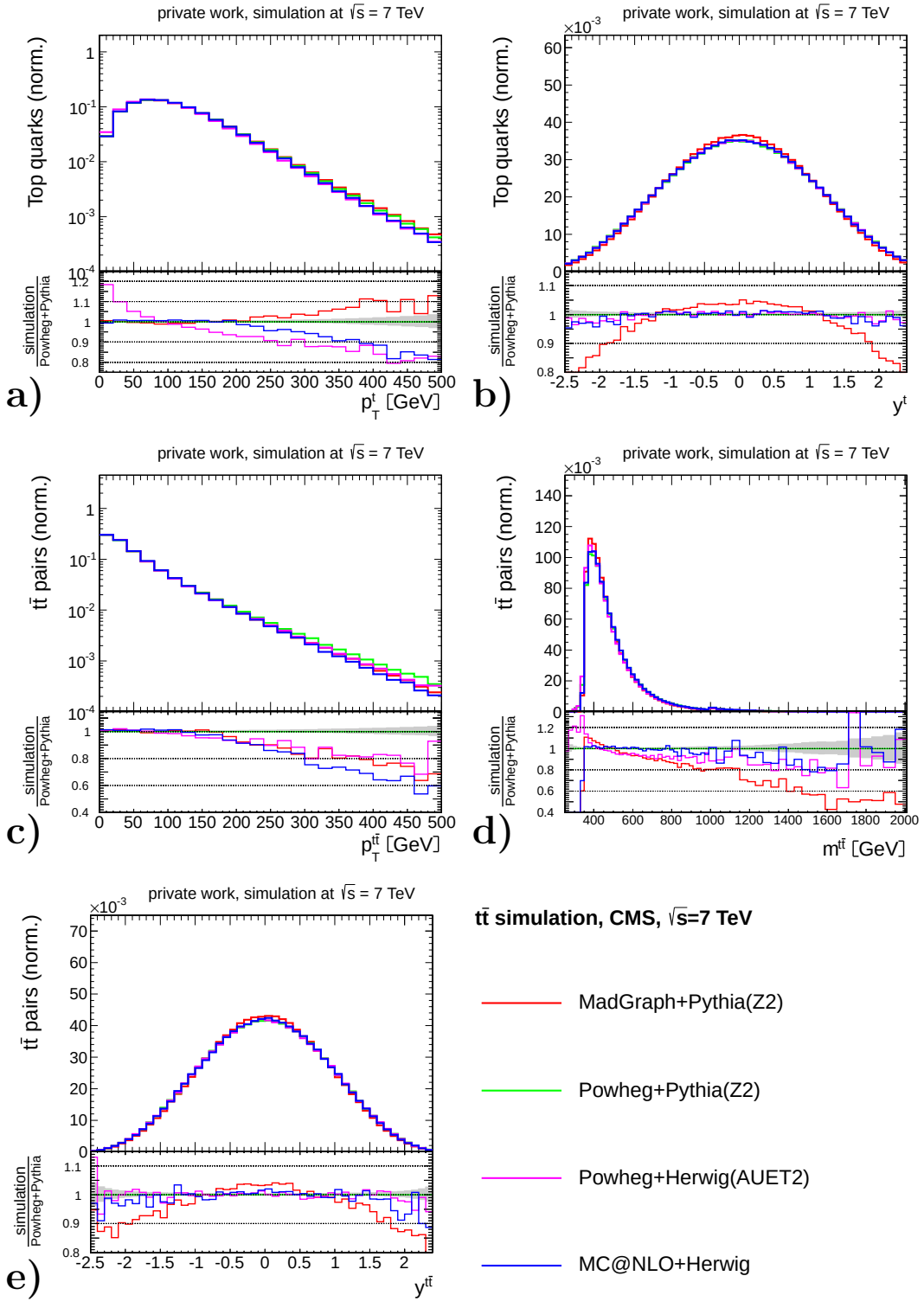

Figure F.1: Comparison of normalised differential distributions predicted by different MC event generators of CMS in the extrapolated parton level phase space for the transverse momentum (p_T^t, a) and rapidity (y^t, b) of the top quarks as well as for the transverse momentum ($p_T^{t\bar{t}}$, c), invariant mass ($m^{t\bar{t}}$, d) and rapidity ($y^{t\bar{t}}$, e) of the $t\bar{t}$ system. The grey band indicates the statistical uncertainty of a single MC prediction.

Appendix G

Uncertainties of the MadGraph+Pythia Prediction

The results of the normalised differential cross sections presented in Chapter 11.2 do not include uncertainty bands for the MC predictions. Therefore, the impact of the main modelling parameters (Q^2 scale, ME-PS matching scale and top-quark mass) on the MadGraph+pythia MC prediction are shown in Figure G.1 and G.2.

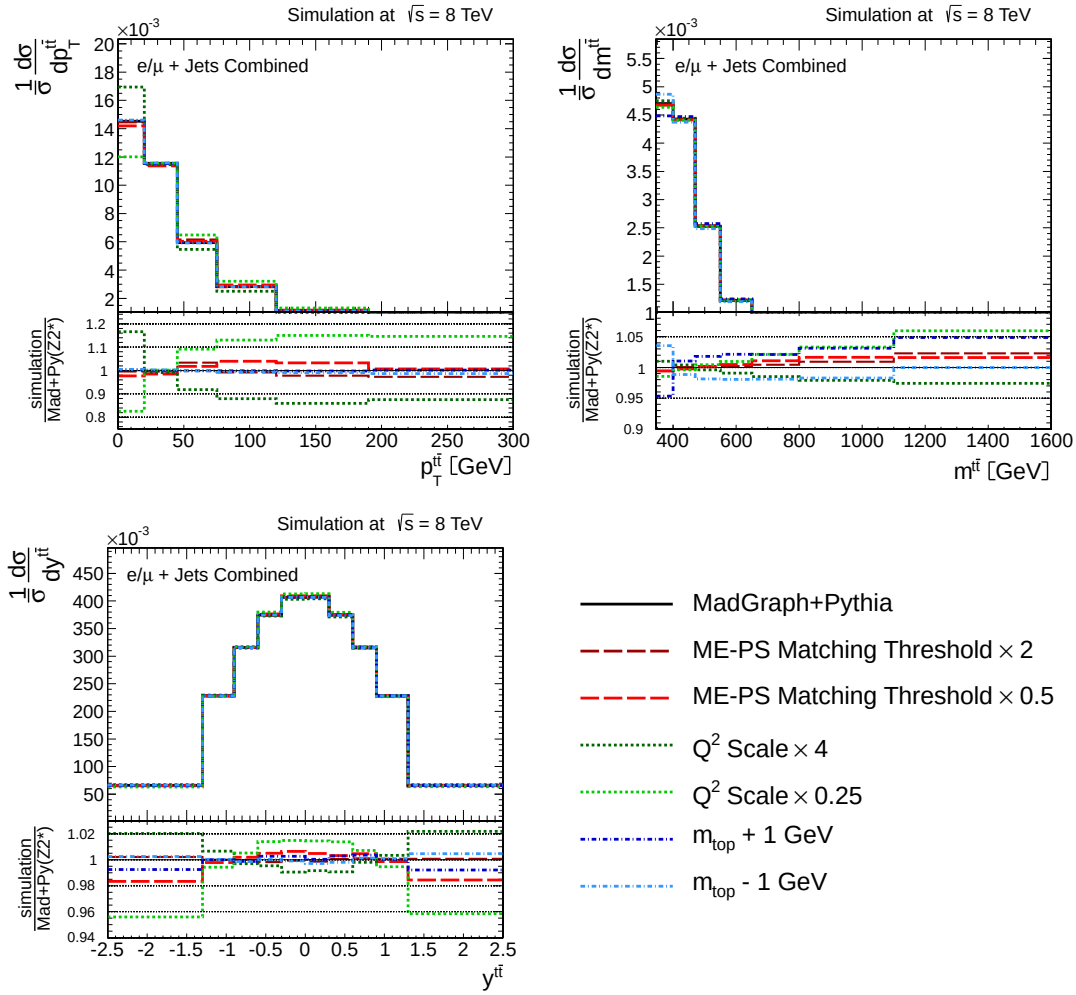

Figure G.1: Impact of the top-quark mass, Q^2 scale and ME-PS matching scale on the normalised differential cross section predicted by MadGraph+pythia for the $t\bar{t}$ system quantities.

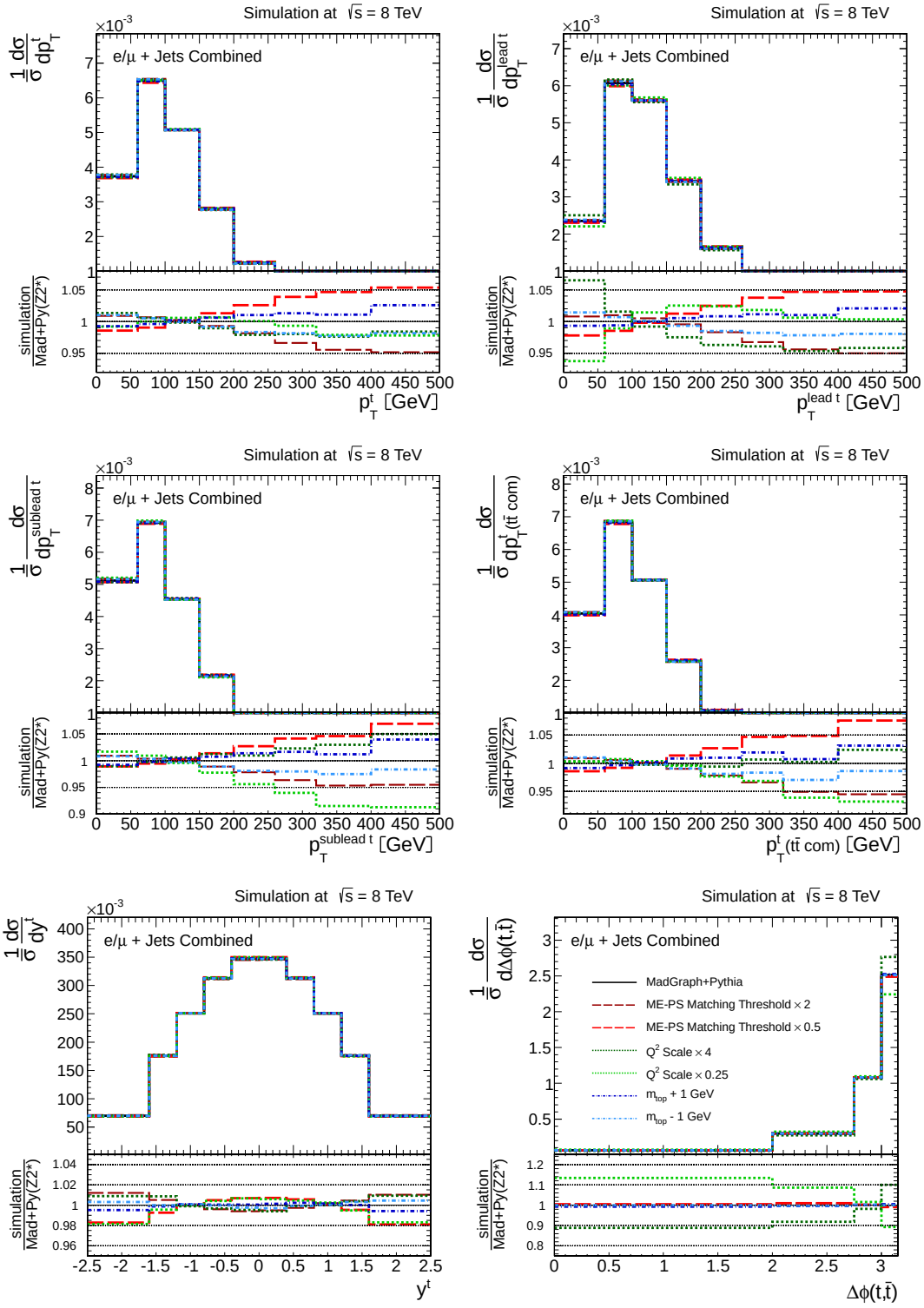

Figure G.2: Impact of the top-quark mass, Q^2 scale and ME-PS matching scale on the normalised differential cross section predicted by MadGraph+pythia for the top-quark quantities.

Appendix H

Compatibility of Data and the MadGraph+Pythia Prediction

Quantifying the agreement between data and the MadGraph+pythia prediction, the difference in the normalised differential cross section relative to the uncertainties are shown in Figure H.1 and H.2. The results of the measurement and the prediction as well as the uncertainties of the measurement are detailed in Appendix D. The uncertainties of the prediction are detailed in Appendix G and the following uncertainties sources are taken into account:

- $m_{top} \pm 1$ GeV
- ME-PS matching threshold
- Q^2 scale

} see Chapter 10

The typical difference between data and simulation is at the level of one to two standard deviations of the combined uncertainty of data and prediction. However, the measurements related to the top-quark transverse momentum differ by typically two to three (two to five) standard deviations of the combined (data) uncertainty.

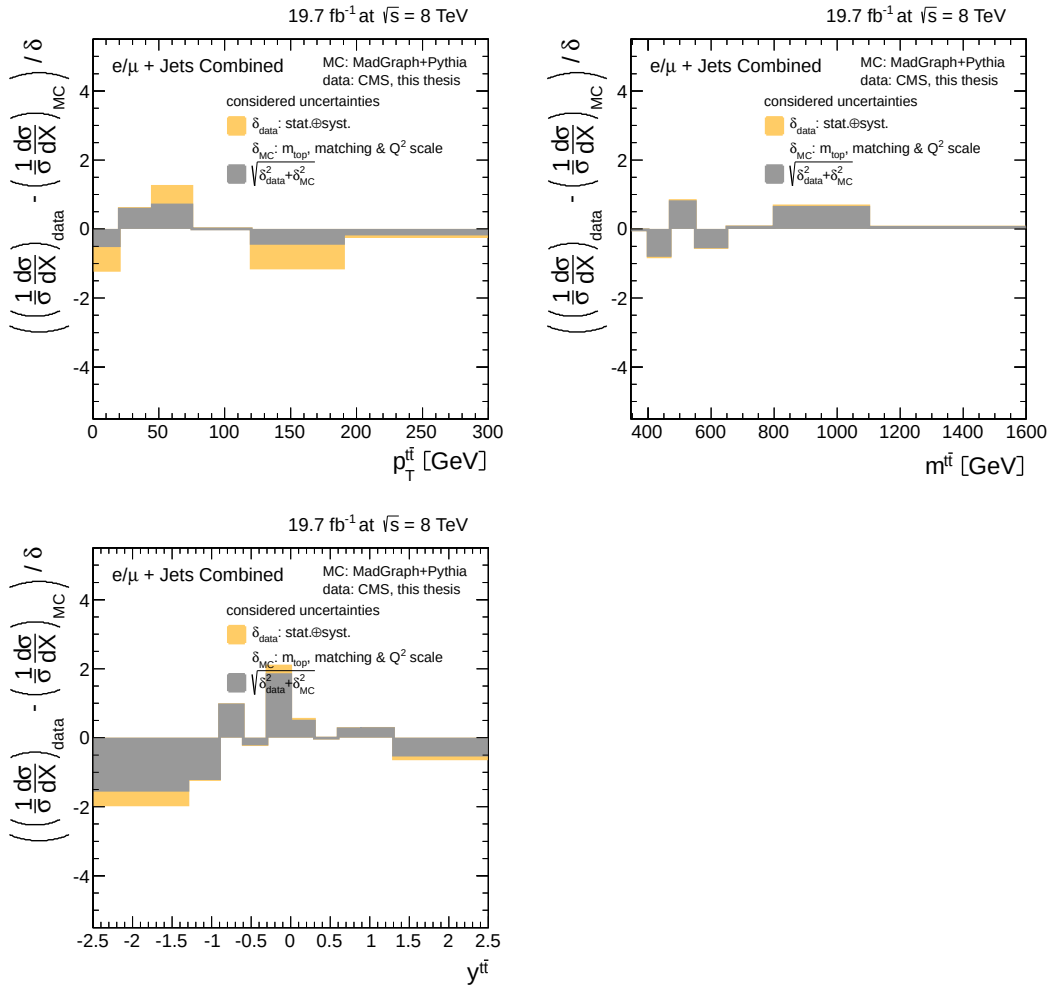

Figure H.1: Difference of the normalised differential cross section between data and the MadGraph+pythia prediction for the $t\bar{t}$-system cross section quantities.

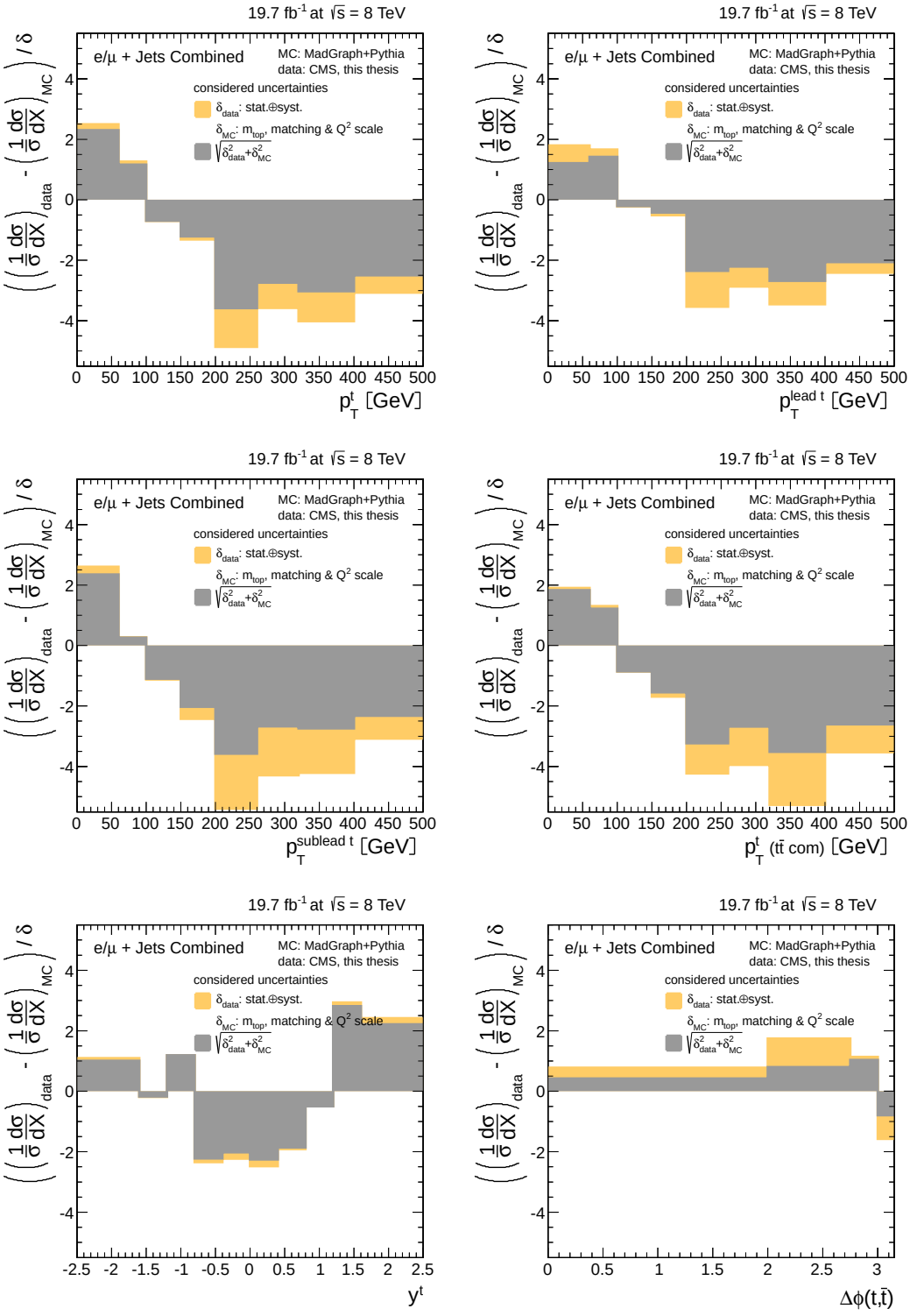

Figure H.2: Difference of the normalised differential cross section between data and the MadGraph+pythia prediction for the top-quark cross section quantities.

Appendix I

χ^2 Comparison With Correlations

In Chapter 11.2.3, the compatibility between data and the different theory predictions is evaluated for each measured normalised differential cross section by calculating a χ^2/ndof expression. The correlation between the different bins is neglected in this approach. However, the expected correlation of the normalised differential cross sections are presented in Appendix E and found to be sizeable in several cases and a χ^2/ndof is calculated using the total (statistical and systematic) covariance matrix of the measurement, following the approach in [22]:

$$\frac{\chi^2_{\text{theory}}}{\text{ndof}} = \left(\frac{1}{\text{N}_{\text{bins}} - 1}\right) \cdot \left(\vec{\Delta}^{\text{T}}_{\text{N}_{\text{bins}}-1} \cdot \mathbf{COV}^{-1}_{\mathbf{sub}} \cdot \vec{\Delta}_{\text{N}_{\text{bins}}-1}\right). \tag{I.1}$$

Due to the normalisation, ndof and the rank of the covariance matrix is lowered by one and one bin is discarded for the calculation to receive an invertible $((\text{N}_{\text{bins}} - 1) \times (\text{N}_{\text{bins}} - 1))$ sub-matrix ($\mathbf{COV_{sub}}$) of the full covariance matrix. Analogously, the corresponding entry in the vector of differences between data and theory prediction ($\vec{\Delta} = \left(\frac{1}{\sigma}\frac{\vec{\text{d}\sigma}}{\text{dX}}\right)_{\text{data}} - \left(\frac{1}{\sigma}\frac{\vec{\text{d}\sigma}}{\text{dX}}\right)_{\text{theory}}$) is also discarded.

The calculated χ^2/ndof expression is expected to be independent of the discarded bin. In addition, the p-value, which corresponds to the probability that the χ^2 is larger than or equal to the observed value assuming the prediction to be correct, is calculated using Equation 7.5. The results can be found in Table I.1 for the extrapolated parton level phase space and in Table I.2 for the visible particle level phase space.

Note, that the result for χ^2/ndof depends heavily on the calculated systematic covariance matrices. The used method (see Chapter 10.3) has been mainly checked for the measured cross sections in the extrapolated parton level phase space. For some quantities measured in the visible particle level phase space like e.g. N_{jets}, the convergence of the iterative procedure is worse and the changes of the total systematic uncertainties by the procedure are observed to be larger than for the quantities in the extrapolated parton level phase space. The obtained result is found to be not completely independent of the choice of the removed bin for some quantities like e.g. $\text{p}^{\text{l}}_{\text{T}}$, indicating that the normalisation condition is not perfectly fulfilled for all considered systematic variations in the applied procedure due to the symmetrisation of up and

down variations.

quantity	$\frac{\chi^2_{\text{theory}}}{\text{ndof}}$ (p-value)				
	MadGraph + pythia	mc@nlo + herwig	powheg + pythia	powheg + herwig	pert.QCD
p_T^t	4.9 (0.00)	4.8 (0.00)	4.8 (0.00)	1.4 (0.21)	1.9 (0.06)
$p_T^{\text{lead t}}$	3.8 (0.00)	3.5 (0.00)	4.2 (0.00)	1.2 (0.31)	-
$p_T^{\text{sublead t}}$	8.4 (0.00)	4.3 (0.00)	4.7 (0.00)	0.8 (0.61)	-
p_T^t ($t\bar{t}$ com)	8.5 (0.00)	2.7 (0.01)	3.6 (0.00)	0.7 (0.67)	-
y^t	3.2 (0.00)	2.3 (0.02)	2.0 (0.03)	2.2 (0.02)	2.9 (0.00)
$\Delta\phi\,(t,\bar{t})$	1.0 (0.38)	1.2 (0.30)	2.3 (0.07)	1.9 (0.12)	-
$m^{t\bar{t}}$	0.6 (0.74)	2.3 (0.03)	3.0 (0.01)	1.3 (0.27)	3.0 (0.01)
$p_T^{t\bar{t}}$	2.0 (0.07)	1.7 (0.12)	2.5 (0.03)	1.1 (0.38)	6.8 (0.00)
$y^{t\bar{t}}$	1.1 (0.35)	1.7 (0.07)	2.1 (0.03)	2.0 (0.03)	-
average	3.7 (0.17)	2.7 (0.06)	3.2 (0.02)	1.4 (0.29)	-

Table I.1: χ^2/ndof expressions and p-values for all normalised differential cross sections measured in the extrapolated parton level phase space for all investigated predictions using the total (statistical and systematic) covariance matrix. The last column corresponds to a perturbative QCD calculation of NLO+NNLL precision [202] for $m^{t\bar{t}}$ [203] and $p_T^{t\bar{t}}$ [204] and approx. NNLO precision for p_T^t and y^t [205].

quantity	$\frac{\chi^2_{\text{theory}}}{\text{ndof}}$ (p-value)			
	MadGraph + pythia	mc@nlo + herwig	powheg + pythia	powheg + herwig
p_T^l	1.5 (0.18)	1.1 (0.33)	2.0 (0.05)	1.0 (0.40)
η^l	1.7 (0.06)	1.0 (0.42)	0.7 (0.73)	0.9 (0.57)
p_T^b	2.0 (0.11)	0.8 (0.51)	1.8 (0.14)	0.5 (0.66)
η^b	2.4 (0.02)	1.1 (0.39)	1.0 (0.40)	0.8 (0.57)
$p_T^{b\bar{b}}$	1.0 (0.43)	1.2 (0.33)	1.4 (0.24)	1.0 (0.42)
$m^{b\bar{b}}$	0.6 (0.75)	0.6 (0.74)	0.5 (0.83)	0.9 (0.54)
m^{lb}	0.2 (0.80)	0.3 (0.71)	0.2 (0.83)	0.2 (0.81)
N_{jets}	0.3 (0.89)	3.4 (0.01)	0.9 (0.47)	0.7 (0.59)
ρ_S	0.4 (0.73)	0.6 (0.64)	0.9 (0.43)	1.0 (0.40)
average	1.1 (0.44)	1.1 (0.45)	1.1 (0.46)	0.8 (0.55)

Table I.2: χ^2/ndof expressions and p-values for all normalised differential cross sections measured in the visible particle level phase space for all investigated predictions using the total (statistical and systematic) covariance matrix.

List of Figures

Chapter 7

Chapter 8

Chapter 9

Chapter 11

Chapter 12

Chapter 13

Appendix A

Appendix B

Appendix C

Appendix G

Appendix H

List of Tables

Appendix I

List of Acronyms

BSM	beyond the Standard Model
BR	branching ratio
CHS	charged hadron subtraction
CMS	Compact Muon Solenoid
CMSSW	CMS software
CSV	combined secondary vertex
DQM	data quality monitoring
ECAL	electromagnetic calorimeter
EWK	electroweak
EWSB	electroweak symmetry breaking
FEWZ	fully exclusive W and Z production
gen	generator truth
gg fusion	gluon-gluon fusion
GT	global tag
HCAL	hadronic calorimeter
HLT	high level trigger
ID	identification
ISR/FSR	initial- and final-state radiation
JES	jet energy scale
JER	jet energy resolution
JEC	jet energy correction
JSON format	JavaScript object notation file format
kin-reco	after successful kinematic reconstruction
LHC	Large Hadron Collider
light jets	jets initiated by gluons or quarks of flavour u, d, or s
ℓ+jets final state	$\mathrm{t\bar{t}} \rightarrow \mathrm{WWb\bar{b}}$,$\mathrm{WW} \rightarrow \mathrm{l}\nu\mathrm{q\bar{q}}'$
matching scale	ME-PS matching threshold
ME	matrix element
MC	Monte Carlo
MCFM	Monte Carlo for FeMtobarn processes
MVA	multivariate analysis
ndof	number of degrees of freedom
N$^{\mathrm{i}}$LO	(next-to)^i leading order
N$^{\mathrm{i}}$LL	(next-to)^i leading logarithm
OF	overflow
PDFs	parton-density functions
PDG	Particle Data Group

PF	particle flow
POG	physics object group
pp collision	proton-proton collision
pretagged	after lepton and jet event selection
prob	χ^2-probability
PS	parton shower
PU	pile-up
PV	primary vertex
QCD	Quantum Chromodynamics
$q\bar{q}$ annihilation	quark-antiquark annihilation
Q^2 scale	hard-scattering scale
rec or reco	reconstructed
SF	scale factor
SM	Standard Model of particle physics
SUSY	Supersymmetry
tagged	after lepton, jet and b-jet identification event selection
$t\bar{t}$ BG / $t\bar{t}$ other	$t\bar{t}$ events decaying not into the ℓ+jets final state ($\ell = e/\mu$)
$t\bar{t}$ SG / $t\bar{t}$ signal	$t\bar{t}$ events decaying into the ℓ+jets final state ($\ell = e/\mu$)
TopLHCWG	Top-quark LHC working group
T&P	tag-and-probe
TQAF	Top-Quark Analysis Framework
UE	underling event
UF	underflow
WP	working point

Bibliography

[1] Johann Wolfgang von Goethe, "Faust. Der Tragödie erster Teil", 1808. Cited on page 1.

[2] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321 – 323. doi:10.1103/PhysRevLett.13.321. Cited on page 2 and 14.

[3] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508 – 509. doi:10.1103/PhysRevLett.13.508. Cited on page 2 and 14.

[4] Nobelprize.org, "The Nobel Prize in Physics 2013". Nobel Media AB 2013. Web. 22. [Online, accessed 13 March 2014]. Cited on page 2.

[5] Rudolf Haag and Jan T. Łopuszański and Martin Sohnius, "All possible generators of supersymmetries of the S-matrix", *Nuclear Physics B* **88** (1975), no. 2, 257 – 274. doi:10.1016/0550-3213(75)90279-5. Cited on page 2.

[6] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, "The hierarchy problem and new dimensions at a millimeter", *Physics Letters B* **429** (1998), no. 3–4, 263 – 272, arXiv:hep-ph/9803315. doi:10.1016/S0370-2693(98)00466-3. Cited on page 2 and 29.

[7] Joseph Polchinski, "String Theory, Vol. 1: An Introduction to Bosonic String". Cambridge University Press, 2007. Cited on page 2.

[8] Joseph Polchinski, "String Theory, Vol. 2: Superstring Theory and Beyond". Cambridge University Press, 2007. Cited on page 2.

[9] G. Altarelli, B. Mele, and M. Ruiz-Altaba, "Searching for new heavy vector bosons in $p\bar{p}$ colliders", *Z. Phys. C* **45** (1989) 109. doi:10.1007/BF01556677. Cited on page 2 and 258.

[10] G. Degrassi, S. Di Vita, J. Elias-Miro et al., "Higgs mass and vacuum stability in the Standard Model at NNLO", *JHEP* **1208** (2012) 098, arXiv:1205.6497. doi:10.1007/JHEP08(2012)098. Cited on page 3 and 28.

[11] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", *Phys. Rev. Lett.* **74** (1995) 2626–2631, arXiv:hep-ex/9503002. doi:10.1103/PhysRevLett.74.2626. Cited on page 3 and 17.

[12] D0 Collaboration, "Observation of the top quark", *Phys. Rev. Lett.* **74** (1995) 2632–2637, arXiv:hep-ex/9503003. doi:10.1103/PhysRevLett.74.2632. Cited on page 3 and 17.

[13] ATLAS Collaboration, CMS Collaboration, "Top pair cross section measurements at the LHC", arXiv:1301.1158. Cited on page 3.

[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", *Phys.Lett.* **B717** (2012) 330–350, arXiv:1205.3130. doi:10.1016/j.physletb.2012.09.031. Cited on page 3.

[15] CMS Collaboration, "Measurement of the single-top-quark t-channel cross section in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **1212** (2012) 035, arXiv:1209.4533. doi:10.1007/JHEP12(2012)035. Cited on page 3.

[16] CDF Collaboration, "First Measurement of the $\mathrm{t\bar{t}}$ Differential Cross Section $d\sigma/dM_{t\bar{t}}$ in $\mathrm{p\overline{p}}$ Collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. Lett.* **102** (2009) 222003, arXiv:0903.2850. doi:10.1103/PhysRevLett.102.222003. Cited on page 3.

[17] CDF Collaboration, "Measurement of the Top Quark $\mathrm{p_T}$ Distribution", *Phys. Rev. Lett.* **87** (2001) 102001. Cited on page 3.

[18] D0 Collaboration, "Dependence of the $t\bar{t}$ production cross section on the transverse momentum of the top quark", *Phys. Lett. B* **693** (2010) 515, arXiv:1001.1900. doi:10.1016/j.physletb.2010.09.011. Cited on page 3.

[19] D0 Collaboration, "Measurement of differential $t\bar{t}$ production cross sections in $p\bar{p}$ collisions", arXiv:1401.5785. Cited on page 3.

[20] Holger Enderle, "Momentum Bias Determination in the Tracker Alignment and First Differential $\mathrm{t\bar{t}}$ Cross Section Measurement at CMS". PhD thesis, University of Hamburg, Germany, 2012. DESY-THESIS-2012-001. Cited on page 3.

[21] Martin Görner, "Messung von Myonen, Jets und Top-Quarks mit dem CMS-Detektor am LHC", Diploma thesis, University of Hamburg, 2010. CERN-THESIS-2010-176. Cited on page 3.

[22] ATLAS Collaboration, "Measurements of normalized differential cross-sections for $\mathrm{t\bar{t}}$ production in pp collisions at $\sqrt{\mathrm{s}} = 7$ TeV using the ATLAS detector", arXiv:1407.0371. Cited on page 3, 170, 256 and 303.

[23] Jörn Lange, "Differential Top-Quark-Pair Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV with CMS and Charge Multiplication in Highly-Irradiated Silicon Sensors". PhD thesis, University of Hamburg, 2013. DESY-THESIS-2013-042. Cited on page 3, 5, 21, 67, 71, 83, 88, 99, 120, 128, 165, 177, 179, 184, 193, 207, 211, 215, 255 and 256.

[24] Wolf Behrenhoff, "Measurement of Differential $\mathrm{t\bar{t}}$ Cross Sections in the Dilepton Decay Channels with the CMS Detector at 7 TeV Centre-of-Mass-Energy". PhD thesis, University of Hamburg, Germany, 2012. DESY-THESIS-2012-037. Cited on page 3 and 128.

[25] CMS Collaboration, "Measurement of differential top-quark pair production cross sections in pp colisions at $\sqrt{s} = 7$ TeV", *Eur.Phys.J.* **C73** (2013) 2339, arXiv:1211.2220. doi:10.1140/epjc/s10052-013-2339-4. Cited on page 3, 71, 75, 88, 94, 128, 211, 215, 232, 234, 255 and 256.

[26] CMS Collaboration, "Measurement of differential top-quark pair production cross sections in the lepton+jets channel in pp collisions at 8 TeV", CMS Physics Analysis Summary TOP-12-027, CERN, Geneva, (2013). Cited on page 3, 94, 147, 170, 175, 207, 232 and 255.

[27] CMS Collaboration, "Measurement of the differential ttbar cross section in the dilepton channel at 8 TeV", CMS Physics Analysis Summary TOP-12-028, CERN, Geneva, (2013). Cited on page 3, 211, 232, 234, 235 and 256.

[28] CMS Collaboration, "Measurement of Top Quark Pair Differential Cross Sections at $\sqrt{s} = 7\,\text{TeV}$", CMS Physics Analysis Summary TOP-11-013, CERN, Geneva, (2012). Cited on page 3 and 255.

[29] D. Griffiths, "Introduction to Elementary Particles". John Wiley & Sons, 1987. Cited on page 5.

[30] C. Berger, "Elementarteilchenphysik". Springer-Verlag Berlin Heidelberg, 2006. Cited on page 5.

[31] M. E. Peskin and D. V. Schroeder, "An Introduction to Quantum Field Theory". Westview Press, 1995. Cited on page 5.

[32] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **86** (2012) 010001. and 2013 partial update for the 2014 edition. doi:10.1103/PhysRevD.86.010001. Cited on page 5, 7, 8, 13, 14, 24, 67 and 191.

[33] Matthias Schröder, "Quality of Jet Measurements and Impact on a Search for New Physics at CMS". PhD thesis, University of Hamburg, 2012. DESY-THESIS-2012-042. Cited on page 5 and 63.

[34] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Physics Letters B* **716** (2012), no. 1, 30–61, arXiv:hep-ph/1207.7235. doi:10.1016/j.physletb.2012.08.021. Cited on page 8 and 14.

[35] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Physics Letters B* **716** (2012), no. 1, 1–29, arXiv:hep-ph/1207.7214. doi:10.1016/j.physletb.2012.08.020. Cited on page 8 and 14.

[36] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585 – 587. doi:10.1103/PhysRevLett.13.585. Cited on page 14.

[37] S. Dawson, "Introduction to electroweak symmetry breaking", arXiv:hep-ph/9901280. Cited on page 14.

[38] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, LEP Electroweak Working Group, SLD Heavy Flavor and Electroweak Groups Collaboration, "A Combination of preliminary electroweak measurements and constraints on the standard model", arXiv:hep-ex/0112021. Cited on page 14.

[39] Super-Kamiokande Collaboration Collaboration, "Atmospheric neutrino oscillation analysis with sub-leading effects in Super-Kamiokande I, II, and III", *Phys.Rev.* **D81** (2010) 092004, arXiv:1002.3471. doi:10.1103/PhysRevD.81.092004. Cited on page 14.

[40] Super-Kamiokande Collaboration, "A Measurement of the Appearance of Atmospheric Tau Neutrinos by Super-Kamiokande", *Phys.Rev.Lett.* **110** (2013) 181802, arXiv:1206.0328. doi:10.1103/PhysRevLett.110.181802. Cited on page 14.

[41] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction", *Progress of Theoretical Physics* **49** (1973), no. 2, 652–657. doi:10.1143/PTP.49.652. Cited on page 17.

[42] W. Bernreuther, "Top quark physics at the LHC", *J. Phys.* **G35** (2008) 083001, arXiv:0805.1333. doi:10.1088/0954-3899/35/8/083001. Cited on page 17.

[43] M. Beneke, I. Efthymiopoulos, M. L. Mangano et al., "Top quark physics", arXiv:hep-ph/0003033. Cited on page 17, 27 and 254.

[44] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_S^4)$", *Phys. Rev. Lett.* **110** (2013) 252004, arXiv:1303.6254. doi:10.1103/PhysRevLett.110.252004. Cited on page 19, 49, 108, 191 and 222.

[45] H1 and ZEUS Collaboration, "Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA", *JHEP* **1001** (2010) 109, arXiv:0911.0884. doi:10.1007/JHEP01(2010)109. Cited on page 19.

[46] H1 and ZEUS Collaboration, "PDF fits including HERA-II high Q^2 data (HERAPDF1.5)", *H1 and ZEUS reports* **H1prelim-10-142, ZEUS-prel-10-018**. [Online, accessed 20 June 2014]. Cited on page 19.

[47] N. Kidonakis, "Differential and total cross sections for top pair and single top production", arXiv:1205.3453. doi:10.3204/DESY-PROC-2012-02/251. Cited on page 20, 21 and 50.

[48] CMS Collaboration, "Jets at CMS and the determination of their energy scale". http://cms.web.cern.ch/news/jets-cms-and-determination-their-energy-scale. CMS public website. [Online, accessed 31 March 2014]. Cited on page 26 and 62.

[49] S. Alioli, P. Fernandez, J. Fuster et al., "A new observable to measure the top-quark mass at hadron colliders", *Eur.Phys.J.* **C73** (2013) 2438, arXiv:1303.6415. doi:10.1140/epjc/s10052-013-2438-2. Cited on page 26, 123, 184, 201, 244, 245, 252 and 257.

[50] A. Buckley, J. Butterworth, S. Gieseke et al., "General-purpose event generators for LHC physics", *Phys.Rept.* **504** (2011) 145–233, arXiv:1101.2599. doi:10.1016/j.physrep.2011.03.005. Cited on page 27.

[51] S. Naumann-Emme, "Simultaneous measurement of top quark mass and jet energy scale using template fits at the CMS experiment". PhD thesis, University of Hamburg, 2011. DESY-THESIS-2011-022. Cited on page 27.

[52] W. A. Bardeen, A. J. Buras, D. W. Duke et al., "Deep-inelastic scattering beyond the leading order in asymptotically free gauge theories", *Phys. Rev. D* **18** (Dec, 1978) 3998–4017. doi:10.1103/PhysRevD.18.3998. Cited on page 27.

[53] S. Moch, S. Weinzierl, S. Alekhin et al., "High precision fundamental constants at the TeV scale", arXiv:1405.4781. Cited on page 27 and 243.

[54] C. T. Hill, "Topcolor assisted technicolor", *Physics Letters B* **345** (1995) 483–489, arXiv:hep-ph/9411426. doi:10.1016/0370-2693(94)01660-5. Cited on page 29.

[55] R. M. Harris, C. T. Hill, and S. J. Parke, "Cross Section for Topcolor Z' decaying to top-antitop", arXiv:hep-ph/9911288. Fermilab-FN-687. Cited on page 29.

[56] R. M. Harris and S. Jain, "Cross Sections for Leptophobic Topcolor Z' decaying to top-antitop", *Eur. Phys. J. C* **72** (2012) 2072, arXiv:hep-ph/1112.4928. doi:10.1140/epjc/s10052-012-2072-4. Cited on page 29 and 165.

[57] D. Choudhury, R. M. Godbole, R. K. Singh et al., "Top production at the Tevatron/LHC and nonstandard, strongly interacting spin one particles", *Physics Letters B* **657** (2007), no. 1–3, 69–76, arXiv:hep-ph/07051499. doi:10.1016/j.physletb.2007.09.057. Cited on page 29.

[58] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo, "Phenomenology of the Randall-Sundrum Gauge Hierarchy Model", *Phys. Rev. Lett.* **84** (2000) 2080–2083, arXiv:hep-ph/9909255. doi:10.1103/PhysRevLett.84.2080. Cited on page 29.

[59] K. Agashe, H. Davoudiasl, S. Gopalakrishna et al., "CERN LHC signals for warped electroweak neutral gauge bosons", *Phys. Rev. D* **76** (2007) 115015, arXiv:hep-ph/0709.0007. doi:10.1103/PhysRevD.76.115015. Cited on page 29.

[60] K. Agashe, A. Belyaev, T. Krupovnickas et al., "CERN LHC signals from warped extra dimensions", *Phys. Rev. D* **77** (2008) 015003, arXiv:hep-ph/0612015. doi:10.1103/PhysRevD.77.015003. Cited on page 29.

[61] L. Randall and R. Sundrum, "Large Mass Hierarchy from a Small Extra Dimension", *Phys. Rev. Lett.* **83** (1999) 3370–3373, arXiv:hep-ph/9905221. doi:10.1103/PhysRevLett.83.3370. Cited on page 29.

[62] S. Choi and H. S. Lee, "Azimuthal decorrelation in $t\bar{t}$ production at hadron colliders", *Phys.Rev.* **D87** (2013), no. 3, 034012, arXiv:1207.1484. doi:10.1103/PhysRevD.87.034012. Cited on page 29, 196 and 258.

[63] I. Aitchison, "Supersymmetry in Particle Physics – An Elementary Introduction". Cambridge University Press, 2007. ISBN 978-0-521-88023-7. Cited on page 29.

[64] L. Evans, (ed.) and P. Bryant, (ed.), "LHC Machine", *JINST* **3** (2008) S08001. doi:10.1088/1748-0221/3/08/S08001. Cited on page 31.

[65] CMS Collaboration, "CMS physics: Technical design report", Technical Report CERN-LHCC-2006-001, (2006). Cited on page 31, 38 and 41.

[66] C. Lefèvre, "The CERN accelerator complex". http://cdsweb.cern.ch/record/1260465, 2008. CERN-DI-0812015. [Online, accessed 25 June 2014]. Cited on page 32.

[67] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002. doi:10.1088/1748-0221/3/08/S08002. Cited on page 32.

[68] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003. doi:10.1088/1748-0221/3/08/S08003. Cited on page 32.

[69] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004. Cited on page 32.

[70] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005. doi:10.1088/1748-0221/3/08/S08005. Cited on page 32.

[71] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *JINST* **3** (2008) S08006. Cited on page 32.

[72] TOTEM Collaboration, "The TOTEM Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08007. Cited on page 32.

[73] M. Lamont, "Status of the LHC", *J.Phys.Conf.Ser.* **455** (2013) 012001. doi:10.1088/1742-6596/455/1/012001. Cited on page 33.

[74] CMS Collaboration, "CMS Luminosity - Public Results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, 2014. [Online, accessed 25 June 2014]. Cited on page 33.

[75] CMS Collaboration, "Standard image of the CMS detector with labels for the sub-detectors". https://cms-docdb.cern.ch/cgi-bin/DocDB/ShowDocument?docid=2716. Online, accessed 26 June 2014]. Cited on page 34.

[76] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **03** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004. Cited on page 35.

[77] CMS Collaboration, "The CMS tracker system project", volume 5 of *Technical Design Report CMS*. CERN, Geneva, 1997. CMS-TDR-005. CERN-LHCC-98-006. Cited on page 35.

[78] CMS Collaboration, "The CMS tracker: addendum to the Technical Design Report", volume 5-add-1 of *Technical Design Report CMS*. CERN, Geneva, 2000. CMS-TDR-005-add-1. CERN-LHCC-2000-016. Cited on page 35.

[79] CMS Collaboration, "Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays", *JINST* **5** (2010) T03009, arXiv:0910.2505. doi:10.1088/1748-0221/5/03/T03009. Cited on page 36.

[80] CMS Collaboration Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", arXiv:1405.6569. Cited on page 36 and 37.

[81] CMS Collaboration, "The CMS electromagnetic calorimeter project", volume 4 of *Technical Design Report CMS*. CERN, Geneva, 1997. CMS-TDR-004. CERN-LHCC-97-033. Cited on page 37.

[82] P. Adzic, "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter", *Journal of Instrumentation* **2** (2007), no. 04, P04004. Cited on page 38.

[83] CMS Collaboration, "The CMS hadron calorimeter project", volume 2 of *Technical Design Report CMS*. CERN, Geneva, 1997. CMS-TDR-002. CERN-LHCC-97-031. Cited on page 38.

[84] CMS Collaboration, "Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data", *JINST* **5** (2010) T03012, arXiv:0911.4991. doi:10.1088/1748-0221/5/03/T03012. Cited on page 39.

[85] V. D. Elvira, "Measurement of the Pion Energy Response and Resolution in the CMS HCAL Test Beam 2002 Experiment", Technical Report CMS-NOTE-2004-020, CERN, Geneva, (Sep, 2004). Cited on page 40.

[86] CMS Collaboration, "The CMS muon project", volume 3 of *Technical Design Report CMS*. CERN, Geneva, 1997. CMS-TDR-003. CERN-LHCC-97-032. Cited on page 40.

[87] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7\,\text{TeV}$", *JINST* **7** (2012) P10002, arXiv:1206.4071. doi:10.1088/1748-0221/7/10/P10002. Cited on page 41 and 59.

[88] CMS Collaboration, "Interactive Slice of the CMS detector". https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=4172&version=2&filename=CMS_Slice.gif. CMS-doc-4172-v2. [Online, accessed 29 March 2014]. Cited on page 42 and 47.

[89] CMS Collaboration, "CMS TriDAS project: the trigger system", volume 6 of *Technical Design Report CMS*. CERN, Geneva, 2000. CMS-TDR-006-1. CERN-LHCC-2000-038. Cited on page 42.

[90] CMS Collaboration Collaboration, S. Cittolin, A. Rcz, and P. Sphicas, "CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project". Technical Design Report CMS. CERN, Geneva, 2002. Cited on page 43.

[91] CMS Collaboration, "CMS computing : Technical Design Report", volume 7 of *Technical Design Report CMS*. CERN, Geneva, 2005. CMS-TDR-007. CERN-LHCC-2005-023. Cited on page 43.

[92] CMS Collaboration, "Official CMS software framework". https://github.com/cms-sw/cmssw/tree/CMSSW_5_3_X. [Online, accessed 26 March 2014]. Cited on page 45.

[93] CMS Top-Quark Group, "Top-Quark-Analysis Framework (TQAF)". https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideTQAF. [Online, accessed 12 Nov. 2013]. Cited on page 45.

[94] CMS Top Analysis Groups of DESY and the University of Hamburg, "Private TopAnalysis Software Package". https://git.cern.ch/web/?p=TopAnalysis.git;a=tree. [Online, accessed 26 March 2014]. Cited on page 45 and 139.

[95] "DOXYGEN Documentation of the TopAnalysis software package". http://wwwiexp.desy.de/groups/cms/topphysics/TopAnalysis/annotated.html. [Online, accessed 26 March 2014]. Cited on page 45.

[96] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", CMS Physics Analysis Summary LUM-13-001, (2013). Cited on page 45 and 178.

[97] L. Tuura, A. Meyer, I. Segoni et al., "CMS data quality monitoring: Systems and experiences", *Journal of Physics: Conference Series* **219** (2010), no. 7, 072020. Cited on page 46.

[98] CMS Data Quality Monitoring Group, "Centrally provided good run JSON list ". https://cms-service-dqm.web.cern.ch/cms-service-dqm/CAF/certification/Collisions12/8TeV/Reprocessing/Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt. [Online, accessed 16 March 2014]. Cited on page 46.

[99] V. N. Gribov and L. N. Lipatov, "Deep inelastic *ep* scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438–450. Cited on page 46.

[100] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys. B* **126** (1977) 298. doi:10.1016/0550-3213(77)90384-4. Cited on page 46.

[101] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics", *Sov. Phys. JETP* **46** (1977) 641–653. Cited on page 46.

[102] M. Dobbs and J. Hansen, "HepMC 2 — a C++ Event Record for Monte Carlo Generators: User Manual Version 2.06", *Comput. Phys. Commun.* **134** (2001) 41. Cited on page 47.

[103] Cornell University, "Simulation of an electron shower in a calorimeter crystal.". http://www.lepp.cornell.edu/Research/EPP/CMS/rsrc/LEPP/Research/EPP/CMS/EventReconstruction/fe24elwhite_small.gif. [Online, accessed 29 March 2014]. Cited on page 47.

[104] CMS Top-Quark Analysis Group, "Overview about Top-Analysis-related Samples of the CMS Summer12 Simulation production cycle at $\sqrt{s} = 8$ TeV". https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopSamplesSummer12. [CMS internal, Online, accessed 26 March 2014]. Cited on page 48.

[105] J. Alwall, M. Herquet, F. Maltoni et al., "MadGraph 5 : Going Beyond", *JHEP* **1106** (2011) 128, arXiv:1106.0522. doi:10.1007/JHEP06(2011)128. Cited on page 48.

[106] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175. Cited on page 48.

[107] M. L. Mangano, M. Moretti, F. Piccinini et al., "Matching matrix elements and shower evolution for top- quark production in hadronic collisions", *JHEP* **01** (2007) 013, arXiv:hep-ph/0611129. doi:10.1088/1126-6708/2007/01/013. Cited on page 48.

[108] P. Artoisenet et al., "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations", *JHEP* **03** (2013) 015, arXiv:1212.3460. doi:10.1007/JHEP03(2013)015. Cited on page 48.

[109] N. Davidson, G. Nanava, T. Przedzinski et al., "Universal Interface of TAUOLA Technical and Physics Documentation", *Comput.Phys.Commun.* **183** (2012) 821–843, arXiv:1002.0543. doi:10.1016/j.cpc.2011.12.009. Cited on page 48.

[110] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **0411** (2004) 040, arXiv:hep-ph/0409146. doi:10.1088/1126-6708/2004/11/040. Cited on page 48.

[111] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **0711** (2007) 070, arXiv:0709.2092. doi:10.1088/1126-6708/2007/11/070. Cited on page 48.

[112] S. Alioli, P. Nason, C. Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **1006** (2010) 043, arXiv:1002.2581. doi:10.1007/JHEP06(2010)043. Cited on page 48.

[113] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **0206** (2002) 029, arXiv:hep-ph/0204244. doi:10.1088/1126-6708/2002/06/029. Cited on page 49.

[114] S. Frixione, P. Nason, and B. R. Webber, "Matching NLO QCD and parton showers in heavy flavor production", *JHEP* **0308** (2003) 007, arXiv:hep-ph/0305252. doi:10.1088/1126-6708/2003/08/007. Cited on page 49.

[115] G. Corcella et al., "HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)", *JHEP* **01** (2001) 010, arXiv:hep-ph/0011363. Cited on page 49.

[116] CMS Collaboration, "Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV", *JHEP* **1109** (2011) 109, arXiv:1107.0330. doi:10.1007/JHEP09(2011)109. Cited on page 49.

[117] ATLAS Collaboration, "New ATLAS event generator tunes to 2010 data", Technical Report ATL-PHYS-PUB-2011-008, CERN, Geneva, (Apr, 2011). Cited on page 49 and 217.

[118] J. Butterworth, J. R. Forshaw, and M. Seymour, "Multiparton interactions in photoproduction at HERA", *Z.Phys.* **C72** (1996) 637–646, arXiv:hep-ph/9601371. doi:10.1007/s002880050286. Cited on page 49.

[119] J. Butterworth and J. R. Forshaw, "Photoproduction of multi - jet events at HERA: A Monte Carlo simulation", *J.Phys.* **G19** (1993) 1657–1663. doi:10.1088/0954-3899/19/10/026. Cited on page 49.

[120] J. Pumplin, D. Stump, J. Huston et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **0207** (2002) 012, arXiv:hep-ph/0201195. doi:10.1088/1126-6708/2002/07/012. Cited on page 49 and 182.

[121] H.-L. Lai, M. Guzzi, J. Huston et al., "New parton distributions for collider physics", *Phys.Rev.* **D82** (2010) 074024, arXiv:1007.2241. doi:10.1103/PhysRevD.82.074024. Cited on page 49.

[122] J. M. Campbell and R. K. Ellis, "$t\bar{t}W^{+-}$ production and decay at NLO", *JHEP* **1207** (2012) 052, arXiv:1204.5678. doi:10.1007/JHEP07(2012)052. Cited on page 50.

[123] M. Garzelli, A. Kardos, C. Papadopoulos et al., "t $\bar{t}$ W^{+-} and t $\bar{t}$ Z Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects", *JHEP* **1211** (2012) 056, arXiv:1208.2665. doi:10.1007/JHEP11(2012)056. Cited on page 50.

[124] CMS Collaboration, "Measurement of the inclusive top-quark pair + photon production cross section in the muon + jets channel in pp collisions at 8 TeV", CMS Physics Analysis Summary TOP-13-011, CERN, Geneva, (2014). Cited on page 50.

[125] W. Kilian, T. Ohl, and J. Reuter, "WHIZARD: Simulating Multi-Particle Processes at LHC and ILC", *Eur.Phys.J.* **C71** (2011) 1742, arXiv:0708.4233. doi:10.1140/epjc/s10052-011-1742-y. Cited on page 50.

[126] K. Melnikov, M. Schulze, and A. Scharf, "QCD corrections to top quark pair production in association with a photon at hadron colliders", *Phys.Rev.* **D83** (2011) 074013, arXiv:1102.1967. doi:10.1103/PhysRevD.83.074013. Cited on page 50.

[127] J. M. Campbell and R. K. Ellis, "An Update on vector boson pair production at hadron colliders", *Phys.Rev.* **D60** (1999) 113006, arXiv:hep-ph/9905386. doi:10.1103/PhysRevD.60.113006. Cited on page 51.

[128] R. Gavin, Y. Li, F. Petriello et al., "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order", *Comput.Phys.Commun.* **182** (2011) 2388–2403, arXiv:1011.3540. doi:10.1016/j.cpc.2011.06.008. Cited on page 51.

[129] R. Gavin, Y. Li, F. Petriello et al., "W Physics at the LHC with FEWZ 2.1", *Comput.Phys.Commun.* **184** (2013) 208–214, arXiv:1201.5896. doi:10.1016/j.cpc.2012.09.005. Cited on page 51.

[130] P. Lenzi, S. Padhi, G. G. Ceballos Retuerto et al., "Collection of Standard Model cross sections at $\sqrt{s} = 8$ TeV to be used in CMS analyses". https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat8TeV. [CMS internal, Online, accessed 25 March 2014]. Cited on page 51.

[131] S. Agostinelli, J. Allison, K. Amako et al., "Geant4 – a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003), no. 3, 250 – 303. doi:10.1016/S0168-9002(03)01368-8. Cited on page 51.

[132] J. Allison, K. Amako, J. Apostolakis et al., "Geant4 developments and applications", *Nuclear Science, IEEE Transactions on* **53** (feb., 2006) 270 –278. doi:10.1109/TNS.2006.869826. Cited on page 51.

[133] CMS MC Simulation Group, "Pile-Up Modelling in CMS Simulations". https://twiki.cern.ch/twiki/bin/view/CMS/PdmVPileUpDescription#Pileup_Profiles_in_Simulation. [CMS internal, Online, accessed 17 June 2013]. Cited on page 53.

[134] CMS Luminosity Group, "The official CMS Luminosity Calculation Tools". https://twiki.cern.ch/twiki/bin/view/CMS/LumiCalc#How_to_use_script_estimatePileup. [CMS internal, Online, accessed 17 June 2013]. Cited on page 53.

[135] CMS Physics Validation Group, "Utilities for Accessing Pile-Up Information for Data". https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData. [CMS internal, Online, accessed 17 June 2013]. Cited on page 53.

[136] CMS Physics Validation Group, "Pile-Up Reweighting Utilities". https://twiki.cern.ch/twiki/bin/view/CMS/PileupMCReweightingUtilities. [CMS internal, Online, accessed 17 June 2013]. Cited on page 53.

[137] CMS Physics Validation Group, "Pile-Up Reweighting– Uncertaintiy Recommendation". https://twiki.cern.ch/twiki/bin/view/CMS/PileupSystematicErrors. [CMS internal, Online, accessed 17 June 2013]. Cited on page 54 and 178.

[138] CMS Top-Quark Analysis Group, "Top Physics Analysis Group Reference Selections". https://twiki.cern.ch/twiki/bin/view/CMS/TWikiTopRefEventSel. [CMS internal, Online, accessed 31 March 2014]. Cited on page 57.

[139] P. Lenzi, C. Genta, and B. Mangano, "Track reconstruction of real cosmic muon events with CMS tracker detector", *J. Phys. Conf. Ser.* **119** (2008) 032030. doi:10.1088/1742-6596/119/3/032030. Cited on page 58.

[140] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Meth. A* **262** (1987) 444–450. doi:10.1016/0168-9002(87)90887-4. Cited on page 58.

[141] CMS Collaboration, "Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary PFT-09-001, CERN, (2009). Cited on page 58.

[142] CMS Collaboration, "CMS public Software Guide - Description of the Particle Flow Algorithm". https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideParticleFlow#Description_of_the_Algorithm. [Online, accessed 30 March 2014]. Cited on page 58.

[143] F. Beaudette, "Performance of the Particle Flow Algorithm in CMS". http://indico.cern.ch/event/73513/session/45/contribution/825/material/slides/0.pdf, 22.07.1010. Presentation at ICHEP - 35th International Conference on High Energy Physics, Paris. July 2010. [Online, accessed 31 March 2014]. Cited on page 59.

[144] CMS Collaboration, "CMS public Software Guide - Description of the Particle Flow Implementation in the Physics Analysis Toolkit". https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuidePF2PAT. [Online, accessed 30 March 2014]. Cited on page 59.

[145] M. Mulders, I. Bloch, E. James et al., "Muon Identification in CMS", CMS Analysis Note 2008/098, CERN, (2008). Cited on page 59 and 60.

[146] S. Baffioni, C. Charlot, F. Ferri et al., "Electron reconstruction in CMS", Technical Report CMS-NOTE-2006-040, CERN, Geneva, (Feb, 2006). Cited on page 60.

[147] W. Adam, R. Fruhwirth, A. Strandlie et al., "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC", *J. Phys. G: Nucl. Part. Phys.* **31** (2005) N9. doi:10.1088/0954-3899/31/9/N01. Cited on page 61.

[148] CMS E/Gamma Physics Objects Group, "Tools for conversion rejection (electron ID) and electron vetoing (photon ID)". `https://twiki.cern.ch/twiki/bin/view/CMS/ConversionTools`. [CMS internal, Online, accessed 30 March 2014]. Cited on page 61.

[149] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, `arXiv:0802.1189`. `doi:10.1088/1126-6708/2008/04/063`. Cited on page 62.

[150] CMS Collaboration, "Plans for Jet Energy Corrections at CMS", CMS Physics Analysis Summary JME-07-002, CERN, 2008. Geneva, (Jul, 2008). Cited on page 62.

[151] CMS Jet Energy Resolution and Corrections Group, "Jet Energy Corrections in CMS". `https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetEnergyScale`. [CMS internal, Online, accessed 01 April 2014]. Cited on page 62.

[152] CMS Collaboration, "Jet Energy Corrections determination at 7 TeV", CMS Physics Analysis Summary JME-10-010, CERN, Geneva, (2010). Cited on page 62.

[153] CMS Collaboration, "8 TeV Jet Energy Corrections and Uncertainties based on $19.8\,\text{fb}^{-1}$ of data in CMS", CMS Detector Performance Note CMS-DP-2013-033, CERN, Geneva, (2013). Cited on page 62 and 257.

[154] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets", *JHEP* **0804:005** (2007). `doi:10.1088/1126-6708/2008/04/005`. Cited on page 62.

[155] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2007) 119–126. `doi:10.1016/j.physletb.2007.09.077`. Cited on page 62.

[156] CMS Jet Energy Resolution and Corrections Group, "Recommended Jet Energy Corrections and Uncertainties For Data and MC". `https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECDataMC`. [CMS internal, Online, accessed 01 April 2014]. Cited on page 63.

[157] A. B. Galtieri, F. Margaroli, and I. Volobouev, "Precision measurements of the top quark mass from the Tevatron in the pre-LHC era", *Reports on Progress in Physics* **75** (2012), no. 5, 056201. Cited on page 64.

[158] CMS B Tag & Vertexing Physics Object Group, "B-jet Identification in CMS". `https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagPOG`. [CMS internal, Online, accessed 02 April 2014]. Cited on page 65.

[159] CMS Collaboration, "Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events", CMS Physics Analysis Summary BTV-13-001, CERN, Geneva, (2013). Cited on page 65 and 71.

[160] CMS Collaboration, "CMS Software Guide– B-Tagging". `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookBTagging`. [Online, accessed 02 April 2014]. Cited on page 65.

[161] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** (2011) P09001, `arXiv:1106.5048`. `doi:10.1088/1748-0221/6/09/P09001`. Cited on page 65.

[162] CMS Collaboration, "MET performance in 8 TeV data", CMS Physics Analysis Summary JME-12-002, CERN, Geneva, (2013). Cited on page 65.

[163] C. Clopper and E. Pearson, "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial", *Biometrika* **26** (1934) 404. Cited on page 68 and 259.

[164] Jörn Lange. Private Communication. Cited on page 69 and 70.

[165] CMS B Tag & Vertexing Physics Object Group, "Recommendation for b/c-tagging and mistagging Data/MC Scale Factors". `https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagPOG#Recommendation_for_b_c_tagging_a`. [CMS internal, Online, accessed 10 April 2014]. Cited on page 71.

[166] CMS Collaboration Collaboration, "CMS reconstruction improvement for the muon tracking by the RPC chambers", *PoS* **RPC2012** (2012) 045, `arXiv:1209.2646`. `doi:10.1088/1748-0221/8/03/T03001`. Cited on page 77.

[167] CMS Collaboration, "Fitting of Event Topologies with External Kinematic Constraints in CMS", CMS Note 2006/023, Geneva, (2006). Cited on page 84 and 85.

[168] CMS Top-Quark Analysis Group, "Implementation of a constrained Kinematic Fitting tool for the reconstruction of $t\bar{t}$ events". `https://github.com/cms-sw/cmssw/tree/CMSSW_5_3_X/TopQuarkAnalysis/TopKinFitterr`. [Online, accessed 12 April 2014]. Cited on page 84.

[169] CMS Collaboration, "Measurement of the top-quark mass in $t\bar{t}$ events with lepton+jets final states in pp collisions at $\sqrt{s} = 8$ TeV TeV", CMS Physics Analysis Summary TOP-14-001, CERN, Geneva, (2014). Cited on page 95, 105, 181, 183, 243, 252 and 254.

[170] ATLAS, CMS Collaboration, "Combination of ATLAS and CMS results on the mass of the top quark using up to 4.9 inverse femtobarns of data", CMS Physics Analysis Summary TOP-13-005, CERN, Geneva, (2013). Cited on page 95 and 243.

[171] Tevatron Electroweak Working Group, CDF Collaboration, D0 Collaboration, "Combination of CDF and DO results on the mass of the top quark using up to 8.7 fb^{-1} at the Tevatron", `arXiv:1305.3929`. Cited on page 95 and 243.

[172] ATLAS Collaboration, CDF Collaboration, CMS Collaboration, D0 Collaboration, "First combination of Tevatron and LHC measurements of the top-quark mass", `arXiv:1403.4427`. Cited on page 95, 243 and 251.

[173] D0 Collaboration, "Precision measurement of the top-quark mass in lepton+jets final states", `arXiv:1405.1756`. Cited on page 95, 243, 252 and 257.

[174] CMS Collaboration, "Measurement of the top-quark mass in $t\bar{t}$ events with all-jets final states in pp collisions at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary TOP-14-002, CERN, Geneva, (2014). Cited on page 95, 105, 183, 243 and 252.

[175] Eike Schlieckau, "Measurement of Top-Quark Mass and Inclusive Top-Quark-Pair Production Cross Section in pp Collisions at $\sqrt{s} = 7/8$ TeV with CMS". PhD thesis, University of Hamburg, 2014. Cited on page 95 and 243.

[176] Holger Enderle, "Resolutions for the Kinematic Fit in the Top-Quark-Analysis Framework". https://github.com/cms-sw/cmssw/tree/CMSSW_5_3_X/TopQuarkAnalysis/TopObjectResolutions/python/stringResolutions_etEtaPhi_Fall11_cff.py. [Online, accessed 26 May 2014]. Cited on page 99.

[177] M. Mangano, "Interpreting the top quark mass: theoretical and MC aspects". https://indico.desy.de/getFile.py/access?contribId=30&sessionId=9&resId=0&materialId=slides&confId=7095. Presentation at the TOP 2013 - 6th International Workshop on Top Quark Physics, Durbach. Sep. 2013. [Online, accessed 21 April 2014]. Cited on page 118 and 243.

[178] Maria Aldaya and Francesco Spano for the ATLAS & CMS top-quark pair differential cross section groups , "Review of ATLAS & CMS top-quark pair differential cross sections ". https://indico.cern.ch/event/280522/session/2/material/slides/0?contribId=7. Presentation at the Top-Quark LHC Working Group Meeting- CERN, 28 November 2013. [Online, LHC internal, accessed 21 April 2014]. Cited on page 119, 120, 217 and 258.

[179] G. Corcella, I. G. Knowles, G. Marchesini et al., "HERWIG 6.5 Manual". http://webber.home.cern.ch/webber/hw65_manual.html#htoc96. [Online, accessed 21 April 2014]. Cited on page 119.

[180] Top-Quark LHC Working Group, "Particle level obects and pseudo top definitions". https://twiki.cern.ch/twiki/bin/view/LHCPhysics/ParticleLevelTopDefinitions. [Online, accessed 22 April 2014]. Cited on page 121.

[181] V. Blobel, "A survey of unfolding methods for particle physics". https://www.wiki.terascale.de/images/a/a8/Blobel_Unfold_Note_20100524.pdf, 2002. Prepared for Alliance Workshop on Unfolding and Data Correction, DESY, Hamburg, Germany, 27-28 May 2002. [Online; accessed 24 April 2014]. Cited on page 136 and 139.

[182] A. Hocker and V. Kartvelishvili, "SVD approach to data unfolding", *Nucl.Instrum.Meth.* **A372** (1996) 469–481, arXiv:hep-ph/9509307. doi:10.1016/0168-9002(95)01478-0. Cited on page 138 and 170.

[183] K. Tackmann, A. Hoecker, and H. Lacker, "TSVDUnfold Software Package". http://root.cern.ch/root/html/TSVDUnfold.html. [Online; accessed 25 June 2014]. Cited on page 138.

[184] F. James, "Statistical Methods in Experimental Physics". World Scientific, 2nd edition, 2006. Cited on page 140.

[185] V. Blobel, "Some Comments on χ^2 Minimization Applications", *eConf* **C030908** (2003) MOET002. Cited on page 140.

[186] David-Johannes Fischer, "Inclusive neutral current ep cross sections with HERA II and two-dimensional unfolding". PhD thesis, DESY, 2011. DESY-THESIS-2011-020. Cited on page 140.

[187] CMS Collaboration, "8 TeV Jet Energy Corrections and Uncertainties based on 19.8 fb^{-1} of data in CMS", cms detector performance note, (Oct, 2013). CMS-DP-2013-033. Cited on page 178.

[188] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV with lepton + jets final states", *Phys.Lett.* **B720** (2013) 83–104, arXiv:1212.6682. doi:10.1016/j.physletb.2013.02.021. Cited on page 181.

[189] M. Gosselink, "Top quark signal modeling uncertainties at CMS". http://indico.cern.ch/contributionDisplay.py?contribId=17&confId=180665. Presentation at the TOP 2012 - 5th International Workshop on Top Quark Physics, Winchester. Sep. 2012. [Online, accessed 02 May 2013]. Cited on page 181.

[190] L. Hung-Liang et al., "New parton distributions for collider physics", *Phys. Rev. D* **82** (2010) 074024, arXiv:hep-ph/1007.2241. doi:10.1103/PhysRevD.82.074024. Cited on page 182.

[191] J. Rojo, "Top quark production at the LHC as a gluon luminometer". https://indico.cern.ch/getFile.py/access?contribId=12&sessionId=1&resId=0&materialId=slides&confId=217721. Presentation at the TOP LHC WG meeting, CERN. Nov. 2012. [Online, accessed 04 July 2014]. Cited on page 184 and 258.

[192] M. Czakon, M. L. Mangano, A. Mitov et al., "Constraints on the gluon PDF from top quark pair production at hadron colliders", *JHEP* **1307** (2013) 167, arXiv:1303.7215. doi:10.1007/JHEP07(2013)167. Cited on page 184 and 258.

[193] Jochen Ott and Thomas Peiffer. Private Communication. Cited on page 184.

[194] CMS Collaboration, "Top pair cross section in e/mu+jets at 8 TeV", CMS Physics Analysis Summary TOP-12-006, CERN, Geneva, (2012). Cited on page 187, 188 and 190.

[195] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **1402** (2014) 024, arXiv:1312.7582. doi:10.1007/JHEP02(2014)024. Cited on page 187, 188 and 190.

[196] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, arXiv:1211.4462. doi:10.1088/1748-0221/8/04/P04013. Cited on page 190.

[197] Holger Enderle, "b-tagging Efficiencies with Different Event Generators". https://indico.cern.ch/event/243715/session/0/material/slides/0?contribId=10. Presentation at the b-tagging Subgroup Meeting- CERN, 25 April 2013. [Online, CMS internal, accessed 05 May 2014]. Cited on page 190.

[198] Annika Vanhoefer, "D-hadron generator studies". https://indico.cern.ch/event/303434/session/2/contribution/5/material/slides/0.pdf. Presentation at the CMS b-tagging Subgroup Meeting- CERN, 13 March 2014. [Online, CMS internal, accessed 06 May 2014]. Cited on page 190.

[199] Daniel Bloch, Caroline Collard, Kirill Skovpen, "Measurement of b- and c-jet efficiency scale factors with the Lifetime tag method". https://indico.cern.ch/event/303434/session/2/contribution/3/material/slides/0.pdf. Presentation at the BTAG POG Meeting- CERN, 13 March 2014. [Online, CMS internal, accessed 05 May 2014]. Cited on page 190.

[200] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", arXiv:1112.5675. Cited on page 191.

[201] M. Botje et al., "The PDF4LHC Working Group Interim Recommendations", arXiv:1101.0538. Cited on page 191.

[202] Ahrens, Valentin and Ferroglia, Andrea and Neubert, Matthias and Pecjak, Ben D. and Yang, Li Lin, "Renormalization-Group Improved Predictions for Top-Quark Pair Production at Hadron Colliders", *JHEP* **1009** (2010) 097, arXiv:1003.5827. doi:10.1007/JHEP09(2010)097. Cited on page 193, 205 and 304.

[203] Ferroglia, Andrea and Pecjak, Ben D. and Yang, Li Lin, "Top-quark pair production at high invariant mass: an NNLO soft plus virtual approximation", *JHEP* **1309** (2013) 032, arXiv:1306.1537. doi:10.1007/JHEP09(2013)032. Cited on page 193, 194, 205 and 304.

[204] Li, Hai Tao and Li, Chong Sheng and Shao, Ding Yu and Yang, Li Lin and Zhu, Hua Xing, "Top quark pair production at small transverse momentum in hadronic collisions", *Phys.Rev.* **D88** (2013) 074004, arXiv:1307.2464. doi:10.1103/PhysRevD.88.074004. Cited on page 193, 194, 205 and 304.

[205] N. Kidonakis, "NNLL threshold resummation for top-pair and single-top production", arXiv:1210.7813. Cited on page 193, 195, 196, 205, 221, 232, 256 and 304.

[206] CMS Collaboration, "Measurement of jet multiplicity distributions in $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV", arXiv:1404.3171. Cited on page 201.

[207] CMS Collaboration, "Measurement of the Jet Multiplicity in dileptonic Top Quark Pair Events at 8 TeV", CMS Physics Analysis Summary TOP-12-041, (2013). Cited on page 201.

[208] ATLAS Collaboration, "Measurement of the jet multiplicity in top anti-top final states produced in 7 TeV proton-proton collisions with the ATLAS detector", ATLAS Conference Note ATLAS-CONF-2012-155, ATLAS-COM-CONF-2012-164, (2012). Cited on page 201.

[209] G. D. Lafferty and T. R. Wyatt, "Where to stick your data points: The treatment of measurements within wide bins", *Nucl. Instrum. Meth. A* **A355** (1995) 541–547. doi:10.1016/0168-9002(94)01112-5. Cited on page 210.

[210] ATLAS Collaboration, "Measurement of top-quark pair differential cross-sections in the l+jets channel in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector", Technical Report ATLAS-CONF-2013-099, CERN, Geneva, (Sep, 2013). Cited on page 215, 216 and 256.

[211] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes", *Phys.Rev.* **D82** (2010) 074018, arXiv:1005.3457. doi:10.1103/PhysRevD.82.074018. Cited on page 217 and 231.

[212] R. Field, "Min-bias and the underlying event at the LHC", *Acta Physica Polonica B* **42** (2011) 2631. doi:10.5506/APhysPolB.42.2631. Cited on page 217.

[213] Paolo Nason, "Issues on the top pt distribution in ME+PS and NLO+PS interfaced to Herwig". https://indico.cern.ch/event/301787/session/10/contribution/23/material/slides/0.pdf. Presentation at the Top-Quark LHC Working Group Meeting- CERN, 22 May 2014. [Online, accessed 22 May 2014]. Cited on page 217, 222 and 256.

[214] CMS Collaboration, "Measurement of the top quark mass using the B-hadron lifetime technique", CMS Physics Analysis Summary TOP-12-030, (2013). Cited on page 236 and 257.

[215] CMS Collaboration, "Measurements of $t\bar{t}$ spin correlations and top-quark polarization using dilepton final states in pp collisions at $\sqrt{s} = 7$ TeV", arXiv:1311.3924. Cited on page 236 and 257.

[216] CMS Collaboration, "Search for Higgs Boson Production in Association with a Top-Quark Pair and Decaying to Bottom Quarks or Tau Leptons", CMS Physics Analysis Summary HIG-13-019, CERN, Geneva, (2013). Cited on page 236 and 257.

[217] CMS Collaboration, "Determination of the top-quark pole mass and strong coupling constant from the production cross section in pp collisions at $\sqrt{s} = 7$ TeV", *Physics Letters B* **728** (2014), no. 0, 496 – 517. doi:10.1016/j.physletb.2013.12.009. Cited on page 243 and 247.

[218] G. D'Agostini, "Probability and measurement uncertainty in physics: A Bayesian primer", arXiv:hep-ph/9512295. Cited on page 247.

[219] CMS Collaboration, "Measurement of the top-quark mass in all-jets $t\bar{t}$ events in pp collisions at $\sqrt{s}$=7 TeV", *Eur.Phys.J.* **C74** (2014) 2758, arXiv:1307.4617. doi:10.1140/epjc/s10052-014-2758-x. Cited on page 254.

[220] CMS Jet Energy Resolution and Corrections Group, "Jet energy scale uncertainty sources". https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECUncertaintySources. [CMS internal, Online, accessed 02 May 2014]. Cited on page 257.

[221] J. Shelton, "TASI Lectures on Jet Substructure", arXiv:1302.0260. Cited on page 258.

[222] T. Plehn and M. Spannowsky, "Top Tagging", *J.Phys.* **G39** (2012) 083001, arXiv:1112.4441. doi:10.1088/0954-3899/39/8/083001. Cited on page 258.

[223] W. H. Bell, "Common top (pair) acceptance: definitions". https://indico.cern.ch/getFile.py/access?contribId=9&resId=0&materialId=slides&confId=189617. [Online, accessed 02 May 2014]. Cited on page 258.

[224] R. Chierici et al., "Particle level objects and pseudo top definitions". https://twiki.cern.ch/twiki/bin/view/LHCPhysics/ParticleLevelTopDefinitions. [Online, accessed 02 May 2014]. Cited on page 258.

Danksagung

Wissenschaft ist etwas großartiges. Einerseits konzeptionell großartig, weil man in einem internationalen Arbeitsumfeld über alle kulturellen Grenzen hinweg an den spannenden und großen Fragen der Menschheit arbeiten darf. Dabei befriedigt man nicht nur seine eigene Neugier und trägt einen (ganz kleinen) Teil zum Fortschritt der Menschheit bei, sondern baut auch Vorurteile ab und leistet somit einen kleinen Beitrag zu einem friedlichen Umgang miteinander.
Vor allen Dingen aber ist Wissenschaft großartig, weil die Menschen, denen man begegnet, einzigartig sind. Ich möchte daher die Gelegenheit nutzen, um all den Leuten zu danken, die mich in den letzten Jahren zumindest ein Stück meines Weges begleitet und auf Ihre Weise zum Gelingen dieser Arbeit beigetragen haben.
Ich danke Prof. Dr. Peter Schleper, der es mir ermöglicht hat, diese Arbeit zu schreiben, und sich darüber hinaus noch bereit erklärt hat, sie ganz zu lesen und zu begutachten. Danke auch an meine Gutachter PD Dr. Olaf Behnke, PD Dr. Andreas Meyer und Jun.-Prof. Christian Sander, die vieles ermöglicht haben.
Weiterhin danke ich den vielen Leuten, die auf ihre Weise dazu beigetragen haben, diese Arbeit zu dem zu machen, was sie ist. Die Inspiration, das Zusammengehörigkeitsgefühl und die Unterstützung, die ich in den letzten Jahren sowohl innerhalb der Arbeitsgruppe als auch von privater Seite von meiner Lebensgefährtin, Familie und Freunden erfahren durfte, sind einzigartig.
Wenn ich nun am Ende meiner Doktorarbeit zurückblicke, dann sehe ich nicht nur Betreuer und Arbeitskollegen, sondern vor allem Ratgeber und Freunde über die Grenzen der Arbeit hinweg. Danke meine lieben Betreuer: Roger, Thomas, Martijn, Maria und Hartmut – ich habe von jedem von Euch etwas unterschiedliches gelernt und jeder war auf seine Weise als Betreuer großartig. Danke Jörn, meinem Bruder im Geiste, und Holger, der großen Konstanten in den letzten Jahren – wir drei waren in unserer gemeinsamen Zeit ein unschlagbares Team und uns verbindet mehr als nur die Arbeit. Danke Eike, dass du es vom ersten bis zum letzten Tag in einem Büro mit mir ausgehalten hast. Unseren Spaß konnte uns auch an schlechten Tagen niemand nehmen! Danke Andrey – für deinen einzigartigen Humor und deine unerschütterliche moralische Integrität. Danke an all die vielen Leute in der Arbeitsgruppe, die mit Rat, Tat und manchmal einfach mit guter Laune zur Seite standen: Alex, Annika, Bene, Dennis, Henning, Jochen, Markus, Matthias, Roman, Thomas und viele mehr.
Danke auch an die ganze DESY-Bande für die zahlreichen Jahre der erfolgreichen Zusammenarbeit: Carmen, Ivan, Sebastian, Wolf und all die anderen.
Gerade die Endphase der Doktorarbeit war ein bißchen wie der Kopfstand auf einer Rasierklinge – meine Arbeit, das Paper und natürlich das kleine Wunder zu Hause. Da hätte ich mir auch das ein oder andere Mal den Zustand des Milchkomas herbeigewünscht, der kleine Babys nach dem Stillen friedlich schlummern lässt. An dieser Stelle ein dickes Dankeschön an meinen kleinen Sohnemann Collin, dass du Papa Nachts genug Schlaf gegönnt hast und an Franzi für die uneingeschränkte Unterstützung und zahlreichen Entbehrungen. Danke an meine Familie, insbesondere meine Eltern, die mich immer nach Kräften unterstützt haben und in allen Lebenslagen hinter mir stehen und an Franzis Eltern, die mir in vielen Dingen den Rücken frei gehalten haben. Danke an all die Freunde, die einen Menschen erst zu dem machen, was er ist – Danke Aila, Andrey, Asi, Eike, Flo, Holger, Jan, Julia, Jörn, Markus, Max, Yasar und alle, die ich jetzt vergessen habe.
Die letzten Zeilen dieser Arbeit sollen meinem Vater gelten: Auch wenn unser gemeinsamer Weg enden musste, danke ich dir für alles, was du mir in deiner unerreichten Art alles gegeben hast – es wird mich immer begleiten.