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

Dissertation
zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von

Wolf Behrenhoff

aus Buxtehude

Hamburg
2012

Gutachter der Dissertation:	PD Dr. Achim Geiser Prof. Dr. Peter Schleper
Gutachter der Disputation:	PD Dr. Achim Geiser JProf. Dr. Christian Sander
Datum der Disputation:	14. September 2012
Vorsitzender des Prüfungsausschusses:	Dr. Georg Steinbrück
Vorsitzender des Promotionsausschusses:	Prof. Dr. Peter Hauschildt
Leiter des Departments Physik:	Prof. Dr. Daniela Pfannkuche
Dekan der MIN-Fakultät:	Prof. Dr. Heinrich Graener

Abstract

After commissioning of the Large Hadron Collider at the European nuclear research centre CERN, the CMS detector had recorded data with an integrated luminosity of 1.1 fb^{-1} at a centre-of-mass energy of 7 TeV already in Summer 2011. With this, it was possible for the first time to accurately measure properties of the top quark with high statistics.

The work presented here focuses on the measurement of normalised differential top quark pair production cross sections in the decay channel into final states with two oppositely charged leptons. The measurement is presented in dependence of kinematic properties of the top quarks, the top quark pair, the leptons, and the lepton pair.

The distributions allow for detailed testing of perturbative quantum chromodynamics and thus the Standard Model of particle physics. Many of the distributions are also sensitive to processes beyond the Standard Model.

In this work, comparisons of the measured cross sections with three different predictions are shown. The result is that the predictions are in accordance with the measurement, only a slight deviation of the prediction from the measurement is visible in few distributions. In the outlook it is shown that a fourth, more accurately calculated prediction for the measurement of the top quark transverse momentum normalised differential cross section is in even better agreement with the data. Thus it can be concluded that the Standard Model was tested successfully and that no hints for new physics are found.

Zusammenfassung

Nach Inbetriebnahme des Large Hadron Colliders am europäischen Kernforschungszentrum CERN wurden mit dem CMS-Detektor bereits bis zum Sommer 2011 Daten mit einer integrierten Luminosität von 1.1 fb^{-1} bei einer Schwerpunktsenergie von 7 TeV aufgezeichnet. Damit war es erstmals möglich, Eigenschaften des Top-Quarks mit hoher Statistik detailliert zu vermessen.

Die hier vorgestellte Arbeit beschäftigt sich mit der Messung normierter differentieller Produktionswirkungsquerschnitte von Top-Antitop-Paaren im Zerfallskanal in zwei entgegengesetzt geladene Leptonen. Die Messung wird in Abhängigkeit von kinematischen Eigenschaften der Top-Quarks, des Top-Quark-Paars, der Leptonen und des Leptonen-Paars durchgeführt.

Die Verteilungen ermöglichen es, die perturbative Quantenchromodynamik und damit das Standardmodell der Teilchenphysik detailliert zu testen. Viele der Verteilungen sind zudem sensitiv auf Prozesse jenseits des Standardmodells.

In der Arbeit wird ein Vergleich der gemessenen Wirkungsquerschnitte mit drei verschiedenen Vorhersagen durchgeführt. Als Ergebnis zeigt sich, dass die Vorhersagen generell sehr gut mit der Messung übereinstimmen, nur in wenigen Fällen ist eine kleine Abweichung der Vorhersagen von der Messung sichtbar. Im Ausblick wird dann für die Wirkungsquerschnittsmessung in Abhängigkeit des Transversalimpulses der Top-Quarks gezeigt, dass eine vierte, genauer gerechnete Vorhersage noch besser mit den Daten übereinstimmt. Somit wurde das Standardmodell erfolgreich getestet und es haben sich keine Hinweise auf neue Physik ergeben.

Contents

Chapter 1

Introduction

Ideas about the existence of the top quark date back to the year 1973 when a third generation of quarks was predicted by Kobayashi and Maskawa [1]. While it took only four years to discover the bottom quark in 1977 [2], more than two decades passed until the top quark was finally discovered [3, 4] by the CDF and DØ experiments at the Tevatron accelerator in 1995. Several studies followed, for example about the mass or the production cross section.

Seventeen years later the top quark is still the heaviest known elementary particle. The Tevatron has been shut down and the LHC is in operation. Due to its higher centre-of-mass energy the production cross section for top quarks at the LHC is larger than at the Tevatron, therefore the LHC provides much better opportunities to study the top quark in more detail.

Top quarks are dominantly produced in pairs. Each top quark decays into a bottom quark and a W boson which decays further into quarks or a lepton and a neutrino. In this work only the dilepton channel is analysed, i.e. those top quark pairs where both W bosons decay into a final state with a lepton. This is done with the first 1.14 fb^{-1} of data with 7 TeV centre-of-mass energy, taken in 2011.

The result of this work is the first normalised differential top quark pair production cross section measurement at the LHC. Nine different variables are shown and compared to theory predictions.

This work is structured in the following way. Chapter 2 gives a brief introduction to the Standard Model of particle physics. The production and the decay of top quarks are explained and also the relevance of top quark physics for the discovery of the Higgs boson is presented.

Chapter 3 describes the experimental setup. This includes an introduction about the LHC and a more detailed description of the CMS detector, its sub-detectors, and the data taking process.

The next chapter, chapter 4, summarises the technical work that has been done for the collaboration in the computing field. It first explains how the CMS experiment software is installed and then covers the development of a program to automatically install the CMS software on grid sites.

Chapter 5 returns to physics and explains how collisions of two protons can be simu-

lated. Several event generators with different levels of accuracy are available, and their main features are listed.

Many particles are created in proton-proton collisions. Chapter 6 explains which signals are measured in the detector and how the reconstruction of the leptons, jets, and transverse momentum imbalance works.

The top quark events examined in this work need to be extracted from the huge amount of all recorded events. Chapter 7 describes the necessary steps to select only interesting events. It also explains how the properties of the top quarks are reconstructed from the measured particles.

Studies on the selection efficiency are presented in chapter 8. For the estimation of background events, several scaling factors are applied to correct for differences between the simulation and the data. In addition, this chapter explains how the uncertainties of the final result have been determined.

The final result of this work is presented in chapter 9. First the total inclusive $\mathrm{t\bar{t}}$ cross section is calculated as a cross check, then normalised differential cross sections are presented in nine different variables. Meanwhile this analysis has been extended to a dataset containing the full 2011 dataset. These updated results are shown chapter 10.

Finally a summary about all the work achieved in this thesis is given in chapter 11.

Additional information about the analysis is given in the appendices. Differential cross sections for the individual ee, $e\mu$ and $\mu\mu$ channel are given in appendix A, their uncertainties in appendix B. The technical steps to retrieve smooth predicted differential cross sections are explained in appendix C. Closure tests about the binwise unfolding and comparisons to SVD unfolding are shown in appendix D. This work ends with appendix E, a short description of a self-developed software to split analysis jobs.

Chapter 2

The Standard Model

2.1 History

Already in ancient times people were interested in the constituents of matter. Back then it was more a philosophical question if one could split a piece of matter infinitely often or if at some point one would reach elementary particles which are indivisible. The famous Greek philosopher Demokrit (about 460–370 B.C.) [5] established the term *atomos* (Greek for *uncuttable*). More then one thousand year later, in the beginning of the 19th century, John Dalton explained that chemical reactions of elements always happen with a fixed ratio of small natural numbers by using the concept of atoms. Each element consists of a certain type of atoms and these atoms can react with other types of atoms to form compound substances. In 1897 Joseph John Thomson discovered the electron [5] and found that it had negative charge and only a small mass. Atoms were neutral and heavier, so Thomson assumed that the mass was distributed homogeneously in the atom with the electrons in between (plum pudding model). This was found to be incorrect with Ernest Rutherford's interpretation of the Gold Foil Experiment. Its results could be explained by a small and heavy core in the centre of the atom which was orbited by electrons. Today we also know that the cores of atoms are not elementary but consist of protons and neutrons and that these nucleons are made of quarks and gluons.

The Standard Model of particle physics [6] tries to describe everything that is known about elementary particles and their interactions in a quantum field theory. It is however incomplete because, for example it doesn't include gravitation. The masses of the particles are not predicted and also neutrino oscillations are not described.

2.2 Particle types

Our present knowledge is that all matter consists of elementary fermions while the interactions are mediated by gauge bosons.

A fermion is an elementary or compound particle with half-integer spin ($\frac{1}{2}\hbar, \frac{3}{2}\hbar, \dots$) while a boson has integer spin ($0\hbar, 1\hbar, \dots$). A wave function ψ of two identical particles behaves differently for fermions and bosons when exchanging both particles: for bosons,

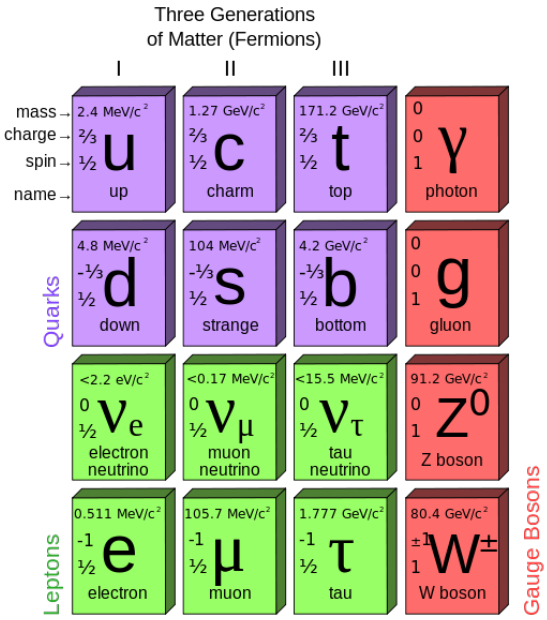


Figure 2.1: Partices in the Standard Model [7]. The diagram shows quarks (purple), leptons (green), and gauge bosons (red) with their mass, charge and spin.

ψ is transformed into itself while for fermions, ψ is transformed into $-\psi$. A system of fermions follows the Fermi-Dirac statistics while bosons follow Bose-Einstein statistics.

Known particles are quarks and leptons. There are two types of quarks, positively charged *up-type* and negatively charged *down-type* quarks. The lightest quarks are the u (up) and d (down) quarks which form the first generation. There are two more generations, consisting of the c (charm) and s (strange) quark in the second and the t (top) and b (bottom) quark in the third generation.

There are also two types of leptons which come in three generations. The negatively charged e^- (electron), μ^- (muon), and τ^- (tau) and the electrically neutral electron-, muon- and tau neutrinos ν_e, ν_μ and ν_τ.

Figure 2.1 shows these fermions with their properties ordered by generation. In addition, the mass of these particles is shown. The u and d are the lightest quarks and form protons (2 u and 1 d) and neutrons (1 u and 2 d). In 1947 new, by that time seemingly strange particles were discovered in cosmic rays in bubble chamber experiments [8], hence the name strange quark. The c and b quarks were discovered in 1974 and 1977 respectively [8]. Due to its very large mass, the top quark was only discovered in 1995 at the Tevatron by the CDF and DØ collaborations [3,4]. The top quark is also the topic of this thesis which presents the measurement of top quark pair differential cross sections.

In 1928, Dirac tried to describe the behaviour of fermions, taking the theory of rela-

tivity into account. The equation (Dirac equation [9]) he found however has solutions with positive and also with negative energy. These negative energy solutions were then interpreted as antiparticles, which have to exist for all fermions. The six quarks and six leptons thus also have corresponding antiparticles. All properties of the antiparticles are the same as for the particles, but all charges get inverted. Antiparticles are written either with a bar (e.g. $\bar{u}$) or marked using their charge (e.g. μ^+).

The charge of all observed stable particles is an integer multiple of the elementary electron charge e. Quarks have fractional charges: the down-type quarks have a charge of $-\frac{1}{3}e$ and the up-type quarks $+\frac{2}{3}e$. Thus when combining three quarks (or three anti-quarks), one gets an integer multiple of the electron charge. These particles are called *Baryons*. Well known examples are the proton with (uud) and the neutron (udd). Three u quarks can be combined to the Δ^{++} baryon and three d quarks to the Δ^- baryon. While the protons have spin 1/2, these Δ baryons have spin 3/2, i.e. all three u respectively d quarks have the same spin. Pauli's exclusion principle however demands that no two identical fermions can occupy the same state, so the existence of these particles can only be explained if there is another quark property in which the three quarks differ, namely the colour charge (see also section 2.3.2).

The quarks in Baryons can originate from any generation, for example the dsb quarks form the Ξ_b^-. All Baryons are fermions, their spin is 1/2 (for example protons or neutrons) or 3/2 (for example the Δs).

Particles consisting of a quark and an anti-quark are called *Mesons*. The lightest examples are the pions, they consist of $u\bar{d}$ (π^+), $\bar{u}d$ (π^-) or a combination of $u\bar{u}$ and $d\bar{d}$ quarks (π^0). The mesons have integer spin (0 or 1) and thus are bosons.

Bound states of quarks, i.e. mesons and baryons, are called *Hadrons*.

2.3 Interactions

The Standard Model describes the interactions between the particles by the exchange of gauge bosons. These bosons carry energy and momentum. During the very fast exchange, the bosons can violate the conservation of energy according to Heisenberg's uncertainty principle $\Delta E \Delta t \geq h$. As an alternative, one can assign a different mass to the bosons, i.e. the mass of the boson is off-shell in intermediate steps.

2.3.1 Electromagnetic interaction

The relevant theory behind this interaction is quantum electrodynamics (QED). Its underlying symmetry group is $U(1)_{\mathrm{EM}}$. Two opposite elementary charges, a positive and a negative, exist where same charges lead to a repulsive force while opposite charges attract each other. The gauge boson of QED is the photon, a massless particle which couples to all other particles with electric charge while itself being electrically neutral. It thus does not couple to other photons.

The electromagnetic interaction becomes weaker with larger distances. Photons can split into an electron-positron pair for a short time (see figure 2.2). When this happens in the field of an electric charge, the pair is polarised and thus effectively shields the charge, an effect called vacuum polarisation [8].

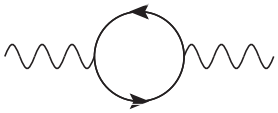


Figure 2.2: Vacuum polarisation: a photon splits into a e^+/e^- loop.

With the electron charge e and the vacuum permittivity ϵ_0, the coupling constant of the electromagnetic interaction can be calculated as:

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137} \tag{2.1}$$

Since $\alpha \ll 1$, perturbation theory can be used to calculate electromagnetic processes precisely.

Due to the mentioned vacuum polarisation, the effective charge is larger for shorter distances. This means that the coupling constant is not a real constant but actually depends on the momentum transfer Q—the value of α given in equation 2.1 is valid at zero energy and gets a little larger for higher Q^2 ($\alpha(M_Z^2) \approx 1/128.95$ [10], still $\ll 1$; $M_Z = 91.2\,\text{GeV}$ is the mass of the Z boson, see section 2.3.4). Measurements of the running of the fine structure constant can for example be found in [11] or [10].

2.3.2 Strong interaction

The strong interaction is the most powerful interaction but has only a very limited range. The theory behind it is called quantum chromodynamics (QCD) [12] which is a quantum field theory with $SU(3)_C$ symmetry. QCD describes particles with a so-called *colour charge*. There are three colours (red (r), green (g) and blue (b)) and their corresponding anti-colours, and the combination of all three colours (or anti-colours) is white or colour neutral. Its gauge boson is the gluon. Gluons are massless and electrically neutral particles carrying a colour and anti-colour charge, for example red and anti-green. In total eight different gluons exist. Gluons couple to all particles with colour charge, i.e. to quarks and also to other gluons.

Just like α, also the strong coupling constant α_S is a running constant. There are however important differences to electromagnetism: the value of α_S is larger for *lower* energy scales, not for larger scales as in the electromagnetic case. Figure 2.3 shows this dependence. One can see that α_S can attain large values so that $\alpha_S \ll 1$ is not true for small Q (or large distances) and thus perturbation theory cannot be applied in this region.

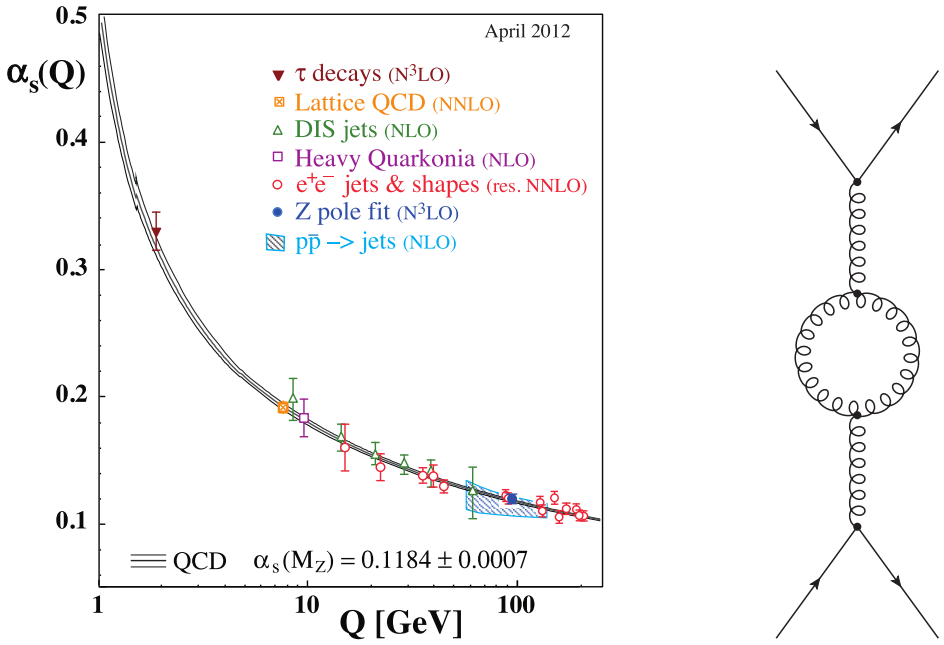


Figure 2.3: Left: Running of the strong coupling constant α_S in dependence of the energy scale Q [13]. Several measurements and their degree of QCD perturbation theory used in the extraction of α_S are shown. Right: a gluon loop.

The coupling constant α_S is formally even greater than 1 for distances greater than the proton diameter (≈ 1 fm), leading to a very strong coupling. In this case however the whole concept of perturbation theory in α_S cannot be applied any more and quoting a value for α_S is no longer useful. The strong coupling at larger distances causes a phenomenon called *confinement*, i.e. no free colour charged particles can be observed in nature. When trying to remove quarks from a proton, new quark-antiquark pairs are produced once the colour potential is large enough. These pairs can shield the colour charge similar to the vacuum polarisation in QED. In contrast to photons, gluons can however couple to other gluons and build gluon loops which even overcompensate this shielding effect, so in summary this is an antishielding effect.

At the other end of the energy scale, i.e. for very short distances or very high energies, α_S is running asymptotically to zero, so the coupling is getting weak. Therefore quarks can be treated as free particles here, this is called *asymptotic freedom*.

2.3.3 Weak interaction

The name *weak interaction* originates from the fact that its typical field strength at low energies is many orders of magnitude smaller than the strong or electromagnetic interaction. It is important nonetheless, and one of its maybe most widely known effects is the radioactive β decay.

The theory of electroweak interaction is based on a $SU(2) \times U(1)$ symmetry [12]. In addition to the photon mediating electromagnetic interactions, there are three further gauge bosons of the weak interaction. Unlike the massless photon, they are massive: the charged W^+ and W^- bosons have a mass of 80.4 GeV and the Z^0 is slightly heavier with 91.2 GeV [12]. The gauge bosons couple to all fermions. To describe the coupling, a new quantum number is introduced for the fermions, the weak isospin T; its third component is called T_3 in the following. The up-type quarks and neutrinos have $T_3 = +1/2$ while down-type quarks and e^-, μ^-, and τ^- have $T_3 = -1/2$. Weak interaction conserves the weak isospin T_3 component in its processes.

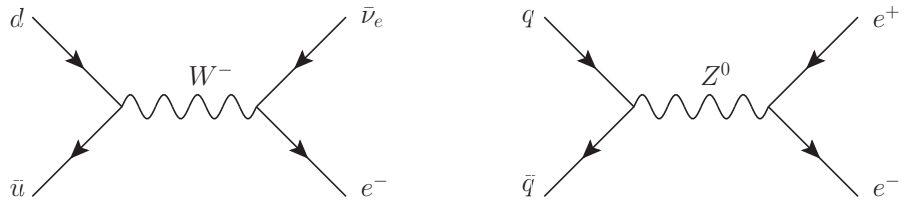


Figure 2.4: Examples of weak processes. A charged current process on the left: $\bar{u}$ and d quark transform into an e^- and a $\bar{\nu}_e$ via a W^- boson; a neutral current process on the right: a quark q and its anti-quark $\bar{q}$ transform into an electron-positron pair.

There are two types of processes: those with a charged W boson and those with a neutral Z boson; examples are given in figure 2.4. It is important to note here that charged current processes change the quark or lepton flavour—and that this is the only way to achieve that. Neither neutral currents nor the processes of the other interactions can change flavours. In addition weak processes are neither conserving parity, i.e. the behaviour of the wave function under the transformation of $(x, y, z) \mapsto (-x, -y, -z)$, nor the combined charge and parity (CP) symmetry.

Changes in the quark flavour always happen within the same generation of the quarks according to the weak eigenstates. However, the weak eigenstates of the quarks are different from their mass eigenstates. One can transform them using the Cabbibo-Kobayashi-Maskawa (CKM) mixing matrix V_{CKM} [1, 12]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.2}$$

$$\text{where} \quad V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.974 & 0.225 & 0.003 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix} \tag{2.3}$$

The CKM matrix is unitary and has only small off-diagonal elements. This means that a W boson only predominantly changes the quark flavour within the same generation

but the off-diagonal elements also allow the flavour change to a different generation. The matrix element V_{tb} is almost equal to one, thus the top quarks decay almost exclusively into a W boson and a bottom quark. All top quark analyses, including this thesis, make use of this specific feature.

2.3.4 Electroweak unification

The electroweak unification describes the combination of the electromagnetic and the weak interaction. The idea behind this unification is that weak and electromagnetic processes can be similar. In the neutral current process shown on the right side of figure 2.4 all incoming and outgoing particles are charged and the exchange boson is the electrically neutral Z^0. In this case, one could also replace the Z^0 with a photon and make this an electromagnetic process (of course, this does not work when the Z goes into neutrinos). Thus for calculations from the initial to the final state, always both processes including their interference have to be taken into account. In later chapters, this is referred to as Z/γ.

Using the weak isospin, fermions can be classified into electroweak multiplets: a left handed doublet, e.g. $\nu_{e,L}$ and e_L^- with $T = 1/2$ and $T_3 = \pm 1/2$ and a right handed singlet e_R^- where the W bosons only couple to the left handed fermions (and their right handed antiparticles). The $W^\pm$ bosons itself now must have $T_3 = \pm 1$ to be able to transform from $T_3 = -1/2$ to $T_3 = +1/2$ and vice versa. Thus they belong at least to a triplet with $T = 1$. In case of a triplet, there should also be a third W boson without charge, a W^0 which cannot change flavour due to its $T_3 = 0$ component. In addition to the W triplet, a singlet with $T = T_3 = 0$ is now introduced with a particle called B^0. The idea behind the unification was introduced by Glashow, Salam and Weinberg [14, 15], who were awarded the Nobel Prize in 1979 for this mechanism [16].

According to the electroweak unification, the photon and the Z^0 are mixed states of the W^0 and the B^0:

$$\begin{pmatrix} |\gamma\rangle \\ |Z^0\rangle \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} |B^0\rangle \\ |W^0\rangle \end{pmatrix} \tag{2.4}$$

The angle θ_W is the *weak mixing angle* or *Weinberg angle*, its value is $\sin^2\theta_W = 0.231$ [12], corresponding to $\theta_W \approx 30°$. A non-zero θ_W also leads to different masses of the W and Z bosons

$$\cos\theta_W = \frac{m_W}{m_Z} \tag{2.5}$$

The combined electroweak unification theory results in a $SU(2)_L \otimes U(1)_Y$ gauge group of the weak isospin and the weak hypercharge $Y = 2(Q - T_3)$ (Q is the electrical charge in units of the elementary charge).

2.3.5 Gravity

Gravity is more than 30 orders of magnitude weaker than the other interactions and hence only plays a negligible role on short distances. It is however relevant on longer distances and for huge masses of large objects, the most prominent example are probably stars and planets. However the masses don't have to be that large, using the Cavendish experiment [17] one can already measure the gravitation constant from smaller masses of the order of kilograms. Gravity is not included in the Standard Model and neglected throughout this thesis.

2.4 Top quark physics

2.4.1 Production

The Tevatron where the top quark was discovered [3,4] was a proton-antiproton collider with a centre-of-mass energy of 1.96 TeV which operated until September 2011. With an integrated luminosity of about $10\,\mathrm{fb}^{-1}$ per experiment[1] measurements of the top quark production cross section and its mass were possible. Even first differential cross section measurements for the transverse momentum of the top quarks [19] and the mass of the top quark pair [20] have been performed.

In figure 2.5 the cross section for top quark production is shown as a function of the centre-of-mass energy of the experiment. The cross section for this process at the LHC is two to three orders of magnitude larger than the one at the Tevatron. Therefore one can call the LHC a top quark factory. Already in the first $1\,\mathrm{fb}^{-1}$ of data, 163 thousand top quark pairs have been produced, assuming a calculated cross section of 163 pb [22] (approximate NNLO, see below). Thus the LHC can not only cross check Tevatron's results but allows for precision measurements of top quark properties.

The production mechanism for $\mathrm{t\bar{t}}$ production depends on the centre-of-mass energy of the collider experiment. At the Tevatron the annihilation of a quark and an anti-quark (figure 2.6(a)) was the dominant process with a ratio of about 85 % while gluon-gluon fusion (figures 2.6(b), 2.6(c), 2.6(d)) contributed about 15 %. At the LHC design energy of 14 TeV the situation is almost reversed with 90 % gluon-gluon fusion and only 10 % quark anti-quark annihilation. At 7 TeV gluon-gluon fusion is still dominant, the ratio is roughly 80 % to 20 % [23]. Not only the energy is responsible for these differences (see section 2.5) but also the fact that the Tevatron collides protons with anti-protons so that the annihilating quarks are both valence quarks while at the LHC a valence quark from one proton must annihilate with a sea quark from the other proton. In figure 2.6 the different leading order (LO) pair production Feynman diagrams are shown.

Leading order QCD diagrams show the simplest possible processes with the lowest possible power of α_S in the calculation of the matrix element. When more powers of

[1] latest DØ publications like [18] use $10.4\,\mathrm{pb}^{-1}$

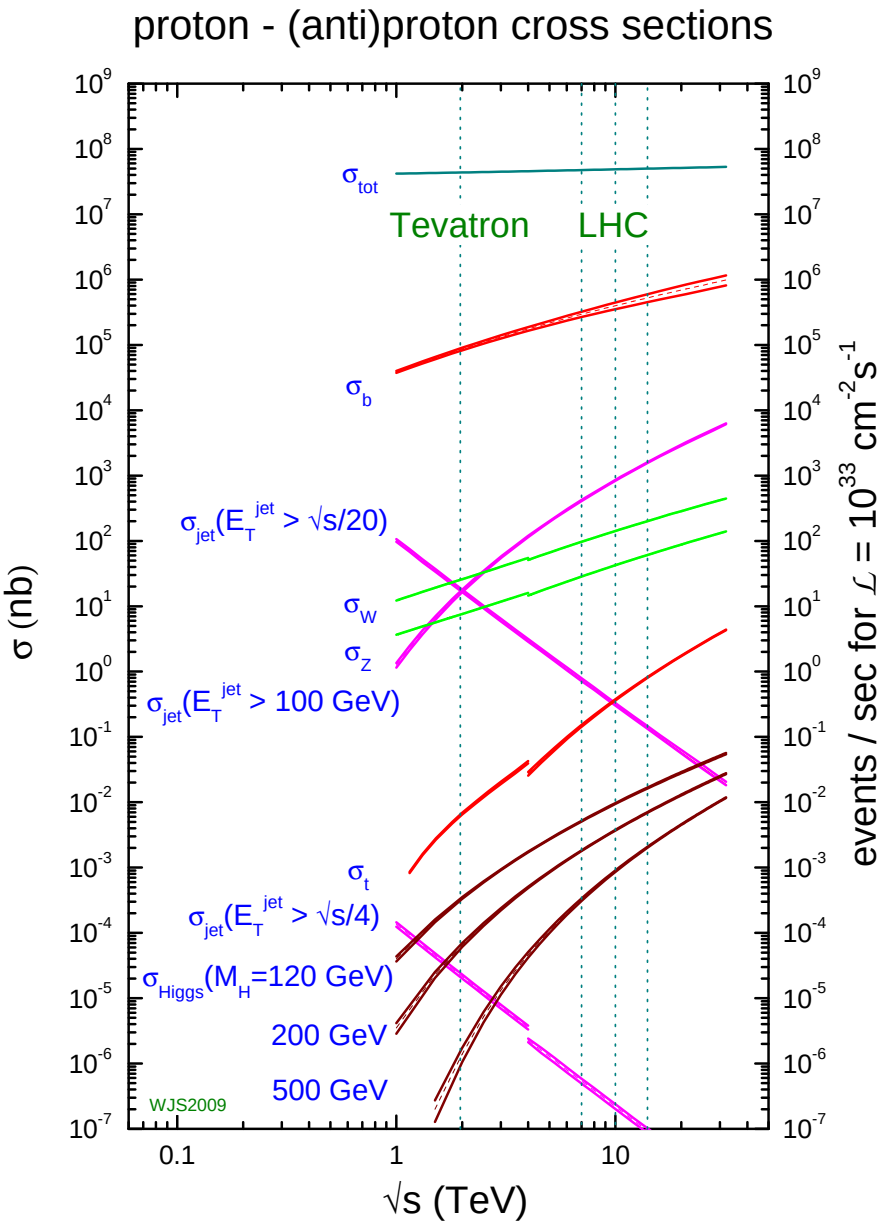
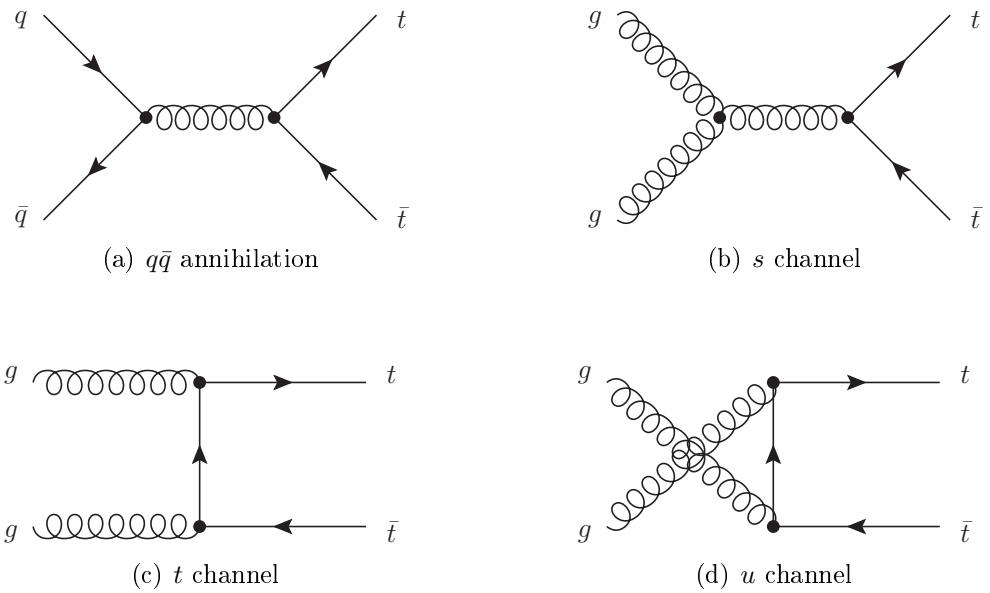


Figure 2.5: Standard Model cross sections (leading order) of different processes depending on the energy [21]. The lower red line shows the top production cross section. The four dashed vertical lines mark from left to right the centre-of-mass energies of the Tevatron (1.96 TeV), the LHC in 2011 (7 TeV), 10 TeV, and LHC design (14 TeV). The step at $\sqrt{s} = 4$ TeV is caused by different calculations for $p\bar{p}$ (< 4 TeV) and pp events (> 4 TeV).



(a) $q\bar{q}$ annihilation (b) s channel

(c) t channel (d) u channel

Figure 2.6: Leading order Feynman diagrams for $t\bar{t}$ production. Quark-antiquark annihilation is shown in the top left plot. Other diagrams show gluon-gluon fusion: s channel (top right), t channel (bottom left) and u channel (bottom right). The time axis is from left to right.

α_S are allowed, more diagrams can be drawn, for example with additional gluons or loops. These are however suppressed with respect to the leading order because of the higher power of α_S. Next-to-leading order (NLO) diagrams and/or calculations allow one more order of α_S, all corresponding Feynman diagrams and the cross section calculation have already been published in 1989 [24]. In next-to-next-to-leading order (NNLO) yet another order of α_S is taken into account. This leads to many possible diagrams and difficult calculations. A result is called *approximate NNLO* if approximations have been used in the calculation.

Two examples each for NLO and NNLO diagrams are shown in figure 2.7. The left diagrams show the radiation of additional gluons, each increasing the order of α_S by one. The plots on the right side are different, they show virtual corrections. Here the additional gluon leads to two additional vertices. However, these diagrams are still treated as NLO/NNLO because of their interference term with the LO diagram (figure 2.6(b)).

Top quarks are not necessarily produced in pairs. Figure 2.8 shows the LO Feynman diagrams for the production of single top quarks. One can already see that in all these diagrams a W boson is present. The single tops are thus not produced via the strong interaction but using a weak process. Therefore the cross section of these diagrams is much lower than for top pair production: the dominant channel here is the t-channel with a cross section of 41.7 pb (22.5 pb for $\bar{t}$ production) [25] while the s-channel only

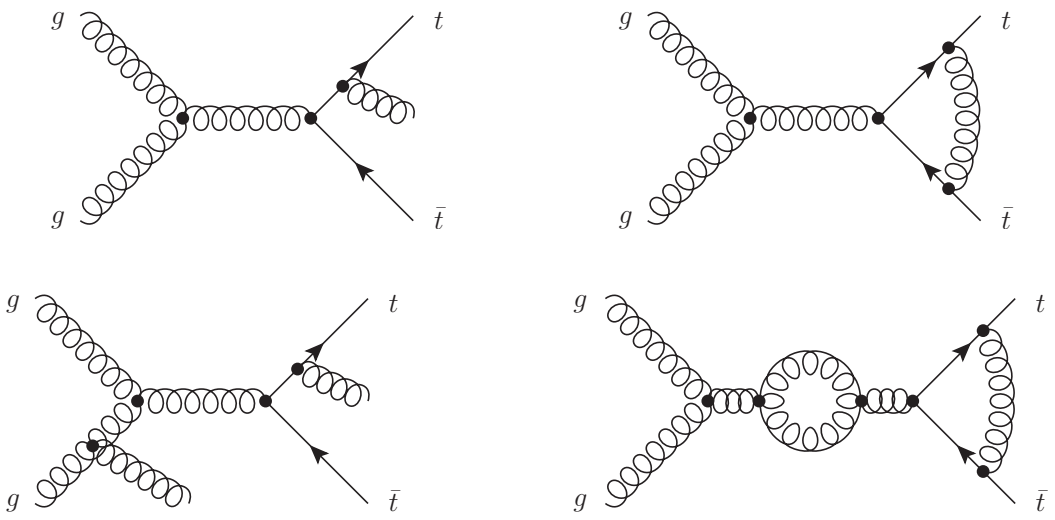


Figure 2.7: Example diagrams for the top quark pair production at NLO (top) and NNLO (bottom). The left diagrams show additional radiation of gluons, the right diagrams show virtual corrections.

contributes 3.2 pb (1.4 pb for $\bar{t}$ production) [26]. The associated tW^- production is calculated to 7.8 pb (same for $\bar{t}W^+$) [27]. All values given here assume $m_t = 173$ GeV.

2.4.2 Decay

As written in in section 2.3.3, almost all top quarks decay into a W boson and a b quark due to the large V_{tb} CKM element. For the decay width, it was found that $\Gamma(Wb)/\Gamma(Wq) = 0.99^{+0.09}_{-0.08}$ where q can be any of d, s, or b [12]. Top quark pair decays are classified into three categories based on the decay modes of the W bosons. These can decay into a quark and an antiquark or into a lepton and a neutrino. To first order, one can assume that the phase space is the same for all decay channels so that all of them have equal weight. In total, there are 9 possible decays, three into an electron, muon, or tau lepton and the corresponding neutrino, and 6 into quark pairs: the two possible quark pairs ($\bar{u}$ and d or $\bar{c}$ and s for a W^-) come in three different colour variants. Thus in $6/9 \times 6/9 = 36/81 = 4/9$ of all cases both Ws decay into quarks (full hadronic channel), in $6/9 \times 3/9 + 3/9 \times 6/9 = 36/81 = 4/9$ one W decays into a quark pair and the other into a lepton and neutrino (ℓ+jets channel) and in only $3/9 \times 3/9 = 9/81 = 1/9$ both Ws decay into a lepton and neutrino (dilepton channel). Figure 2.9 shows this.

Although the full hadronic channel has large statistics due to the high branching ration, it is the most difficult to analyse because it has to face a large QCD multi-jet background. Nevertheless the $\mathrm{t\bar{t}}$ cross section could be measured in this channel, although with a rather large uncertainty. The result is [28]:

$$\sigma_{\mathrm{t\bar{t}}} = 136 \pm 20\,(\mathrm{stat.}) \pm 40\,(\mathrm{syst.}) \pm 8(\mathrm{lumi.})\ \mathrm{pb} \tag{2.6}$$

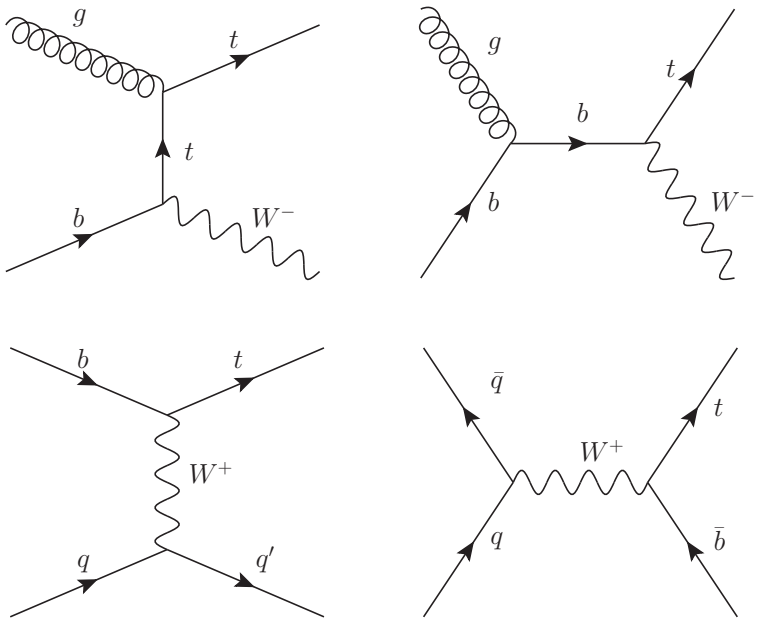


Figure 2.8: Single top production Feynman diagrams on tree level. The two upper diagrams show the tW channel, the lower left diagram shows the t channel and the lower right the s channel.

Both in the ℓ+jets and in the dilepton channel the background can be reduced much better due to the additional leptons in the event. Not only an inclusive cross section, but also differential cross sections can be measured in these channels.

The advantages of the ℓ+jets channel are the large branching fraction and the still good distinction from background. However the events contain at least 4 jets and the reconstruction algorithm needs to assign these correctly. The dilepton channel on the other hand has an even larger signal fraction after the selection and only at least two jets but its branching fraction is smaller and the event kinematics are underconstrained due to two neutrinos. This work focuses on the dilepton channels only, the results for ℓ+jets can be found in [29].

Since the τ leptons decay further, either into a muon or electron and two neutrinos, or into a neutrino and one or three hadrons, one has to decide how to assign these. In this work the events are assigned to a channel based on the final state of the decay, i.e. events where the τ decays into a μ are treated like direct decays into muons. In the end, this leads to three different dilepton channels: the ee channel with two electrons, the $\mu\mu$ channel with two muons, and the $e\mu$ channel with one electron and one muon in the final state.

The real branching fractions for the all different top decays deviate slightly from the naïve value 1/9, and in addition we have to take the decays via τ into account. The exact branching ratios of the relevant processes are listed in table 2.1. The these values

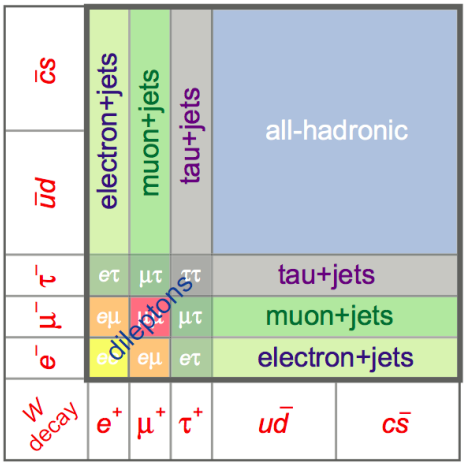


Figure 2.9: Classification of top quark pair decays. [Original source unknown.]

W decay	BR (%)
$W \to e\nu$	10.75
$W \to \mu\nu$	10.57
$W \to \tau\nu$	11.25

τ decay	BR (%)
$\tau \to e\nu\bar{\nu}$	17.85
$\tau \to \mu\nu\bar{\nu}$	17.36

Combined	BR (%)
$W \to \tau \to e$	2.01
$W \to \tau \to \mu$	1.95

Table 2.1: Relevant branching ratios for the decays of the top quark pairs. The left table shows the branching fraction of the W and the middle table of the τ decay [12]. The right table can be calculated from these values.

are then used to calculate the branching ratios of the dilepton channels.

$$\mathrm{BR}^{t\bar{t}\to ee} = (\mathrm{BR}^{W\to e} + \mathrm{BR}^{W\to\tau} \times \mathrm{BR}^{\tau\to e})^2 = 1.63\ \% \tag{2.7}$$

$$\mathrm{BR}^{t\bar{t}\to \mu\mu} = (\mathrm{BR}^{W\to \mu} + \mathrm{BR}^{W\to\tau} \times \mathrm{BR}^{\tau\to \mu})^2 = 1.57\ \% \tag{2.8}$$

$$\begin{aligned}\mathrm{BR}^{t\bar{t}\to e\mu} =& 2 \times (\mathrm{BR}^{W\to e} + \mathrm{BR}^{W\to\tau} \times \mathrm{BR}^{\tau\to e}) \\ & \times (\mathrm{BR}^{W\to \mu} + \mathrm{BR}^{W\to\tau} \times \mathrm{BR}^{\tau\to \mu}) = 3.20\ \%\end{aligned} \tag{2.9}$$

2.4.3 Top quark properties

After its discovery, properties of the top quark have been studied. Maybe the most important finding is its exceptionally large mass of $m_t = 173.5 \pm 0.6\,(\mathrm{stat.}) \pm 0.8\,(\mathrm{syst.})$ GeV [13] which is even larger than the mass of a complete gold atom. The total decay width

is [30]

$$\Gamma_t = \frac{G_F m_t^3}{8\pi\sqrt{2}} \left(1 - \frac{M_W^2}{m_t^2}\right)^2 \left(1 + 2\frac{M_W^2}{m_t^2}\right) \left[1 - \frac{2\alpha_S}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right] \tag{2.10}$$

$$\approx 1.36 \text{ GeV} \quad \text{for } \alpha_S = 0.118 \tag{2.11}$$

Top quarks have a very short lifetime of 0.5×10^{-24} s [31] which is below the timescale for hadronisation. Thus top quarks cannot form bound states like "toponium" before they decay into a W boson and a down-type quark. Due to the large mass, top quarks are the only known particles which decay into an on-shell W boson. The second decay product is a down-type quark which is almost always a bottom quark due to the large CKM matrix element V_{tb} (equation 2.3), the branching ratios to the other down-type quarks are 0.1 % for s and 0.01 % for d quarks.

2.5 Parton Distribution Functions

When two protons are colliding, it is important to understand the structure of the protons themselves. At first glance, a proton is formed by two u and a d *valence quark*, it is colour neutral and has +1 elementary charge. However due to Heisenberg's uncertainty principle, quarks can radiate a gluon and later absorb it. The gluon can also split into a $q\bar{q}$ pair which can then annihilate back into a gluon and be absorbed again. These additional quark pairs in the proton are called *sea quarks*.

The way how a proton appears in inelastic scattering events depends on the momentum transfer Q of the hard process. Parton distribution functions describe the probability to find a certain quark or a gluon with a longitudinal momentum fraction x in the proton. Research at previous collider experiments has lead to a good knowledge of the PDFs. Figure 2.10 shows the results for two fixed scales Q^2 of 10 (left) and 10^4 GeV^2 (right).

One can see that the valance quarks dominate at high x while at low x the gluon content, scaled down by a factor of 10 in the plots, becomes the dominant part. Especially the probability density for heavier quarks like b or $\bar{b}$ are very low for small Q^2 (and even zero in the left plot because $Q^2 = 10$ GeV^2 is below the mass threshold for b quarks), while they are not negligible for larger Q^2.

Several working groups exist to calculate the parton distribution functions, among them MSTW [32], CTEQ [33], and HERAPDF [34].

2.6 Top and Higgs

2.6.1 The Higgs mechanism

There is an apparent problem in the electroweak unification (section 2.3.4), namely that the gauge symmetry requires the bosons to be massless to keep local gauge invariance.

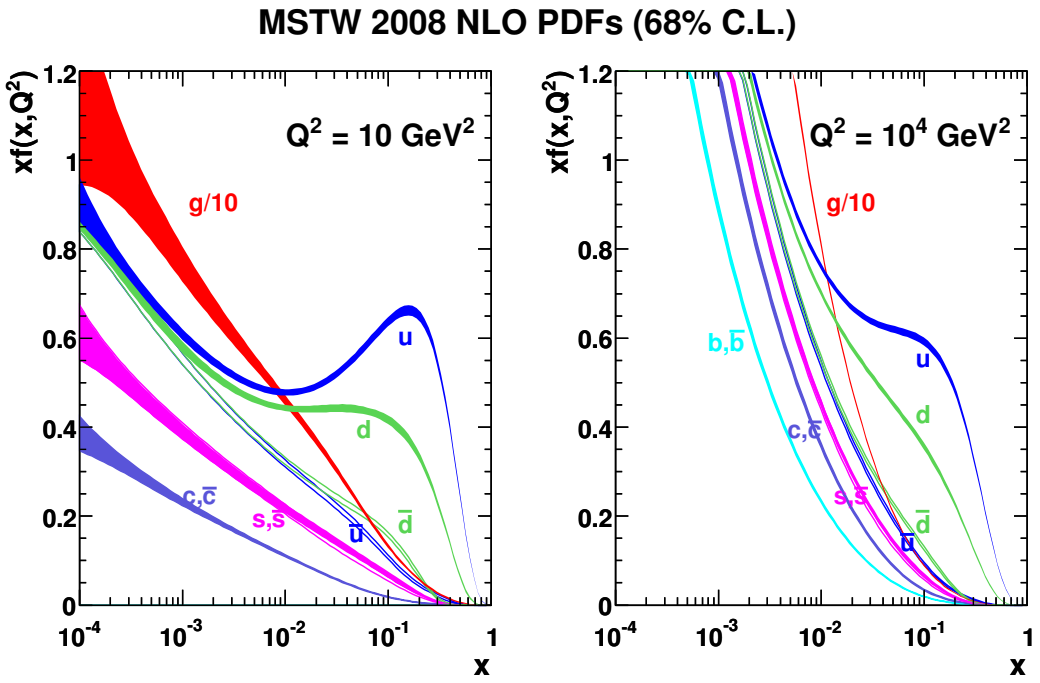


Figure 2.10: Parton Distribution Functions from MSTW [32] at different scales $Q^2 = 10$ GeV^2 (left) and $Q^2 = 10^4$ GeV^2 (right). The uncertainty band represents 68 % confidence level.

As this is obviously not the case, the gauge invariance must be broken spontaneously. This is done by postulating an additional field, the Higgs field ϕ. It interacts with all other fields and also with itself and thereby gives mass to the particles. This mechanism was introduced by Peter Higgs and others [35, 36]. The minimum choice for the Higgs field is a complex scalar isospin doublet [37]

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.12}$$

with the following symmetric potential:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{2.13}$$

The parameter λ describes the self-coupling and has to be a positive real number to ensure that the potential gets large for $\phi \to \pm\infty$, while the parameter μ is a complex number. If $\mu^2 < 0$, the vacuum expectation value v of the Higgs field is non-zero:

$$v = \sqrt{\frac{-\mu^2}{\lambda}} \tag{2.14}$$

Looking at the Higgs potential on a complex plane, it has the form of a champagne bottle, like a rotated w. By choosing a ground state the symmetry is spontaneously broken. It has to be chosen such that that photon is massless which can be achieved by setting $\phi_{1,2,4} = 0$. Remaining degrees of freedom describe the longitudinal polarisation of the Z and Ws. Around the ground state another scalar field $h(x)$ is introduced which finally leads to a new particle, the Higgs boson:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.15}$$

The vacuum expectation value can be calculated via the following formula (using $\hbar = 1$ and $c = 1$)

$$v = \frac{1}{\sqrt{\sqrt{2} G_F}} \approx 246 \text{ GeV} \tag{2.16}$$

using the Fermi coupling constant $G_F = 1.16637 \times 10^{-5}$ GeV^{-2}. This value can be used to calculate the Higgs mass m_H [37]:

$$m_H = \sqrt{\lambda/2} v \tag{2.17}$$

Even though the scale λ cannot be calculated, constraints can be put on the mass because too large masses lead to divergences in the self-coupling at a scale Λ below the Planck scale while too small masses lead to a global minimum of the Higgs potential at large values of the scalar field of the order Λ.

2.6.2 The importance of top quarks for the Higgs

Being the heaviest known fundamental fermion, the top quark is of particular interest for searches for the Higgs boson because the Higgs coupling to fermions is proportional to their mass m_f:

$$g_{Hf\bar{f}} = \frac{m_f}{v} \tag{2.18}$$

The Higgs coupling to vector bosons $V = W^{\pm}/Z$ is proportional to the squared mass, as are the couplings to other Higgs bosons [37]:

$$g_{HVV} = \frac{2m_V^2}{v} \qquad g_{HHVV} = \frac{2m_V^2}{v^2} \tag{2.19}$$

$$g_{HHH} = \frac{3m_H^2}{v} \qquad g_{HHHH} = \frac{3m_H^2}{v^2} \tag{2.20}$$

As the Higgs does not couple directly to massless photons or gluons, $H\gamma\gamma$ is mainly generated via a loop with a virtual $W^{\pm}$ pair while Hgg is is realised via a top quark

loop due to the strong Higgs to top quark Yukawa coupling. The Feynman diagram of the latter process is shown in figure 2.11. It is, among the associated production of a vector boson $W^{\pm}H/ZH$ and associated $t\bar{t}H$ production, one of the main processes for Higgs production at the LHC. Thus good knowledge about the top quark is crucial for the calculation of the Higgs production cross section.

Figure 2.11: Production of a Higgs boson from gluons via a top quark loop (left) and an example for associated $t\bar{t}H$ production (right).

Higgs-top processes have already been studied even before the discovery of the top quark [38]. Today, the calculated NLO cross section for $ttH + X$ is known, the result is $\sigma_{\mathrm{NLO}}(m_H = 125\ \mathrm{GeV}, \sqrt{s} = 14\ \mathrm{TeV}) \approx 0.6$ pb [39].

2.6.3 Higgs discovery

On July 4th, 2012 in the CERN Higgs Seminar the discovery of a new boson with a mass of about 125 GeV was presented which is consistent with the Higgs boson [40]. In the CMS collaboration the new particle was found to have a mass of 125.3 ± 0.6 GeV at 4.9σ significance [41]. The ATLAS collaboration has found the particle at a mass of approximately 126.5 GeV at a significance of 5.0σ [42]. In both experiments in total about 10 fb^{-1} of data recorded in 2011 ($\sqrt{s} = 7$ TeV) and 2012 ($\sqrt{s} = 8$ TeV) were used and many decay channels were exploited.

The quoted results are however preliminary and more data is needed to study the properties of this particle better. More decay channels and more topologies can to be studied, and it of course also has to be made sure that the new particle is really the Standard Model Higgs boson.

First publications on the measurement have been submitted to Phys. Lett. B by the ATLAS [43] and CMS [44] collaborations.

Chapter 3

The CMS experiment at the LHC

The CMS detector is one of the two general-purpose detectors at the Large Hadron Collider (LHC). Therefore this chapter is divided into two parts: the first part explains general LHC features while the CMS detector is explained in the second part beginning with section 3.2.

3.1 The Large Hadron Collider

The LHC [45] is a two-ring superconducting proton-proton collider operated by CERN, the *European Organization for Nuclear Research* (abbreviation from the former name *Conseil Européen pour la Recherche Nucléaire*). The LHC is reusing the tunnel of the LEP experiment, a tunnel with a circumference of 26.7 km located near Geneva. It is built into molasse rock (90 %) and limestone (10 %) between 70 m and 140 m under the surface of Switzerland and France.

LHC is designed to work at a centre-of-mass energy of up to $\sqrt{s} = 14$ TeV and a peak instantaneous luminosity of $\mathcal{L} = 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. For a safe operation and to avoid magnet quenches (like the one which happened in September 2008 [46]), the LHC has operated with $\sqrt{s} = 7$ TeV until the end of 2011. In 2012, the energy could be raised to 8 TeV. A longer maintenance shutdown is needed in 2013 before the energy can finally be raised to 14 TeV.

The instantaneous luminosity $\mathcal{L}$ describes the rate dN/dt of a certain process per its cross section σ:

$$\mathcal{L} = \frac{dN}{dt} \cdot \sigma^{-1} \tag{3.1}$$

It needs to be distinguished from the integrated luminosity L which denotes the total amount of recorded events of a certain process per cross section.

$$L = \int \mathcal{L}\, dt = N \cdot \sigma^{-1} \tag{3.2}$$

Thus the number of produced events can be calculated as:

$$N = L\sigma \tag{3.3}$$

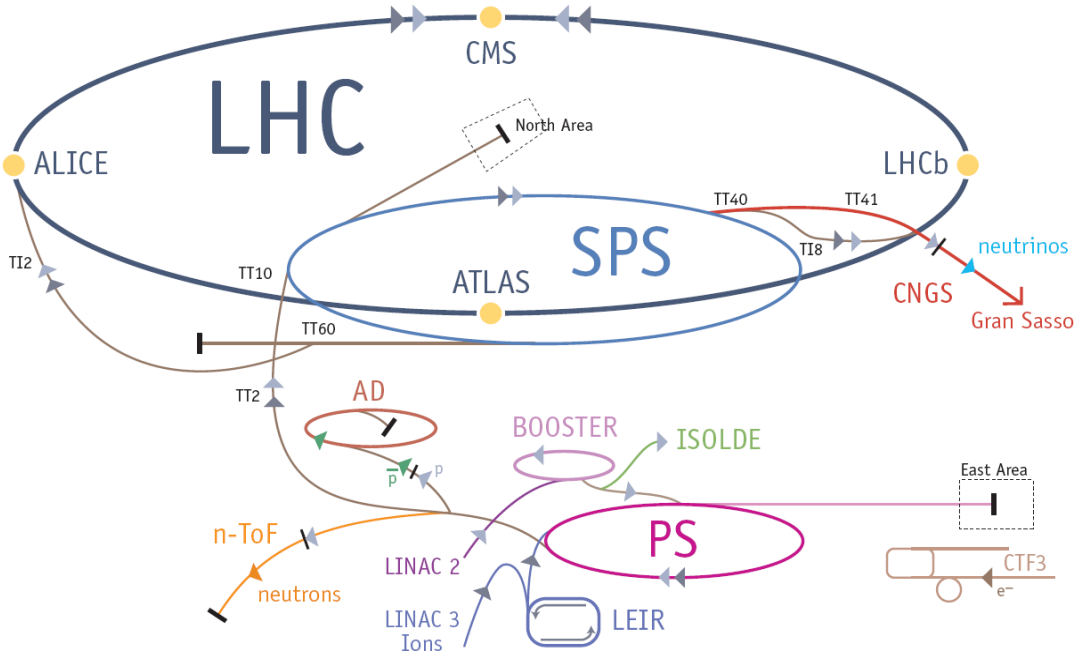


Figure 3.1: Accelerators at CERN [49]. The way of protons (light grey arrows) and heavy ions (dark grey arrows) into the LHC is shown.

While the cross section is specific to a certain event type, the luminosity is a parameter of the LHC. The protons in the LHC are organised in n_b bunches, each of which contain N_b protons. The luminosity can then be written as [45]:

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \tag{3.4}$$

In this equation, $f_{rev} = c/r_{\mathrm{LHC}} = 11.25\,\mathrm{kHz}$ is the revolution frequency, $\gamma_r = 1/\sqrt{1 - v^2/c^2}$ the relativistic gamma factor, ϵ_n the normalised transverse beam emittance, β^* the focus of the beam and F the geometric luminosity reduction factor due to the crossing angle at the interaction point.

In the design, the maximum number of protons per bunch is 1.15×10^{11} in a maximum of 2808 bunches. This corresponds to a total bunch crossing rate of 40 MHz or one bunch crossing in 25 ns. In 2011, the maximum number of bunches was 1374, resulting in an instantaneous luminosity of $3.65 \times 10^{33}\,\mathrm{cm^{-2}s^{-1}}$ with a bunch spacing of 50 ns [47]. Until summer 2012, about 65 % of the design instantaneous luminosity were reached, the number of protons per bunch is 1.5×10^{11} even higher than the design value [48].

Figure 3.1 shows the setup of the accelerators at CERN. The protons and heavy ions first need to be pre-accelerated before they can be injected into the LHC. Both types of particles start in dedicated linear accelerators and are then run through the Proton Synchrotron PS and the Super Proton Synchrotron SPS where they are accelerated to 25 GeV and 450 GeV respectively [50, 51]. Only then they are injected into the LHC.

The figure also shows four large experiments (ALICE, ATLAS, CMS, LHCb) at the LHC in yellow, the two smaller experiments LHCf and TOTEM are not shown. All six experiments are briefly described in the following:

- **ALICE** – A Large Ion Collider Experiment

  ALICE [52] focuses on QCD processes in the Standard Model. It is designed especially for heavy ion collisions which produce a quark-gluon plasma with very high energy densities. Therefore the detector needs to handle particle multiplicities which are about three orders of magnitude higher than the ones in proton-proton collisions. To reconstruct the tracks, ALICE uses a silicon vertex detector, a time projection chamber (TPC) and a transition radiation detector. While the tracking can reliably detect $O(10000)$ charged particles within the acceptance, it is a rather slow system with about 90 µs drift time. The TPC thus limits the luminosity for proton-proton collisions.

- **ATLAS** – A Toroidal LHC ApparatuS

  With its length of 44 m and a diameter of 25 m, ATLAS [53] is the largest particle detector in the world. It is onion shaped with its tracking system close to the interaction point, a high granularity liquid argon electromagnetic sampling calorimeter, hadronic tile calorimeters and a muon spectrometer. The magnet system consists of a rather small 2 T solenoid aligned on the beam axis between the tracking system and the calorimeters as well as large barrel and end-cap toroids (0.5 to 1 T) for the muon system.

  With ATLAS being a general-purpose detector, a variety of searches for the Higgs boson and other new physics have been published alongside precision measurements of the Standard Model.

- **CMS** – Compact Muon Solenoid

  The CMS detector is the other general-purpose detector at LHC. Compared to ATLAS, it has a similar onion shape design but uses different technologies in the detector. It is described in more detail in section 3.2.

- **LHCb** – Large Hadron Collider beauty

  The LHCb detector [54] mainly focuses on CP violation and heavy flavour physics. Rare decays of B and D hadrons are studied to find evidence for new physics. The cross section for $b\bar{b}$ production is so large ($288\,\mu$b at 7 TeV [55]) that the LHCb experiment can afford lowering the luminosity to avoid pile-up by tuning the beams at the interaction point. The b and $\bar{b}$ quarks are predominantly produced either both in forward or both in backward direction, thus it is sufficient to measure only in one direction. To save costs LHCb is therefore built as a single-arm spectrometer and covers only one of the two directions along the beam axis (from 10 up to 250/300 mrad).

- **LHCf** – Large Hadron Collider forward

  LHCf [56] consists of two very forward imaging calorimeters made of tungsten plates, plastic scintillators and position sensitive sensors located at ±140 m from the ATLAS interaction point, corresponding to a pseudorapidity (see equation 3.5) range of $|\eta| > 8.4$. The goal of these detectors is mainly to calibrate hadron interaction models for high-energy cosmic rays.

- **TOTEM** – TOTal Elastic and diffractive cross section Measurement

  The total proton-proton cross-section is measured in a luminosity-independent way by TOTEM [57] using the Optical Theorem. In addition, the proton structure is studied by inspecting elastic scattering with large momentum transfers. This happens in close collaboration with CMS as TOTEM is technically integrated in CMS, covering a pseudorapidity of $3.1 \leq |\eta| \leq 6.5$. It consists of two telescopes to measure charged particles and detectors in movable beam-pipe insertions at ±147 m and ±220 m from the CMS interaction point.

3.2 The Compact Muon Solenoid detector

The Compact Muon Solenoid [58] is located about 100 m underground at one of the interaction points of the LHC. It is 21.6 m long and has a diameter of 14.6 m. With its mass of 12.5 kt, CMS is the heaviest detector at the LHC. A perspective view of CMS and its components can be found in figure 3.2. The detector is built in a cylindrical form around the beam pipe which passes through the middle of the detector, the point where the protons are colliding. CMS is symmetric in the radial direction around the beam pipe and also symmetric along the beam pipe from the centre of the detector.

3.3 Coordinate system and variables

In CMS, a coordinate system is used where the origin is in the centre of the detector, at the nominal interaction point. The x axis points to the centre of the LHC ring, the y axis points vertically upwards and the z axis points along the beam pipe toward the Jura mountains, corresponding to the anticlockwise direction of the beam pipe.

Given the symmetry of the detector, it is often useful to use spherical coordinates: the azimuthal angle ϕ is measured from the x axis in the $x - y$ plane ($0 \leq \phi \leq 2\pi$) and the polar angle θ is measured from the $+z$ axis. The distance from the beam pipe is named r.

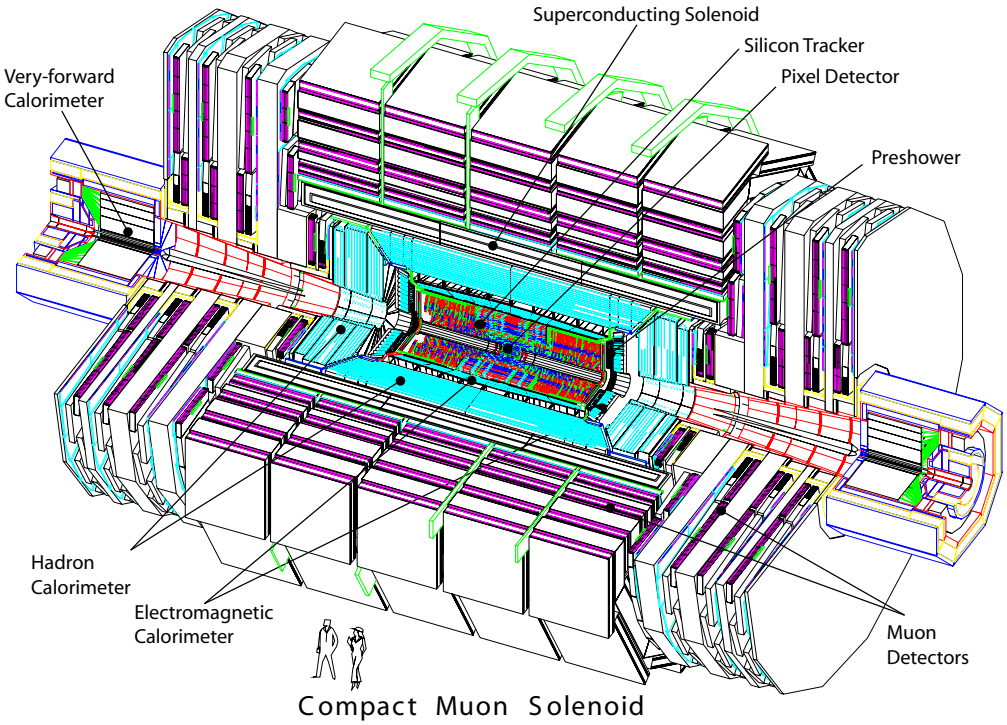


Figure 3.2: Perspective view of the CMS detector [59]. The different components of the detector are explained in section 3.4.

Instead of the polar angle θ, often the pseudorapidity η is given:

$$\eta = -\ln \tan \frac{\theta}{2} \tag{3.5}$$

$$= \frac{1}{2} \ln \frac{|\boldsymbol{p}| + p_L}{|\boldsymbol{p}| - p_L} \tag{3.6}$$

It can not only be expressed as a function of the polar angle (equation 3.5) but also using the momentum $\boldsymbol{p}$ of a particle and its longitudinal component p_L (equation 3.6) where longitudinal is defined as a projection on the z axis. For massive particles like the top quark, the rapidity y is used instead:

$$y = \frac{1}{2} \ln \frac{E + p_L}{E - p_L} \tag{3.7}$$

The difference to the pseudorapidity is that the momentum $\boldsymbol{p}$ is replaced with the energy E, so $\eta = y$ for massless or light particles like leptons. The advantages of the rapidity over the angle θ are that differences in rapidity are invariant under Lorentz boosts in the z direction. Also the particle flux is roughly constant per rapidity unit.

Other useful variables are transverse momentum $p_T = \sqrt{p_x^2 + p_y^2}$ and the transverse energy $E_T = \sqrt{E_x^2 + E_y^2}$ of a particle where the transverse part of a vector is the absolute value of its projection onto the $x - y$ plane. Because both colliding protons have no transverse momentum, the whole initial state has no transverse momentum. Due to energy and momentum conservation, this is also true for the final state after the collision. Due to neutrinos which are not detected at all and due to mismeasurements and detection inefficiencies, there can be transverse energy missing to fulfil the conservation law. This imbalance is called $E_{\mathrm{T}}^{\mathrm{miss}}$, it is defined as follows:

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_i \vec{E}_T^{\,i} \quad \text{where } i \text{ runs over all reconstructed particles} \tag{3.8}$$

$$E_{\mathrm{T}}^{\mathrm{miss}} = |\vec{E}_{\mathrm{T}}^{\mathrm{miss}}| \tag{3.9}$$

3.4 Detector components

3.4.1 Tracker

As shown in figure 3.2 there are several different types of detectors in CMS. The tracking system [58] is the innermost component and has a diameter of 2.5 m and a length in z of 5.8 m. It is used to measure tracks of charged particles in the region up to $|\eta| = 2.5$ very precisely. While the tracker has a constant efficiency of almost 100 % in the central region, this value starts falling from $|\eta| > 2.1$ because there are fewer modules in that region.

Figure 3.3 gives a sketch of the tracker. The silicon pixel detector is placed in the close vicinity of the interaction point. It consists of three cylindrical barrel layers placed at radii of 4.4, 7.3, and 10.2 cm from the beam pipe and four disks, two on each side of the barrel at ±34.5 and ±46.5 cm in z direction. In total there are 66 million pixels on 1440 modules, the pixel size is only 100 µm × 150 µm, resulting in a hit position resolution of 15-20 µm [60]. With a hit efficiency over 99.5 % per layer, typically charged particles generate three hits in the central region ($|\eta| < 2.2$) and two hits in the forward regions ($2.2 < |\eta| < 2.5$) of the pixel detector.

The pixel detector is surrounded by silicon strip trackers which contain 9.3 million strips on 15148 modules divided into four subsystems. The modules in the TIB (Tracker Inner Barrel) and TOB (Tracker Outer Barrel) are aligned in parallel to the beam pipe, the modules in the TID (Tracker Inner Disk) and TEC (Tracker End Cap) are placed perpendicular to the beam pipe. The strip spacing varies from 80 µm in the TIB up to 184 µm in the TEC. Several mainly inner modules are built as double sided modules where two modules are mounted back-to-back with an angle of 100 mrad (5.7°), allowing to measure the z coordinate in the barrel and the r coordinate in the disks and endcaps.

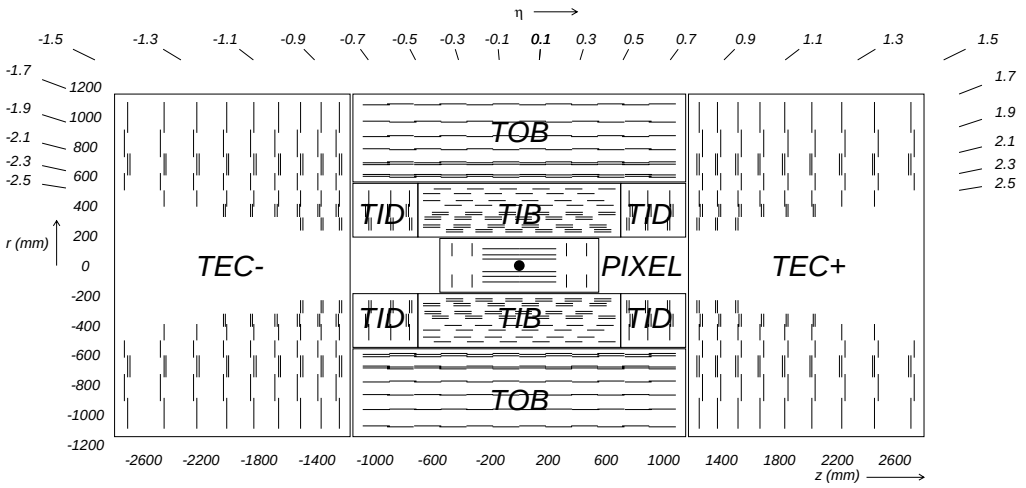


Figure 3.3: The tracking system in CMS [58]. Each single detector module is represented with a line, stereo modules with a double line.

3.4.2 Calorimeters

3.4.2.1 The electromagnetic calorimeter

After passing the tracking system particles fly into the electromagnetic calorimeter (ECAL) [58] where the energy of charged particles and photons is measured. The ECAL consists of a barrel part (EB) covering a range up to $|\eta| < 1.479$ and the two endcaps (EE) in the range $1.479 < |\eta| < 3.0$ as shown in figure 3.4. In addition, there is a preshower detector in front of the endcaps ranging from 1.653 to 2.6 in $|\eta|$.

The ECAL is designed to be fast, radiation hard and to have a fine granularity to be able to discover the Higgs bosons via its decay into two photons. Technically this is achieved with lead tungstate ($PbWO_4$) crystals. The high density of lead tungstate leads to a short radiation length[1] of $X_0 = 0.89$ cm. Also the material emits 80 % of the light within the nominal LHC collision rate of 25 ns. The emitted scintillation light has its maximum intensity at a wavelength of 420-430 mm and is detected with avalanche photodiodes in the EB and with vacuum phototriodes in the EE.

In the EB there are 61200 crystals (360 in ϕ times 170 in η), each positioned at a distance of $r = 1.29$ m from the beam pipe. They have a size of 0.0174×0.0174 in η-ϕ and a length of 230 mm (25.8 X_0), corresponding to 22×22 mm^2 at the front and 26×26 mm^2 at the back. They are tilted by 3° with respect to the direction of the nominal interaction point to avoid particles traversing the calorimeter along the cracks between the modules.

[1] The radiation length X_0 is the typical distance in which an electron has radiated so much bremsstrahlung that its energy is reduced by a factor of $1/e$.

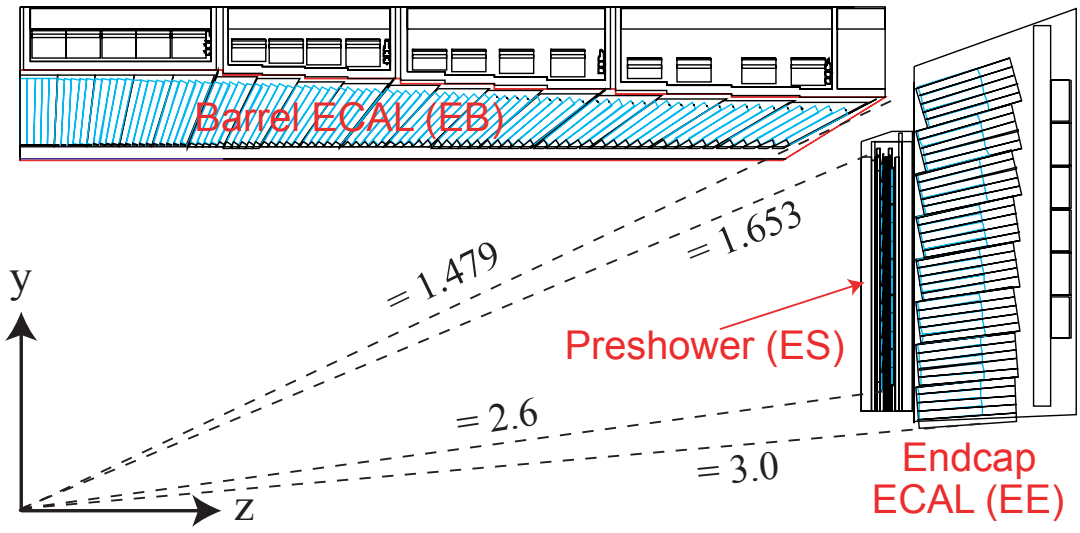


Figure 3.4: A schematic view of the $+y$ and $+z$ part of ECAL [58]. The barrel (top) and and endcap with the preshower detector (right) are shown.

The 7324 crystals per endcap are placed at $z = \pm 3.154$ m. They have a front face cross section of 28.62×28.62 mm^2 (30×30 mm^2 at the rear) and are tilted by 2° to 8° so that they point 1.3 m behind the interaction point. With a length of 220 mm (24.7 X_0) the EE crystals are a bit shorter than the ones in the EB.

The preshower is a two-layered sampling calorimeter in front of the endcaps. It consists of lead to initiate the showers and silicon strip sensors to detect them. The first layer contains material of 2 X_0, the second layer 1 X_0. The silicon strips in the two layers are orthogonal for a better resolution. The main use of the preshower is the detection of π^0 hadrons, the distinction of electrons from muons, and an improved position determination for electrons and photons.

The resolution of the ECAL has been measured in testbeams, however without magnetic field and inert material in front of the crystals. It was found to be [58, 61]:

$$\frac{\sigma(E)}{E} = \frac{2.8\%}{\sqrt{E[\,\mathrm{GeV}]}} \oplus \frac{12\%}{E[\,\mathrm{GeV}]} \oplus 0.3\% \tag{3.10}$$

There are three contributions which need to be added in quadrature (denoted with the $\oplus$): the first term is a stochastic term with contributions from event-to-event fluctuations in the lateral shower containment, photostatistics and energy deposits in the preshower; the second term corresponds to noise from the electronics, digitisation and pile-up; and the third term contains non-uniformity of the longitudinal light collection, intercalibration errors and energy leakage.

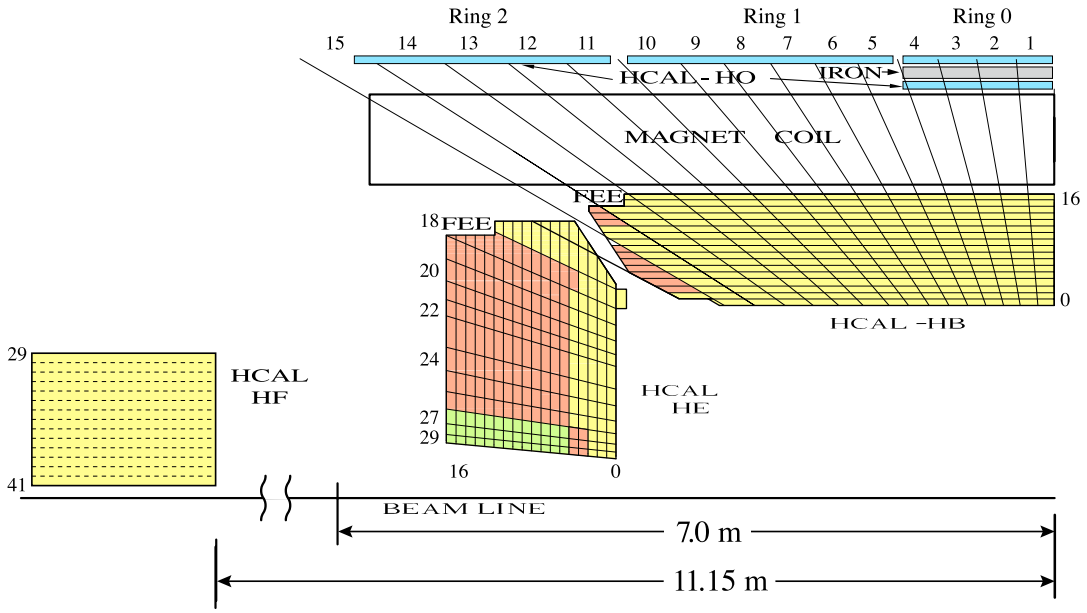


Figure 3.5: Sketch of a quarter of the CMS hadron calorimeter [62] with the interaction point on the bottom right of the image. The barrel (HCAL-HB), the endcap (HCAL-HE), the outer (HCAL-HO) and the forward (HCAL HF) calorimeters are shown and the position of the front end electronics is marked with FEE. Same colours in HB and HE towers show the longitudinal segmentation for the readout.

3.4.2.2 The hadron calorimeter

Particles which were not stopped in the electromagnetic calorimeter enter the hadron calorimeter (HCAL) [58] afterwards. The HCAL consists of four parts, as shown in figure 3.5. An important feature of the HCAL is that large parts of it, i.e. the barrel (HB) and the endcaps (HE), are still inside the solenoid while only the small outer calorimeter (HO) is placed behind behind the magnet. In addition there are forward calorimeters (HF) positioned behind the sides of the solenoid.

The HB covers the pseudorapidity range of $|\eta| < 1.3$. It is a sampling calorimeter with 14 brass absorbers (50.4 mm to 56.6 mm thickness) and two stainless steel absorbers as innermost and outermost plates (40 mm and 75 mm) for stability. The light from the plastic scintillators is converted with wavelength-shifting fibres in the scintillator modules and then transported through clear fibres to hybrid photodiodes. The HB is divided into towers covering a range of 0.087×0.087 in η-ϕ, corresponding to 5×5 crystals in the ECAL. Due to the limited space between the ECAL ($r = 1.77$ m) and the solenoid ($r = 2.95$ m), the hadronic interaction lengths λ_I of the HB ranges from only $5.82\lambda_I$ at the centre ($\eta = 0$) to $10.6\lambda_I$ on the sides ($|\eta| = 1.3$). In addition the hadrons have already passed the ECAL which provides about $1.1\lambda_I$ of material.

Because of the limited thickness, an outer barrel detector, the HO, is attached behind the solenoid. Here the solenoid itself is used as absorber material, its thickness depends on the angle and results in $1.4\lambda_I/\sin\theta$. In the central region the total depth of the HB is smallest, therefore the HO adds two layers of scintillators at $r = 3.82$ m and $r = 4.07$ m with 19.5 cm steel in between. In the non-central region the total depth of the HB is larger and thus the HO only adds one layer of scintillators at $r = 4.07$ m. In total the HO increases the minimum thickness of the calorimeter to $11.8\lambda_I$.

Hadronic particles with pseudorapidities in the range $1.3 < |\eta| < 3.0$ are detected in the HE. The HE has 18 layers with 79 mm brass plates and 9 mm scintillators in between. The granularity is in η-ϕ space is 0.087×0.087 for $1.3 < |\eta| < 1.6$ and 0.17×0.17 for $1.6 \leq |\eta| < 3$.

In the forward directions a very high particle flux is present and thus radiation hardness is crucial in this region. The HF is made of a cylindrical steel structure with a radial extension between $r = 12.5$ cm and $r = 130.0$ cm, its face is placed at $z = \pm 11.2$ m and therefore it covers $3.0 < |\eta| < 5.2$. The resolution is 0.17×0.17 in η-ϕ space. In contrast to the other parts of the HCAL, also many electromagnetic particles can reach the HF. Therefore only half of the quartz fibres which collect Čerenkov light emitted by the showers go through the whole 165 cm of the HF, the other half does not cover the first 22 cm from the face. This way a distinction between electromagnetic and hadronic showers is possible because electromagnetic showers typically deposit their energy within the first part, while hadronic showers start later.

Detailed studies have shown that the systematic uncertainty on the jet energy scale is less than 3 % in the region $|\eta| < 3$ and about 5 % in $3 < |\eta| < 5$ for jets with at least 50 GeV transverse energy [63].

3.4.3 Solenoid

One of the name-giving components of CMS is the solenoid. As already mentioned in the previous section, it is placed between the HB and HO detectors. Its dimensions are large: the magnet has a diameter of 6 m, a length of 12.5 m and a mass of 220 t [58]. It uses four winding layers of a superconducting NbTi conductor operated at a temperature of 4.6 K and a current of 19 kA to produce the designed 4 T magnetic induction in its bore. In operation, 2.6 GJ energy are stored in the magnet, corresponding to a very high energy to mass ratio of 11.6 kJ/kg and thus causing a large mechanical deformation of 0.15 % w.r.t. a zero field. The strong magnetic flux is returned through a large multi-layered iron return yoke as visible in figure 3.6. The barrel part of the return yoke is 13 m long and has an outer diameter of 14 m, it is enclosed with endcaps on both sides. With a total iron mass of 10 kt the return yoke makes up 80 % of the total mass of CMS.

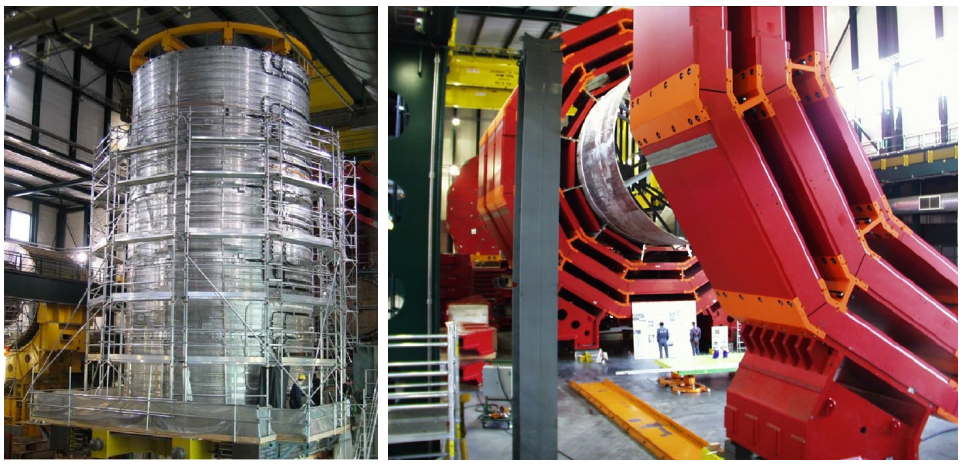


Figure 3.6: On the left side a photo of the solenoid is shown, the right side shows the barrel layers of the red iron return yoke [58].

3.4.4 Muon system

The muon system [58] is the second name-giving component of CMS. It is important to detect muons because they can not only be used in searches for the Higgs boson using the golden channel $H \rightarrow ZZ \rightarrow 4\mu$ but they are also present in many other interesting processes and thus are often an important mean to distinguish signal from background events. They are minimum ionising particles (the stopping power of different materials can be found in [64]) and thus deposit only a small fraction of energy in the calorimeters. Therefore the muon system can be placed outside of the calorimeters. In CMS, the muon system is embedded into the return yoke of the solenoid and is divided into four muon stations in the barrel and four stations in the endcaps. A sketch of its different components is shown in figure 3.7.

Three different types of gaseous particle detectors for muon identification are used: drift tube chambers (DT) in the barrel ($|\eta| < 1.2$), cathode strip chambers (CSC) in the endcaps ($0.9 < |\eta| < 2.4$) and resistive plate chambers (RPC) both in the barrel and the endcaps ($|\eta| < 1.6$).

In the barrel region, drift tubes filled with Argon (15 %) and CO_2 (85 %) are used. Here the magnetic field has the opposite direction compared to the tracker, and it is smaller (2 T). Each muon station consists of three superlayers which have four layers of rectangular drift cells each. To maximise the sensitive area, every second layer is displaced in z by half the length of a cell. There are superlayers which contain DTs with wires parallel to the beam pipe to measure r and ϕ and superlayers where the DTs have wires in the orthogonal direction to measure the z coordinate.

Because of a higher flux with hit rates up to 1 kHz/cm^2 and a stronger non-uniform magnetic field, Cathode Strip Chambers (CSC) are used in the endcaps. The CSCs are multiwire proportional chambers comprised of six anode wire planes interleaved among

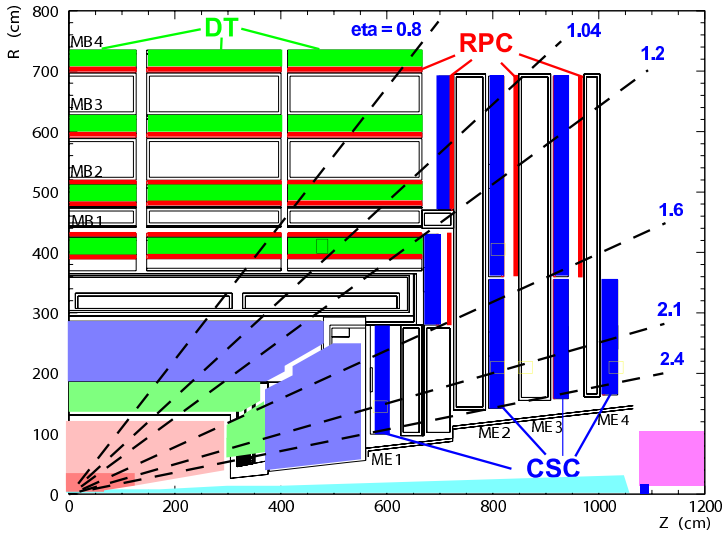


Figure 3.7: Sketch of quarter of the muon detectors [58]. Drift tubes (DT) are shown in green, resistive plate chambers (RPC) in red and cathode strip chambers in blue. MB refers to the barrel and ME to the endcap of the muon system.

seven cathode panels. Wires run azimuthally and define a track's radial coordinate [58]. The spacial resolution achieved with the CSCs ranges from 75 µm to 150 µm in r-ϕ.

The third system to measure the muons are Resistive Plate Chambers (RPC). They complete the muon system with an excellent time resolution and are able to tag muons in much less than 25 ns [58]. Therefore they can identify the bunch crossing from which a muon is originating and provide fast input for the trigger (see section 3.5). The RPCs are parallel plate detectors filled with a gas mixture of 96 % $C_2H_2F_4$, 3.5 % C_4H_{10} and 0.3 % SF_6. In total there are six layers of RPCs in the barrel, one on both sides of the DTs in the first and second muon station and one on the inner side of the DTs in the outer two muon stations. In addition, there are another three RPCs layers in the endcaps.

For the reconstruction of muons the tracking system is used in addition to the different muon detectors. The reconstruction performance has been measured in [65]: the identification efficiency for muons with a transverse momentum of more than a few GeV is greater than 95 % in all detector regions while the misidentification rate lies only between 0.1 % and 1 %, depending on the selection. For muons with 20 GeV $< p_T <$ 100 GeV the relative transverse momentum resolution is between 1.3 % and 2 % in the barrel and better than 6 % in the endcaps. Even for high-energetic muons with a p_T of 1 TeV the resolution is still better than 10 %.

3.5 Trigger

Due to the high collision rate of up to 40 MHz at the LHC and the large number of read out channels, it is not possible to record all events that are produced. Rather only a drastically reduced rate of potentially interesting events can be recorded and processed. The task of the trigger is to decide very quickly whether or not an event is worth recording.

The trigger system in CMS reduces the rate by six orders of magnitude so that roughly 100 to 150 are stored per second. This is done in two steps using a hardware-based Level-1 Trigger (L1) [66] followed by a software-based High-Level Trigger (HLT) [67,68]. The L1 only uses coarsely segmented information from the calorimeters and the muon system to reduce the output rate to a maximum of 100 kHz, corresponding to a data flow of ≈ 100 GB/s. At this stage a fast specialised jet clustering algorithm is run to reconstruct electron and photon candidates as well as jets and the missing transverse energy. All detectors of the muon system and a minimum ionising signature from the calorimeters are used for the muon reconstruction. Finally all information from the global calorimeter trigger and the global muon triggers is collected by the global trigger which finally decides to keep or discard the event.

The L1 mainly uses FPGAs[2] for flexibility; ASICs[3] and programmable lookup tables for speed. Because the L1 takes its decision in up to 3.2 µs, all detector information must be pipelined for a possible decision to accept the event. In that case all the event information is read out and passes to the HLT.

The HLT has three main tasks to perform. At first, it needs to read out the whole event from the front-end electronics of the detector components. The second step is to reconstruct the events using a specialised CMS software version on a large computer farm. The pure software implementation allows for easy changes in the HLT so that the algorithms deciding on the events to accept can be changed. This is done not only when there are better algorithms or implementations but also as the instantaneous luminosity of the LHC grows. The third task is to finally forward all the accepted events, but also a small rate of rejected events for storage and online data quality monitoring.

3.5.1 Prescaling

Because there are limitations on the rate in which events can be recorded, triggers cannot trigger all events but must have relatively high thresholds on the transverse momentum of the triggered objects, depending on the instantaneous luminosity. In the analysed dataset (see chapter 7.2) for example the p_T threshold was 24 GeV for single isolated muons. To also allow for triggers with lower thresholds, one can either use a trigger which requires another object to fire, for example a second lepton or a jet, or one can

[2]Field Programmable Gate Array
[3]Application Specific Integrated Circuit

use a *prescaled* trigger. A prescaled trigger has a low energy threshold but only records a fraction of events in which it has fired. Using a prescale factor N_{pre}, only one in N_{pre} events will get recorded. This is particularly useful for already known processes or analyses which do not need high statistics but which need to go down to low energetic objects.

3.6 Data Quality Monitoring

Before the recorded data can be used for physics analyses, it needs to be made sure that all detector components are working properly and that the reconstruction software has no obvious bugs. This is done in the Data Quality Monitoring (DQM). DQM tries to discover these problems early to ensure a good detector and operation efficiency [69]. The monitoring is organised in two steps, online and offline.

The data in the online monitoring is taken directly from the high level trigger filter units at a rate of 100 kHz for a limited number of control histograms. In addition events from a dedicated DQM data stream at a rate of $10 - 15$ Hz are processed. Finally different algorithms are run on the data and corresponding monitoring elements are produced. In these, for example bad channels or noise can be detected very fast.

The second step is the offline DQM. It monitors fully reconstructed or even re-reconstructed data, i.e. not only raw detector information but also the physics objects such as muons or jets can be monitored; also it combines longer data taking periods than the online DQM. The reconstruction is a lengthy process and usually takes one or two days for the prompt reconstruction and up to weeks for the re-reconstruction which happens in distributed computer centres around the world. The output of the offline DQM can be used to certify the data or to decide that certain periods of data taking must not be considered for analyses.

The DQM histograms can be looked at via a dedicated web service [70]. It allows CMS users to look at the data quality using a web browser. Also the DQM shifters who certify the runs are looking at the same web page. This system has proven to be very reliable, even though it handles approximately 50 thousand histograms per run. In the course of this thesis, I have participated in online DQM shifts from the DESY remote control centre and certified several runs. Currently, all online shifts are done from CERN and DESY participates in offline shifts.

Chapter 4

Computing

In this chapter, the technical work that has been done in the course of this thesis is described: the automation of the deployment of the CMS software on Grid sites.

At first, a general introduction about the CMS software is given and the local installation procedure is explained. Beginning with section 4.4, my own work to automate the installation is summarised.

4.1 The CMS Experiment Software

The CMS Software CMSSW is the essential tool needed to perform any analysis on data recorded by the CMS detector. The software includes code to record data and to reconstruct physics objects like jets or leptons. Even some dedicated analysis code is included.

The amount of data produced by the detector is huge. Analysing the data on a single computer would take much too long, therefore a large computing infrastructure is required. The CMS computing infrastructure [71] is built on top of the *Worldwide LHC Computing Grid* (WLCG), which is organised in a tiered structure [72].

4.2 CMS Experiment Software on the Grid

Grid computing describes a form of distributed computing where the worker nodes usually belong to different computing centres, i.e. where the computing centres have full control over their own nodes but where they have agreed to comply with a certain list of requirements. For the LHC, these requirements and participating sites are defined in a memorandum of understanding [72].

The CMS experiment is supported at the Tier-0 at CERN, 7 Tier-1 centres (ASGC, Taiwan; Fermilab, USA; GridKa, Germany; IN2P3, France; INFN, Italy; PIC, Spain; RAL, United Kingdom), about 50 Tier-2 sites in more than 20 different countries and several Tier-3 sites.

Different requirements, for example response times in case of failures, are defined for the different tiers, with highest requirements for the Tier-0 and Tier-1 sites and looser

requirements for the Tier-2 sites. There is no strict definition or requirement for the Tier-3 sites, those are local computing farms beyond the control of CMS.

At the Tier-0, prompt reconstruction of the recorded data is performed before the data stored and exported to the Tier-1 centres. The Tier-1 centres are used for long-term storage of the data and for the re-reconstruction of the data. Re-reconstruction is needed for example when the alignment of the detector is known more precisely than during prompt reconstruction or in case there were failures in certain detector components which were not accounted for correctly. The Tier-2 centres are used for the production of Monte Carlo events (see 5) and for user analysis. Local computer centres are called Tier-3. They are not necessarily under the central control of CMS. They are also used for Monte Carlo production and user analysis.

To be able to process CMS workflows, all sites require the CMS software being pre-installed. With the exception of only very few sites, all CMS software deployments and deletions are managed centrally. In the following it is described how the deployment process works and the various steps of automation are explained.

4.3 CMSSW: Packaging and Installation

4.3.1 Packaging

The main CMS software stack CMSSW is packaged in the RPM (Redhat Package Manager) (e.g. [73, 74]) format. Dependencies of the RPMs are handled with the `apt-get` tool. A single CMSSW release is a rather monolithic build, which includes almost all base functionalities for Monte Carlo (MC) simulation, event reconstruction and event displaying. A number of external packages can be shared among several CMSSW releases. External packages are for example ROOT and GEANT4, MC generators, parton density functions but also a C++ compiler and various libraries.

CMSSW is available for different architectures, where an architecture is defined by a certain compiler version and the system for which it is compiled, e.g. compiled with GCC 3.4.5 for the `ia32` architecture under Scientific Linux 4 or compiled with GCC 4.6.4 for the `amd64` architecture under Scientific Linux 5. Present CMSSW releases are officially supported on Scientific Linux 5 only. Nevertheless there have been successful installations on various other (mainly Linux and MacOS) platforms.

When looking at the disk space used after installing one version of CMSSW from scratch, one finds that about 5.5 GB of disk space are occupied in 115 thousand files. These files and another 4 thousand symbolic links reside in 11 thousand directories.

About 2.5 GB of this space is used by more than 80 external packages. Further releases require less space because of package sharing with already installed releases. Depending on whether a patch release or major release is added to the installation, a few 100 MB up to several GB of additional disk space are needed. The largest external components of a CMSSW release are shown in figure 4.1.

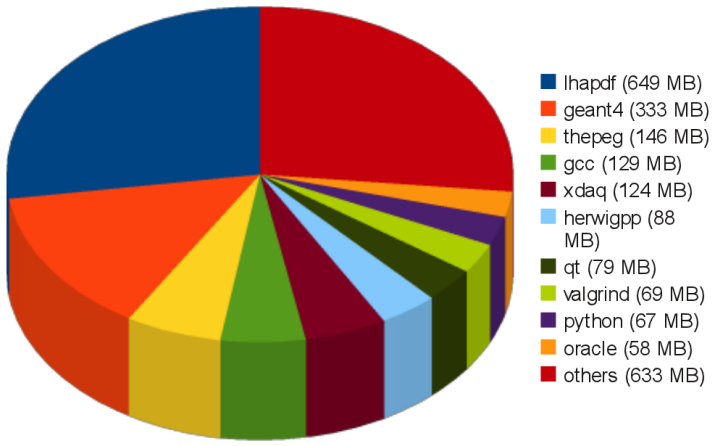


Figure 4.1: The largest external components of a CMSSW installation (determined from a clean installation of version `CMSSW_3_6_0`).

4.3.2 Installation

Installation of the CMS software is done in two steps. The first step is called bootstrap and is done only once per architecture to prepare the installation. It first checks that required libraries on the operating system are installed and sets up the package management software and installs common tools which are needed by all CMSSW versions. Because CMSSW is installed under a user account without system administrator rights, packages cannot be added to the RPM database of the host OS. Thus, a separate RPM database within the CMSSW installation directory is created. To fulfil all dependencies of the CMS software on software which is not shipped with CMSSW, the CMS RPM database must know a list of installed system libraries. Therefore an RPM package with the only purpose to fulfil these dependencies is created and installed into the CMS RPM database. This is shown in figure 4.2.

Once the bootstrap has been done, the actual installation can be started. After ensuring that the environment is set up correctly for the use of the CMS RPM database and the CMS versions of the package manager, the `apt-get` tool can be used to update the package repository and to install a CMSSW release including all required dependencies. Unlike the bootstrap, this step needs to be repeated every time a new release is installed.

4.4 Installation on Grid Sites

The installation of CMSSW on Grid sites is managed centrally by a deployment team (of which I was a member). This way, problems that might occur during installation can often be diagnosed and fixed very fast; site administrators are only involved if no solution is possible within the abilities of the deployment team, e.g. fixing misleading site configurations.

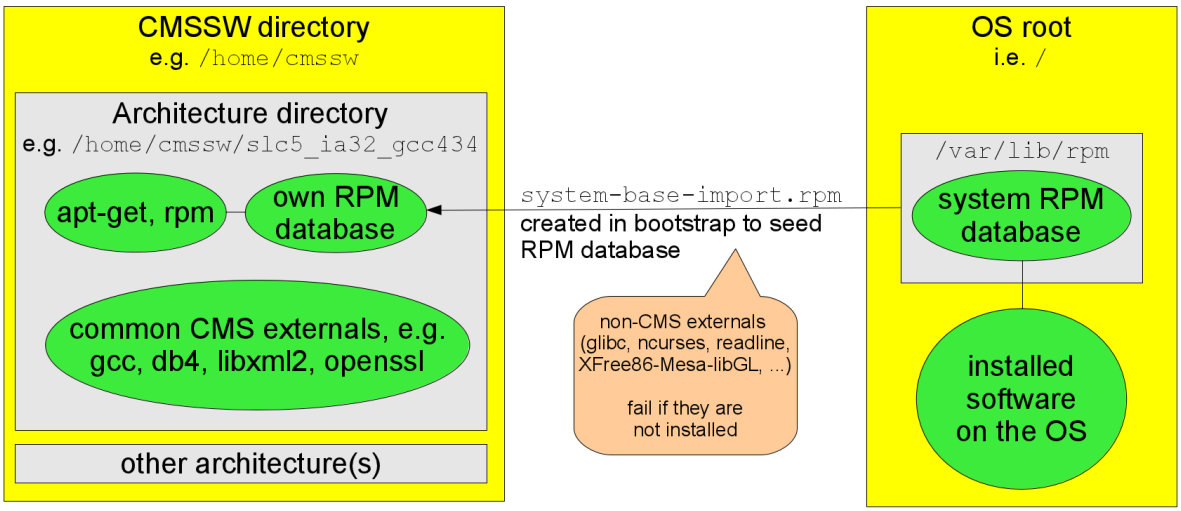


Figure 4.2: Bootstrapping a CMSSW release. The left side shows the CMSSW release directory in which a private RPM database is installed. Required components that are already provided by the operating system are collected in the `system-base-import.rpm` package. It just fulfils these dependencies and does not include any non-metadata.

At Grid sites the CMS software gets installed on a shared file system that is mounted on all Worker Nodes. Prominent choices for the technology of that shared file system are NFS, AFS or Lustre. The installation on that shared file system works in principle like a local installation. The main difference is that the deployment team has no local accounts at the remote sites and all installation jobs have to be sent via the Grid. These jobs are using the VOMS role *lcgadmin* to be prioritised and to gain write privileges to the shared file system. Due to the nature of Grid jobs, the installation is running unattended and thus requires a well-tested installation procedure in order not to leave a broken software installation.

In the following the different procedures used for sites with gLite[1] or ARC[2] middleware (Europe, Asia) and OSG[3] sites (America) are described.

4.4.1 gLite and ARC

The contribution of this work is the development of a program called `release-installer` (RI) for the sites running gLite or ARC middleware to handle the Grid job creation. Not only this single program, but several reusable classes and tools have been written

[1] gLite. `http://glite.web.cern.ch/glite/`
[2] Advanced Resource Connector. `http://www.nordugrid.org/middleware/`
[3] Open Science Grid. `http://www.opensciencegrid.org/`

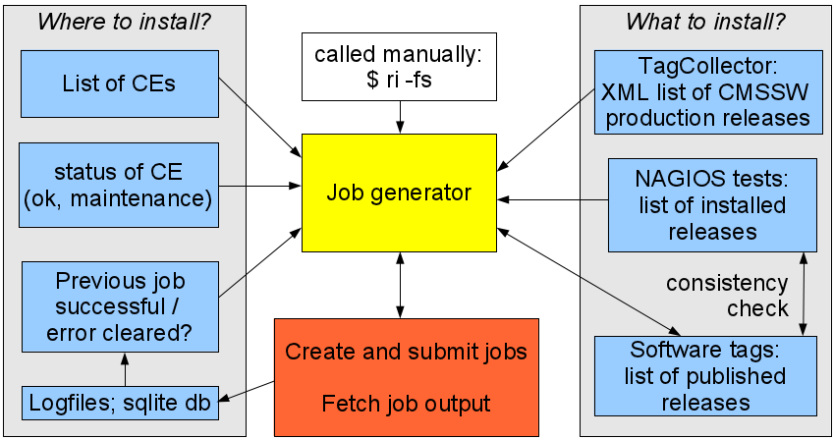


Figure 4.3: Automated installation on gLite and ARC sites

to handle all steps from the determination of the sites and the releases to install over a wrapper around the existing `apt-get` tool to the creation of grid jobs and parsing the their output values.

On the left side of figure 4.3, it is shown how the RI fetches the list of computing elements (CE) it needs to handle. The main input is a manually managed list of CEs. Experience has shown that generating this information automatically is not feasible. In principle a list of all CEs is available from the top-level BDII[4], but certain sites do not want to participate in the central deployment. CMS maintains a special database called siteDB [75] to track available computing resources. Basing the deployment on siteDB would require a high discipline to keep the information up to date, which is often not the case. Therefore the choice to keep a manual list is currently the best.

For all CEs in the list, the RI first checks their status. Only those CEs that are currently available are considered; the CEs that are in maintenance or unavailable for other reasons are skipped. Further all those CEs are skipped where a previous job did not run successfully unless the deployment operator has looked at the logfiles and cleared the error state.

The right side of the figure shows how the RI determines the releases to install. The CMS TagCollector contains information about all CMSSW releases, particularly whether it is a current production release or if it is already deprecated. Site specific requests to install additional releases are read out by the RI from a manually maintained XML file.

NAGIOS tests provide a list of installed releases on a CE, so do the software tags. The list of installed releases is then compared to the list of production releases and manual requests. An installation Grid job is created if there is a difference. Once the job is done, its output is parsed and stored in a SQLite3 database.

[4]Berkeley Database Information Index. `https://twiki.cern.ch/twiki/bin/view/EGEE/BDII`

4.4.2 Object oriented deployment framework

To create the Grid jobs and to install the software on gLite and ARC sites, an object-oriented deployment framework has been newly developed as the technical part of this thesis. It consists of about 10 classes and several thousand lines of code. See figure 4.4 for an UML class diagram. In the following, the different classes and their functionality are described.

4.4.3 Releases

Information about CMSSW releases is handled in the `CMSSWRelease` class. It is a very simple class where the instances contain only the name and architecture of a release. For a given release it can find out if it is announced or deprecated. In addition, it provides static methods to retrieve the full list of production and deprecated releases. Even though one CMSSW release can be available for different architectures, it is only supported officially for exactly one architecture. Using only official releases, the class can also determine the architecture automatically in most cases.

4.4.4 Actual software installation

All installations steps of the CMS software as described in section 4.3.2 are implemented in the class `CMSSWInstall`. Each class instance contains information to maintain a CMSSW installation of a specific architecture in a given local path. The class takes care about setting up the correct environmental variables, does the bootstrap and updates the RPM database if necessary, and installs or removes CMS software. Because the class does not know anything about Grid computing, it can be tested thoroughly on a local directory before being used in Grid jobs.

In addition, a workaround for RPM databases on file systems without locking support is included (see section 4.6.1 for details).

4.4.5 Computing Elements

Information about computing elements is encapsulated in the `LcgCE`[5] class which collects and caches data from several different sources, for example from the CMS siteDB, the CMS dashboard, from NAGIOS tests and from the BDII. Thus all information can be accessed via a single interface, and also differences between ARC, CREAM[6] and LCG CEs are hidden.

For the deployment, the most important information provided is the list of installed releases, the list of published releases and information about the current status of the CE,

[5] The name is historic and is now misleading, it supports also ARC and CREAM CEs.

[6] Computing Resource Execution And Management Computing Element. `http://grid.pd.infn.it/cream/`

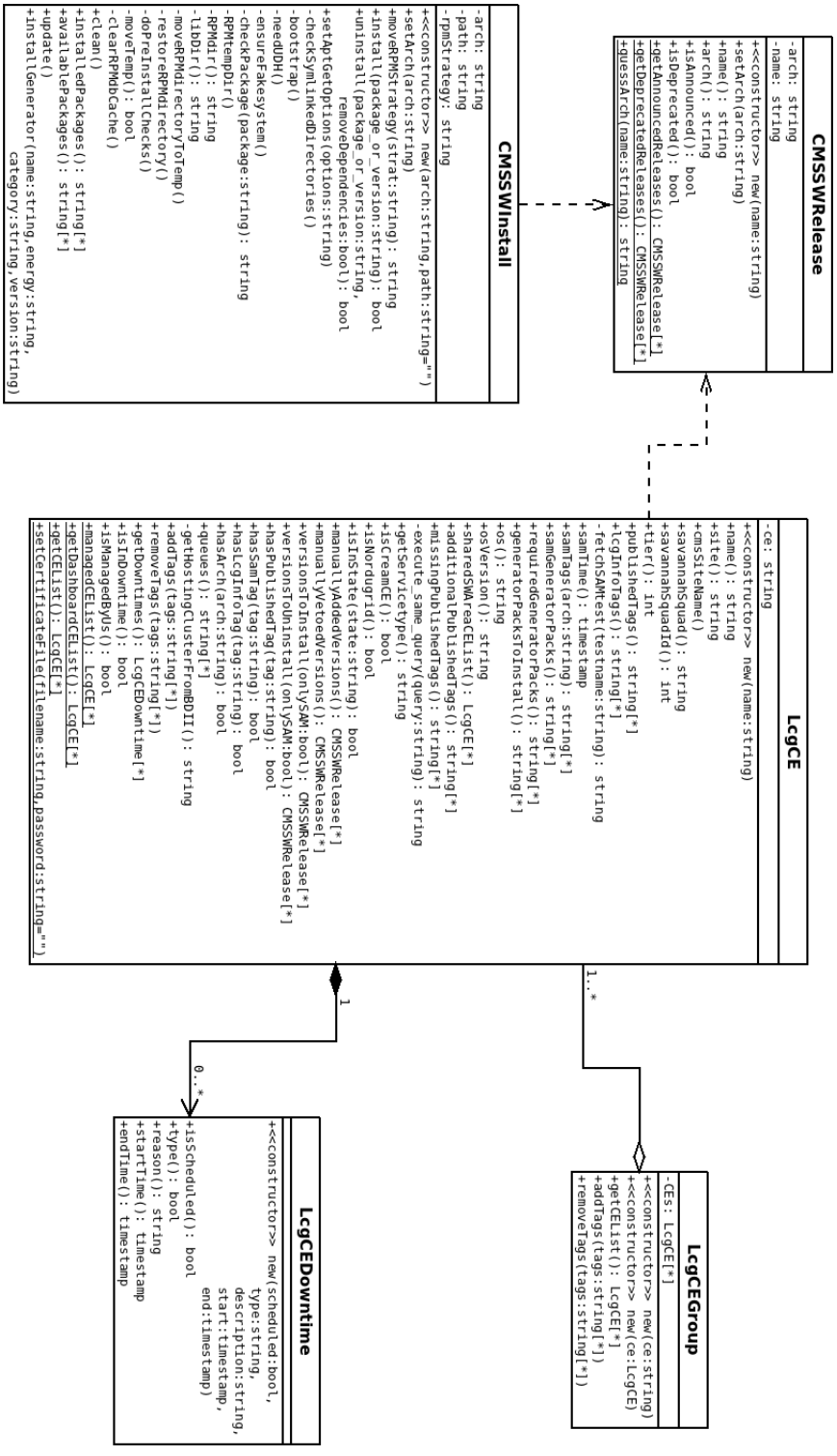


Figure 4.4: Classes developed for the deployment software: Installation and status checking

e.g. if it currently available. Downtimes of the CE are read out and stored in LcgCE-Downtime objects. In addition, many small methods are implemented for debugging and/or statical reasons.

Some sites with multiple CEs share the CMSSW installation among some (not necessarily all) of their CEs. The `LcgCEGroup` class can be used to determine a list of CEs accessing the same installation area. Thus it is possible to send only one installation job to the group of CEs to have the software installed or removed on all CEs of that group.

4.5 Grid job submission

To deploy CMSSW on Grid sites, a Grid job needs to be created for each installation on each site. Thus the `JDL` class (see figure 4.5) has been developed to create, submit and retrieve Grid jobs including support for MonALISA[7] based reporting to the CMS dashboard.

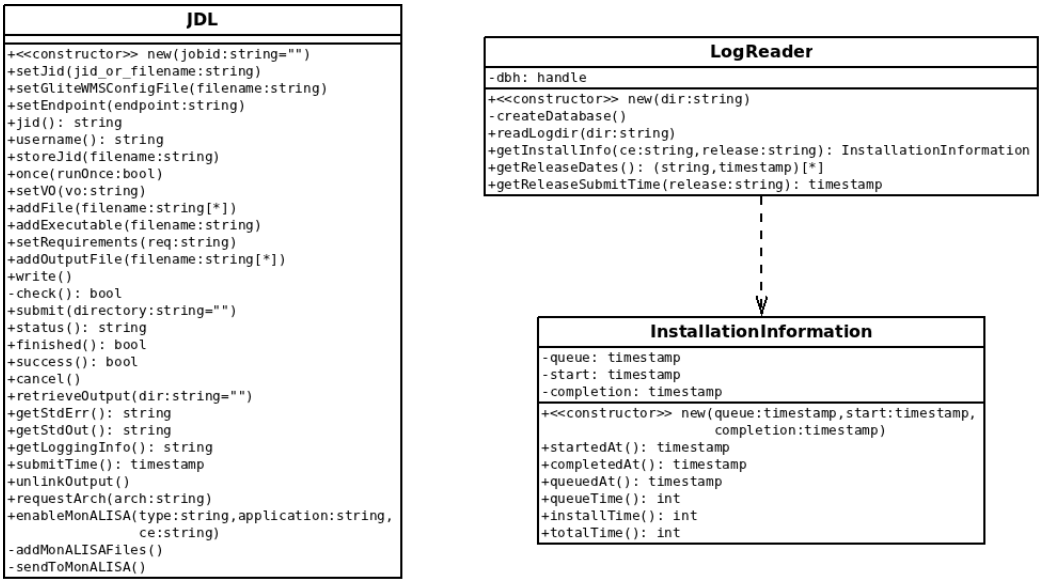


Figure 4.5: Classes developed for the deployment software: Grid job creation and monitoring

The framework is also used by a CGI script to view the deployment status and the logfiles through a web browser. This is especially helpful in case of deployment errors.

[7] Monitoring Agents using a Large Integrated Services Architecture. `http://monalisa.caltech.edu/`

4.5.1 OSG

OSG sites (all US sites and a few others) have a different setup and can execute the installation jobs directly on the CE (i.e. not on a worker node). Also a different framework is used, in which the deployment across Tier-2 and Tier-3 sites is triggered from a CRON job. It reads out the list of production releases from the TagCollector and compares it to the list of installed releases for each site which is stored in a MySQL database. As soon as there is a new release, the installation starts automatically.

The deployment software for OSG sites is not part of this work, it is currently maintained by Bockjoo Kim (University of Florida).

4.5.2 Release deprecation

As the CMS software is constantly developed and improved, old releases become obsolete and need to be removed from the software area. This not only forces physicists to use newer and better releases but also ensures that the amount of disk space needed (see 4.3.1) does not exceed a certain limit.

The deprecation process starts with a proposal sent to the release announcement mailing list. Unless there are objections, the deprecation announcement is sent typically a week later and the TagCollector is updated. After the deprecation, only software tags are removed from all sites so that no new jobs using the deprecated releases can be submitted. Analysis jobs which have already been submitted should not fail because of missing software, thus removal jobs are not sent immediately but only after a grace period of at least five days. While new releases are being installed on a regular basis, deprecation rounds happen only a few times per year and many releases get deprecated at the same time.

4.6 Experiences

4.6.1 File Locking

The CMSSW installation of a Grid site usually lives in a shared file system like NFS that is mounted across the compute nodes. It has turned out that file locking, which is required for the RPM database and certain files used by `apt-get`, often causes problems when used with shared file systems. A workaround has been implemented which copies the RPM database directory into a temporary directory on the local machine where no locking problems occur. The old RPM directory is then renamed and used as a backup and a symbolic link is created to link the temporary RPM copy to its original place. With this setup, RPM can lock its files locally and do the installation. Afterwards the link is removed and the RPM directory is copied back to its original position in the shared file system.

Working on a copy of the database has the additional benefit that we do not end with a corrupt RPM database even if the Grid job is cancelled at any time because a backup of the original version is restored automatically with the next installation job. Thus the whole procedure is done even if the shared file system at a site supports locking properly.

4.6.2 Differences in site configuration

A lot of sites have a simple configuration with only one CE and a single CMS software area. The CE publishes the supported architectures and the installed CMSSW releases via software tags which are set right after the installation has finished. However, sites can have more than one CE. These CEs can either share the software area or have separate areas. If they are separate, the software has to be installed on all CEs, otherwise the installation has to be done on only one of the CEs. In the latter case, it does not matter which of the CEs is selected to install the software, so that in case of a downtime of one CE the installation can be done on a different CE. This is detected automatically by the release installer.

Not only the software area can be shared among the CEs but also the list of software tags. All combinations are possible, also the configuration of a shared file system and separate tags. In this case software tags need to be set on all CEs although installation was only done on one CE. This has also been automated. However, manual tag correction is still required if errors occur while the tags are changed. In principle the tags can be corrected automatically using a CRON job. Using the developed classes, such a program was written and put into operation.

However, soon a problem showed up: Sometimes site administrators had removed some tags intentionally to stop receiving new analysis or production jobs. The CRON job has then restored these tags and new jobs were sent to that CE.

The conclusions from this was to disable the automatic tag correction. Instead, tag setting was made more robust. The most common problems of the `lcg-tags` command was a lock file which was left behind. To change the tags on a CE, `lcg-tags` first downloads the file in which the tags are stored, then modifies it locally and finally reuploads it. To avoid problems with concurrent access, a lock file is created by `lcg-tags` in the beginning and removed at the end. If a lock file exists, `lcg-tags` exits with an error message. The `LcgCE` class now detects this automatically and tries to rerun `lcg-tags` after a delay of three minutes. If the lock file still exists after three tries, it is removed automatically before `lcg-tags` is run once more.

In addition, the release installer also checks for the correct tags after fetching installation jobs and issues a warning if the result looks suspicious, for example when the list of tags is too short.

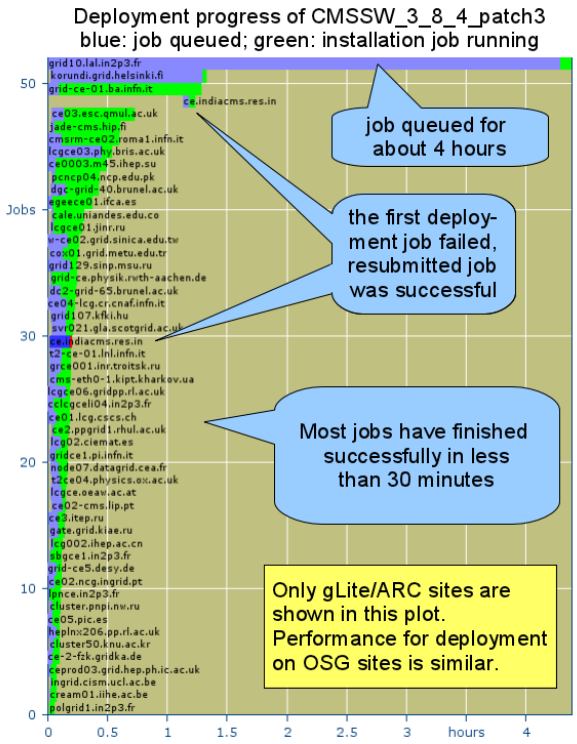


Figure 4.6: Typical deployment plot of a patch release (here: `CMSSW_3_8_4_patch3`). Installations during the first 5 hours are shown. The blue bars indicate jobs in a queue, the green bars show the timespan of the actual installation. A failing job is shown in red (time from execution to the failure), its queueing time is shown in darker blue.

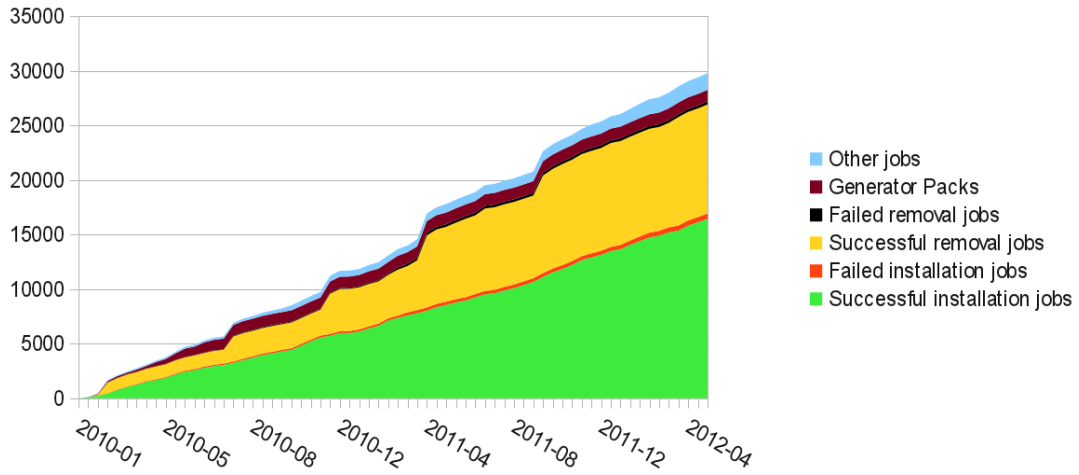


Figure 4.7: Overview of all grid jobs sent with the deployment tools since 2010. Jobs which have installed or removed more multiple releases are counted multiple times.

4.6.3 Typical deployment cycle

Typically, the time from the release announcement to complete installation ranges between 30 minutes and a few hours for most sites. The time variation is due to the size of the release, the performance of the file system, the internet connection of the site, and the site availability. Figure 4.6 shows a typical deployment progress plot for a patch release. Many jobs have finished in about half an hour, but there is also one site where the first installation attempt has failed and only the second attempt was successful. Also one site with a long queuing time (here: about 4 hours) is visible.

Due to the large number of sites, it sometimes happens that a site is in a downtime, that an error occurs during the installation, or simply that the job is queued for a very long time. On gLite/ARC, deployment jobs are queued for a median of only 9 minutes, although a few outliers with up to 60 hours queuing time lead to a much higher average. Once the job is running, the installation is done in less than 15 minutes on average.

These numbers show that the completion of a deployment round can take a few days even if the releases have been deployed to 90% of the sites within an hour.

The deployment system is in operation since January 2010. As of May 11th, 2012, the software has dealt with 416 different CMSSW releases. Figure 4.7 shows how many installations and removals have been performed over time. Jobs to fix problems or to install software which is not included in the CMSSW framework are shown as *Other jobs*.

A description about the whole deployment process has been presented at the CHEP 2010 conference and proceedings have been published in [76].

4.7 Future plans and conclusions

The installation procedure as it is done now has several disadvantages. There are two independent deployment teams, one for OSG sites and one for all other sites. In addition, there is a separate installation area at CERN managed by CMSSW developers. The installation faces a variety of problems, among them the ones mentioned in section 4.6.1. However, one must not only consider issues during installation but also issues during operation. Grid sites can have thousands of worker nodes and all of them are accessing the same CMSSW installation area. The file server, whether it is using AFS or NFS, can thus easily become a bottleneck. Even parallel filesystems like Lustre do not solve the problem since the the CMS software needs to make many metadata requests (i.e. `stat` calls, e.g. to find out if a file in a given path exists). Also these filesystems are usually optimised to handle a large amount of data in relatively few files and not a large amount of files (metadata).

To save the manpower for the deployment teams, to avoid all installation problems and also to find a solution for the metadata bottleneck, it is planned to use the CernVM-FS in CMS. CernVM-FS is a scalable filesystem designed for read-only access. It runs locally on each worker node and downloads required files on demand. It uses the `http` protocol for the download and therefore it is easy to cache the downloads per site using a proxy server. A detailed description would be out of scope for this thesis, see [77] for details. Several CMS sites are already using CernVM-FS or are in an evaluation phase.

Even though CernVM-FS makes the installation tool developed in this work obsolete, a lot of time has been saved since it was put into operation in January 2010, now more than 2.5 years ago. It will still be in use for a longer time until all sites have migrated. Even then, parts of the tools can still be used, for example as an easy way to collect information about computing elements or to submit special jobs to any and/or all sites.

Chapter 5

Event simulation

Modelling events is a very important task in particle physics and is used for several different purposes. By simulating a certain signal, it is possible to predict its detailed behaviour such as the distributions of different observables and also the kinematics of non-observable intermediate particles. Some generators (POWHEG, MC@NLO) [78, 79] also predict a cross section, while the output of other generators needs to be normalised to a pre-calculated cross section. When searching for new processes, a good understanding of all background processes is required—also these are usually modelled using a simulation.

The simulation of a process is usually divided into several simulation steps. When the hard process is calculated, its *matrix element* (ME) is usually not known to all orders, but it is rather calculated in a particular order in perturbation theory, e.g. in Leading Order (LO) or Next-to-Leading Order (NLO).

To cover different initial and final states according to their probability, the incoming partons are simulated with random momenta based on the measured PDFs (see 2.5) while outgoing particles are distributed randomly in their phase space. After the simulation of the hard process, *parton showers* (PS) are simulated using DGLAP evolution equations [80] to account for initial and final state radiation.

The equations not only depend on α_S but also on the PDFs. In addition, *multiple interactions* (MI) in the same events and particles from beam remnants need to be taken into account.

While perturbation theory can be applied for the hard scatter ($Q \gg \Lambda_{\mathrm{QCD}}$), this is not true any more for the outgoing particles. At the energy scale of $O(\Lambda_{\mathrm{QCD}})$ they *hadronise*, a process where the partons need to be converted to colour neutral hadrons. The implementation for this step relies on phenomenological ideas and needs to be tuned to the data. This also means that not only many measurements depend on the simulation, but also that the simulation depends on the measurements. The measured results can be put back into the event generators to create a new, improved tune. A different way to improve the simulations is the use of higher order calculations as done in POWHEG and MC@NLO.

Finally the outgoing particles are flying into the detector where they interact with the material and may decay. This step requires a good simulation of the detector and its

components.

The whole chain of the event generation can be written in a figurative sense as [81]:

$$(\text{ME} \oplus \text{MI}) \otimes \text{PS} \otimes \text{Hadronisation} \otimes \text{Detector simulation} \tag{5.1}$$

In the following, the Monte Carlo method and the event generators used in this thesis are explained.

5.1 Monte Carlo integration

The Monte Carlo integration is a general way to calculate integrals numerically using random numbers based on calculating the integral as an average. It is mainly used for the integration over a multi-dimensional space where other numerical integration methods such as Gaussian integration or Simpson's rule are no longer useful [82].

In the use case of a cross section calculation, a given function $f(\mathbf{x})$ could represent a differential cross section (which is proportional to the matrix element $|\mathcal{M}|^2$) which needs to be integrated over a volume V in the n dimensional $\mathbf{x}$ space. The volume V would correspond to the phase space Ω.

Thus the following integral I could be used to calculate the total cross section:

$$I[f] = \int_V d^n x f(\mathbf{x}) \tag{5.2}$$

The Monte Carlo integration method makes use of the central limit theorem by estimating the integral using the mean value $\langle f(\mathbf{x}) \rangle$:

$$I[f] \approx \langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i) \tag{5.3}$$

The estimated error of this method depends on the variance of f and on the number of simulated events:

$$E[f] = \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N-1}} \tag{5.4}$$

To reduce the uncertainty on the simulation, one simply needs to increase the number of simulated events in which case the error is proportional to $1/\sqrt{N}$, irrespective of the dimension. Alternatively one can try to reduce the variance. This can be done using a mapping $\mathbf{x} \to \mathbf{y}(\mathbf{x})$ which flattens the integrand and then integrating over the corresponding volume in $\mathbf{y}$ space. This method is called *importance sampling*.

Detailed descriptions of Monte Carlo integration methods can be found in the appendix of [82] as well as in [83].

5.2 Pythia

PYTHIA is a multi-purpose event generator for e^-e^+, ep, and $pp/p\bar{p}$ colliders. Its history dates back to the Jetset generator from 1978. In more than 30 years of development, PYTHIA has been constantly improved with new physics models, more options, and better tunes to data [82]. The version used for the simulated samples used in this thesis is PYTHIA6.

In general, PYTHIA is easy to use and it can not only produce all Standard Model processes but also several processes beyond the Standard Model. Furthermore it can simulate all the steps shown in equation 5.1. Its only drawback is that the hard scattering process is only calculated to leading order while higher orders are simulated with parton showers. PYTHIA provides different tunes which modify the showering algorithms, the main difference being the ordering of the parton showers by the invariant mass (Q^2 ordered showering) or by the transverse momentum (p_T^2 ordered showering). Also the modelling of the underlying event depends on the showering. Only for the p_T^2 ordered showers a model with interleaved initial state radiation and multiple interactions is available. The tune Z2 which is used in this work contains a p_T^2 ordered showering model and uses CTEQ6L PDFs [33].

A full description of PYTHIA6 can be found in [84].

5.3 Herwig

Also HERWIG [85] is a general-purpose event generator. It supports hard lepton-lepton, lepton-hadron, and hadron-hadron scattering as well as soft hadron-hadron processes. HERWIG also implements the parton showering, taking into account colour coherence effects and azimuthal correlations within and between jets. One difference to PYTHIA is that HERWIG uses a cluster model for jet hadronisation while PYTHIA uses a string model [86, 87].

5.4 MadGraph

Unlike PYTHIA, MADGRAPH [88] is not a general-purpose generator but specialised to pp and $p\bar{p}$ collisions. It also does not include the hadronisation step, i.e. a final state consists of leptons, quarks, and gluons. These particles are then passed on to a different generator for the hadronisation, in this work always to PYTHIA.

To be exact, MADGRAPH itself is only a matrix element generator. It calculates the matrix element on tree level up to a given order and calculates the squared amplitudes of the corresponding Feynman diagrams. MADGRAPH is limited in that it can produce a maximum of eight final state particles [89]. The integration over the phase space is subsequently done by the MADEVENT event generator and only then the event is passed

to a hadronisation generator. In this work, the MadGraph/MadEvent generation is referred to as MadGraph.

This way of generating events separately for the hard process plus 0, 1, ...jets leads to a precise description of the event topology. One problem with it is that one has to avoid double counting due to the parton showering. This is done using the k_T MLM scheme [86] with a threshold of 20 GeV. Another problem is that the calculated cross section is strongly scale dependent and needs to be normalised afterwards, here with predictions from MCFM [90].

When comparing with data it is observed that MadGraph/Pythia describes the data better than Pythia alone, especially at higher jet multiplicities [91]. Therefore it is used as event generator for most samples in this work. This effect is also be visible in this work in the low invariant dilepton mass region of the Drell-Yan simulation where a MadGraph sample agrees much better with data than a Pythia sample (see chapter 10).

5.5 Powheg

In contrast to the leading order calculations in Pythia, Powheg [78, 92] (*POsitive Weight Hardest Emission Generator*) calculates the hard process at next to leading order. However, Powheg does not include the parton showering and thus needs to be interfaced with Pythia or herwig. This is done such that the NLO cross section is preserved after the showering, no additional normalisation is needed.

In this work, Powheg/Pythia is used for the simulation of single tops, and both Powheg/Pythia and Powheg/herwig are used for the $t\bar{t}$ signal to determine differences in the hadronisation modelling.

5.6 MC@NLO

MC@NLO is an event generator which calculates the matrix element in NLO and matches them with a parton shower simulation based on herwig. It allows the simulation of Higgs boson, single vector boson, vector boson pair, heavy quark pair, single top (with and without associated W or $H^{\pm}$), lepton pair and associated $H+W/Z$ production in hadron collisions. Details can be found in [79].

One specific feature of MC@NLO which needs to be taken into account in the analysis is that a small fraction of events with negative event weights are produced.

5.7 Tauola

The decay of τ leptons is handled by the separate package Tauola [93, 94] which was first published in 1990. It contains sub-programs for several different τ decay modes.

TAUOLA takes the spin information into account and applies QED corrections. As it only handles the τ decays, TAUOLA is used in connection with the other event generators described in this chapter.

5.8 Geant4

After an event has been simulated with any of the previously described event generators, it must also be simulated how such an event looks like in the detector. In particular the interactions with matter of the detector need to be calculated. This includes bremsstrahlung, electromagnetism showering, multiple scattering and hadronic interactions. To simulate all that, the complete detector geometry including passive support structures, technical necessities such as cooling pipes and of course also the active detector components must be used as an input to the simulation which is done using GEANT4 (GEometry ANd Tracking) [95]. Apart from the geometry, also the response of the different detectors needs to be simulated so that at the end the same reconstruction algorithms can be used for data and simulation. Furthermore GEANT is tuned, correcting for deviations between data and simulation.

Excellent predictions can be made about the behaviour of physics processes at the end of all simulation steps.

Chapter 6

Event reconstruction

The reconstruction of particles from measured quantities like energies in the calorimeter cells is one of the main tasks for any analysis. As described in chapter 3.2, the CMS detector consists of several components which are each specialised to detect certain particles. Many particles can however be reconstructed in more than one detector component, leading to an improved resolution, higher efficiency or lower misidentification when the information from all components is combined.

For each particle type dedicated reconstruction algorithms exist. In the following the reconstruction details of the relevant particles are explained.

6.1 Muon reconstruction

Muons can usually be unambiguously reconstructed because they are the only particles which are detected in the muon system. This can be seen in figure 6.1 which shows a slice of the CMS detector. The muon (blue line) flies from the tracker on the left side through the calorimeters, the solenoid, and the complete muon system on the right side. As muons are charged, their tracks can also be reconstructed in the inner tracker. In the calorimeters however they only deposit a small amount of energy [64]. In addition one can see the bending of the track caused by the magnetic field.

In CMS the tracks in the tracker and in the muon system are reconstructed independently of each other. To combine them, two different strategies exist [96]: an outside-in and an inside-out algorithm.

The muons reconstructed outside-in are called *global muons*. Here tracks in the muon system are extrapolated to the tracker and compared to all tracks in there propagated to the outside. A Kalman filter is used to fit a combined track. For low transverse momenta the momentum measurement is dominated by the tracking system, while from $p_T \gtrsim 200$ GeV the muon system improves the tracker measurement.

The inside-out method produces *tracker muons*. Here tracks in the tracker with $p > 2.5\,\text{GeV}$ and $p_T > 0.5\,\text{GeV}$ are extrapolated to the muon system where only one muon segment needs to be matched to the inner track. The matching must happen within a distance of 3 cm of the extrapolated track or the difference in the position divided by the summed position uncertainties must be less than 4. The advantage of tracker muons

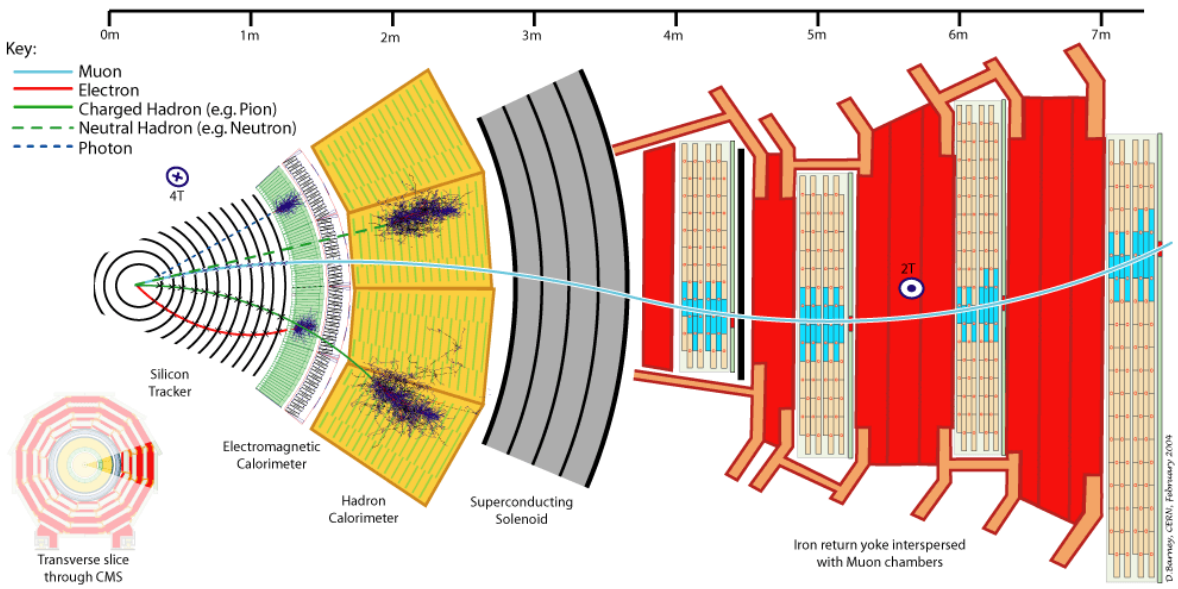


Figure 6.1: A transverse slice through the CMS detector [97]. Muons (light blue), charged hadrons (green), neutral hadrons (dashed green), photons (dashed blue), and electrons (red) are shown.

is that they are also efficient for low muon momenta $p \lesssim 5$ GeV at the price of a higher misidentification, for example due to hadrons which are not completely stopped in the calorimeter (punch-through) and produce a signal in a muon chamber.

Muons used in this analysis fulfil several quality criteria. They need to be reconstructed outside-in and inside-out and the combined track's χ^2 per degree of freedom needs to be smaller than 10. To ensure a well reconstructed track, there have to be at least 11 hits in the inner tracker from which at least one hit is in the pixel detector. The muons also have to originate from the hard process, reflected by the requirement that the transverse impact parameter $|d_{xy}|$ needs to be smaller than 0.2 mm.

6.2 Electron reconstruction

As electrons, shown in red in figure 6.1, produce a signal in the inner tracker and in the electromagnetic calorimeter, there are again two ways to reconstruct electrons: One can use tracker information as a seed and extrapolate into the calorimeter or take energy deposits in the calorimeter and search for a track [98]. The former approach is better suited for electrons with a very low p_T as the latter method requires $p_T \gtrsim 5$ GeV for the ECAL seeding.

The quality criteria applied are optimised to select electrons from W decays and reject fakes from jets or from converted photons. Due to the large magnetic field, due to

the varying amount of material in the tracker and of course also due to barrel/end cap differences, the electron identification has been divided into categories. For this analysis, the criteria defined by the `eidTightMC` working point are used [99, 100].

Apart from the obvious separation into barrel and end caps, three further categories are defined according to different amounts of radiation in the tracker. Finally the transverse energy of the electron candidate splits these categories into three further categories.

The variables used for the selection are the position of the track and the number of (missing) hits, the energy deposit in the calorimeter, the energy measured in the calorimeter vs. the momentum measured by the tracker, the fraction of energy deposited in the hadron calorimeter, the cluster shape and the isolation (see section 6.3). In total there are 180 cut thresholds per working point which are described elsewhere [101].

In addition an impact parameter requirement $|d_{xy}| < 0.4$ mm is imposed and a conversion rejection cut is applied. It allows a maximum of only one hit in the inner tracker which is not reconstructed. Additionally a check is made that no oppositely charged partner track T_2 exists in addition to the electron candidate track T_e. T_2 is considered a partner track of T_e if the distance between T_e and T_2 at the point where both tracks are parallel in the r/ϕ plane is smaller than 0.2 mm and where $\cot\theta(T_e) - \cot\theta(T_2) < 0.02$.

6.3 Isolation

Isolation is an important criterion to distinguish leptons from the hard process, for example from a W or Z boson (*prompt leptons*), from leptons or misidentified hadrons in jets. Prompt leptons may fly through the detector in any direction. Their track and the energy deposit in the calorimeters usually do not overlap with jets or other particles, in contrast leptons which are produced from particles in a jet fly roughly in the same direction as the jet. For misidentified hadrons obviously the same applies.

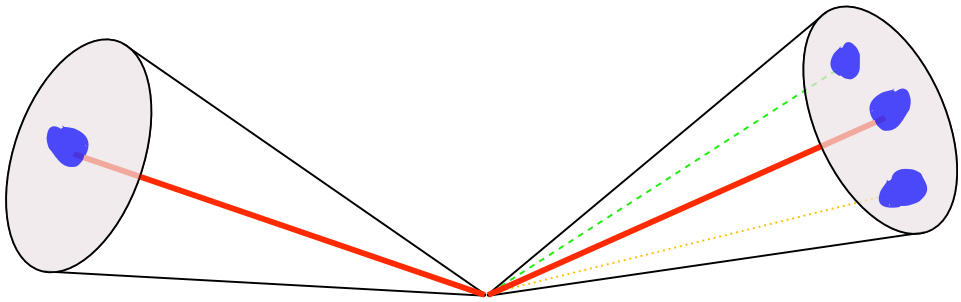


Figure 6.2: Isolation cone around lepton candidates (red thick line). The blue, pear like spots show the energy deposit in the calorimeters. The particle on the left side is isolated; on the right side two other particles (green dashed and yellow dotted) are inside the cone so that the particle isn't isolated.

To tell prompt and non-prompt leptons apart, the amount of additional energy in a

cone around the particle, but without the particle itself, is calculated. If the additional energy is small, the particle is isolated, otherwise it isn't. The isolation can be calculated in the tracker and the electromagnetic and hadron calorimeters. The value is combined (summed) to get a good separation power:

$$I_{comb} = I_{tracker} + I_{ECAL} + I_{HCAL} \tag{6.1}$$

To increase the power even further, the additional energy is put in relation to the transverse momentum of the lepton $p_T(\ell)$:

$$I_{comb}^{rel} = \frac{I_{tracker} + I_{ECAL} + I_{HCAL}}{p_T(\ell)} = \frac{I_{comb}}{p_T(\ell)} \tag{6.2}$$

This is done to identify high energetic leptons as isolated even if there is a little bit of other energy in the cone which would mark a low energetic lepton as not isolated.

6.4 Jet reconstruction

At hadron colliders like the LHC many processes lead to hadronic final states. Because the final state quarks and gluons hadronise, each of these lead to many particles which move mainly in the same direction, they form a *jet*. To exploit these jets one must first reconstruct them. While in events with very little hadronic activity it is rather simple to decide which energy deposition respectively particle belongs to which jet, it is much more difficult in events with much hadronic activity. In addition, pile-up events increase the amount of jets to reconstruct. In the following commonly used jet algorithms and their properties are explained.

6.4.1 Collinear and infrared safety

A very important feature for jet algorithms is that they should be both collinear and infrared safe [102]. Collinear safety means that the resulting jets should not change when there is a collinear splitting of jet components, infrared safety ensures that additional soft emissions from hard jet components do not change the result. This is important because jet algorithms that do not have these properties can produce totally different results just because of hadronisation, the underlying event, or pile-up—even for high energetic jets.

6.4.2 Jet algorithms

Several algorithms have been developed which can reconstruct the jets. They can be divided into two groups: cone-type algorithms and sequential clustering algorithms. The cone-type algorithms usually require a seed, for example the most energetic particle,

and try to find a jet around this particle. These algorithms tend to be rather fast in terms of computing performance and were therefore discussed for earlier hadron colliders. However, typically they are not infrared and collinear safe and involve several complex, non-physical parameters and are therefore disfavoured [102].

In the following, the sequential clustering algorithms are described of which the anti-k_T algorithm is used for jet reconstruction in this thesis.

6.4.3 The anti-k_T algorithm

The main goal of any jet clustering algorithm is to combine many measured particles into several jets. The sequential clustering algorithms iteratively either combine two objects (a particle or a pseudo-jet consisting of several particles) to a new object or decide that a jet has been found. The decisions are done in the following way: Let d_{ij} be the distance between two objects i and j and d_{iB} the distance between an object i and the beam B. All these distances are calculated and the smallest one is looked at. In case it is the distance between two objects i and j, these two objects are combined into a single new pseudo-jet object and added to the event while the source objects are removed. Then the algorithm starts over. In case a distance of type d_{iB} has the smallest value, the corresponding object is identified as a jet and removed from the event. This is repeated until there are no objects left in the event.

How the algorithm works depends on the exact definition of the distances. They are defined as follows [103]:

$$d_{ij} = \min(k_{Ti}^{2p}, k_{Tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \tag{6.3}$$

$$d_{iB} = k_{Ti}^{2p} \tag{6.4}$$

$$\text{where} \quad \Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{6.5}$$

In these equations k_{Ti} is the transverse momentum of particle i, y_i its rapidity and ϕ_i its azimuthal angle. The radius parameter R determines how large the cone of a jet can become, typical values range from 0.4 to 0.7.

The parameter p is crucial in the distance calculation. At first $p = 1$, the k_T algorithm [104], and $p = 0$, the Cambridge/Aachen algorithm [105], were used. While k_T tries to balance energy and distances to approximate inversion of the QCD branching processes, Cambridge/Aachen removes the $\min(\dots)$ factor from equation 6.3 and thus only takes the spatial separation into account.

Also negative values for p turn out to be useful and the algorithm with $p = -1$ was named anti-k_T [103]. Here the shapes of the jet cones are circles. If two jets are closer together than twice the jet cone size, the jets overlap and not both can obtain circular shape. If both jets have the same energy, the cones are cut by a straight line in the middle; if one jet is much more energetic than the other, the higher energetic jet will acquire the overlapping area almost completely. Due to the fact that soft particles

tend to cluster with hard particles instead of other soft particles, regular jet shapes are achieved and anti-k_T behaves considerably better than the other algorithms in events with lots of activity in the underlying event and under high pile-up conditions. A detailed description can be found in [103].

6.4.4 Jet types

When simulating events as described in chapter 5, one can choose the particles the jet algorithm should be applied to: the algorithm can not only run on the fully reconstructed particles after the detector simulation (reconstructed jets) but also on the generated hadrons after the hadronisation step but before the detector simulation (generator jets). The generator jets do not suffer from detector inefficiencies and can for example be used to calculate detector efficiencies or the deviation in the calculated jet energies. For data, there is obviously no way to measure without the detector and thus only reconstructed jets are available.

In addition to this, there are several ways to reconstruct the particles and CMS provides four different input collections for the reconstructed jets [106]. These are *calorimeter jets* which only use information from the energy deposits in the calorimeters, the *jet-plus-tracks* approach exploits CMS's excellent tracker (see 3.4.1) to correct energy and direction of the particles reconstructed in the calorimeters, the *track jets*, being only based on tracker information, are used for cross checks only, and finally the *particle flow jets* which use information from all detector components. In this analysis jets from the particle flow approach are used, it is explained in section 6.6.

6.4.5 Jet Energy Correction

Even though there can already be corrections to the 4-momentum of the particles used for jet clustering, the jet's 4-momentum p_μ also needs to be calibrated. The correction is done using a factor C which is the product of different corrections as shown in figure 6.3.

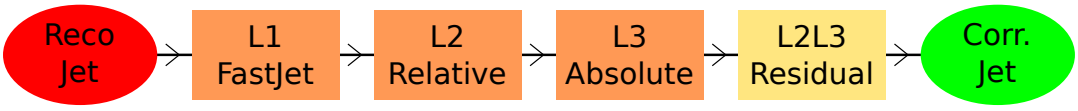


Figure 6.3: Different levels of jet energy corrections. The raw jets on the left are corrected sequentially from left to right where corrected jets are output. The residual corrections are only applied on real data, not on simulation.

These corrections can be expressed using the following formulae:

$$p_\mu^{\text{corrected}} = C \cdot p_\mu^{\text{raw}} \tag{6.6}$$

$$\text{where} \quad C = C_{\text{FastJet}}(p_T^{\text{raw}}) \cdot C_{\text{rel}}(\eta') \cdot C_{\text{abs}}(p_T'') \cdot \ldots \tag{6.7}$$

Using a factorised approach, the primed quantities correspond to the values after previous corrections, i.e. the N^{th} correction is applied on the jet corrected with the previous $N-1$ corrections. The sequence for the calculation of the correction factors is also shown in figure 6.3.

The first correction step is the L1-FastJet correction. It corrects for additional measured energy which does not belong to the hard process. Such energy contributions can for example stem from noise in the electronics or pile-up. To calculate the correction factor, very many ghost particles with infinitesimally small energies are added to the event, they do not influence the hard jets. Then the k_T algorithm with a cone of $R = 0.6$ is used to reconstruct jets. Now the median transverse hadronic activity per area (the p_T density ρ) in the event is determined from the ratio p_T^{jet}/A^{jet} of all jets. After this calculation, the corrected energy of each jet can be calculated by subtracting the part of the underlying hadronic activity corresponding to the jet size. This can be converted into a correction factor which by construction is always smaller than 1 so that the jet's momentum is reduced.

In the following step, the level 2 corrections, the dependency on the pseudorapidity is addressed. The goal of this correction is to achieve a jet response which is flat in η. This is done by selecting clean events with at least two hard jets which are back-to-back in the azimuthal direction ($\Delta\phi > 2.7$) where at least one jet is reconstructed in the barrel region $|\eta| < 1.3$. Additionally a veto on the energy fraction of a possible third jet is imposed [107] before the dijet balance is finally calculated assuming that the uniformity in the barrel region is much better than in the endcaps. Thus the jets are scaled to the energy that they would have in the barrel region. Many of these events are simulated and at the end an η dependent scale factor is derived. In contrast to the level 1 corrections, this scale factor is independent of the event.

In the level 3 correction, the absolute scale of the transverse jet momenta is determined. To do this, a p_T balancing is performed using many $\gamma/Z+$jets events. In case of a photon, the excellent resolution of the electromagnetic calorimeter is exploited to find the correction factor for the jet; in case of a Z, the event is only used if the Z decays into e^+e^- or $\mu^+\mu^-$. Then the tracker and in addition the electromagnetic calorimeter for the electrons or the muon detectors for the muons are used for the reconstruction and determination of the energy of the recoiling jet. The p_T dependent correction factor is then derived from the directly measured jet transverse momentum and the calculated value. Like the level 2 corrections this factor is independent of the event.

Detailed descriptions of the described jet corrections can be found in [63].

Level 2 and 3 corrections are purely determined from the simulation and therefore correct only the simulation to the best possible jet energies. Because the simulation is very good, these jet energy corrections can also be applied to data. However the data does not behave exactly like the simulation and therefore additional corrections are applied on the data only which correct for these small differences. These corrections are called *L2L3Residual* corrections [108]. They are usually of the order of 2 % in the barrel and can go to 10 % for larger $|\eta|$. Like the L2 and L3 corrections, they are also

determined using p_T balancing in dijet, γ+jet, and $(Z \to \mu\mu)$+jet events, but using real data instead of the simulation. The main advantages of using residual corrections on top of simulation based ones are that these corrections are rather small and technically easier to handle and in addition, the systematic uncertainty has proven to be smaller this way.

As shown in formula 6.7, further correction factors could simply be appended. Possible additions include making the jet response uniform with respect to the electromagnetic energy fraction (level 4), jet flavour corrections to account for differences between light flavour, heavy flavour, and gluon jets (level 5), and parton jet corrections which try to correct back to the parton level. These further corrections are currently only used in a few dedicated jet studies. These corrections are mainly important for jets which are only reconstructed in the calorimeters. With the introduction of particle flow (see section 6.6) they are usually not needed any more. This analysis is using particle flow, therefore the corrections up to level 3 (and L2L2Residual for data) are sufficient and remaining corrections are covered by the jet energy scale systematic error (see section 8.8).

6.5 Missing transverse energy

The incoming protons from the beam do not have any transverse momentum. Therefore the sum of the transverse energies of the initial state particles and due to energy conservation also the summed transverse energies of the final state particles are zero. Due to the non-measured neutrinos and detector resolution effects, there is a measured transverse momentum imbalance in the event, see also section 3.3.

The missing energy $E_{\mathrm{T}}^{\mathrm{miss}}$ can be calculated from particles reconstructed by different algorithms, for example by calorimeter information only or by calorimeter information corrected using measurements in the tracker.

In this analysis the particle flow approach (see next section, 6.6) is used to calculate $E_{\mathrm{T}}^{\mathrm{miss}}$ [109]. Therefore in $\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_i \vec{E}_T^i$ (equation 3.8), the index i runs over all objects, i.e. electrons, muons, and jets reconstructed by the particle flow algorithm.

6.6 Particle flow

The particle flow algorithm [110, 111] provides a *global event description*. Instead of reconstructing physics objects only in their dedicated detectors, the particle flow algorithm looks at the event in its entirety. For example, a muon is not only measured in the muon system but also in the tracking system [112]; a jet is not only measured in the calorimeters but its charged constituents can also be seen in the tracking system [109]. This way, higher reconstruction accuracies and efficiencies are achieved.

The exact behaviour of the event reconstruction might need to be analysis specific

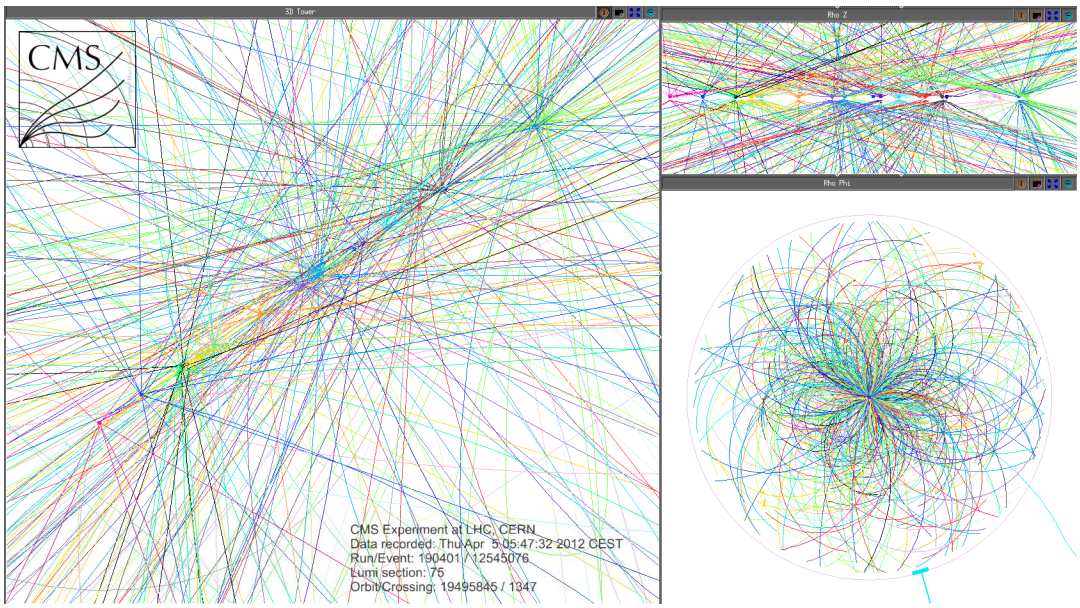


Figure 6.4: Event display of an event with 29 vertices [115].

and is therefore configurable (for example isolation cone sizes for jets and leptons). In this work, the configuration of the CMS particle flow algorithm is implemented using a common Hamburg University/DESY code [113] which simplifies the setup.

6.7 Pile-up removal

For the analysis of top quarks, only the objects coming from the top quark event need to be considered while other objects from a different interaction, i.e. a different vertex can be ignored. The vertex with the highest p_T sum of all tracks coming from this vertex is considered the interesting vertex. All the tracks from other vertices and their corresponding energy deposits in the calorimeters are projected out of the event so that following reconstruction steps do not take them into account any more. This procedure is called *Charged Hadron Subtraction* (implemented using `pfNoPileUp` [114]) because only particles with an electric charge can be seen in the tracking system and thus be assigned to a vertex.

Figure 6.4 shows a full event with multiple vertices. After the Charged Hadron Subtraction, one can treat it like an event with only one primary vertex.

6.7.1 Top projections

One problem in the reconstruction is that there can be ambiguities. For example, a reconstructed muon could be a jet constituent or a high energy deposit in the electro-

magnetic calorimeter could be caused by an electron or by a photon. In order to resolve these ambiguities, the PF2PAT [114] algorithm is used. In this work, the particles are reconstructed following a fixed order. As the muons are the only particles which can be detected in the muon system, we first reconstruct all muons. However, we are only interested in isolated leptons (see section 7.7) which come directly from the W boson (or an intermediate τ), while muons inside a jet (for example originating from a B hadron) are not used. Analogous to the pile-up removal (see section 6.7) all the detector information which is assigned to these isolated muons is then projected out of the event so that for example hits in the tracking system which were caused by the muon are no longer available for the reconstruction of other objects (`pfNoMuon`).

The next objects to be reconstructed are the electrons. Again only isolated electrons are considered in the analysis and thus their entries from the tracking system and from the electromagnetic calorimeter are removed (`pfNoElectron`).

Only after the isolated muons and electrons have been removed, jets can be reconstructed. This is done using the AntiKT algorithm (see section 6.4) with a cone of 0.5 in η-ϕ space ($\Delta\eta^2 + \Delta\phi^2 < 0.5^2$).

Finally, the missing transverse energy in the event can be calculated from the vectorial sum of the momenta of all reconstructed objects.

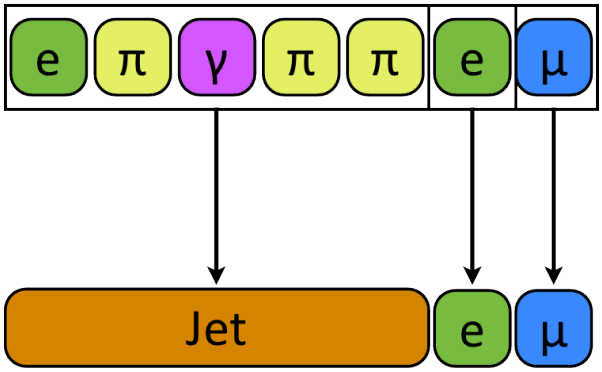


Figure 6.5: Reconstruction of particles in an event. From all seven reconstructed particles (top) the two leptons on the right are projected out as isolated leptons. The remaining five particles (left) then form a jet. (illustration from [114])

The steps are visualised in figure 6.5. All particles in the event are in the top row. The two rightmost particles, a muon and an electron, are isolated. Thus they are projected out of the event, indicated with the arrows to the bottom, where a list of all objects in kept. The remaining particles in the top row, including the electron on the left, are not isolated. They are clustered into jets.

6.8 b tagging

In many analyses it is useful to be able to distinguish jets coming from b quarks from jets coming from light quarks. The generic term for such an algorithm is b tagging. This is of specific interest in top quark analyses as the top quark almost always decays into a b quark, thus the event selection can make use of b tags. There are different tagging algorithms which all exploit the specific properties of these jets. The lifetime of B hadrons is of the order of 10^{-12} s (e.g. $\tau_{B^\pm} = (1.638 \pm 0.011) \times 10^{-12}$ s [12]) and therefore they can travel a measurable distance before they decay ($c\tau_{B^\pm} = 491$ µm; the actual distance $c\gamma\tau$ depends on the relativistic γ factor which depends on the mass and energy of the B hadron). Thus there will not only be tracks in the event pointing to the interaction point (primary vertex) but there will also be tracks coming from the point of the B decay (secondary vertex). The tracker (see 3.4.1) is precise enough to measure this.

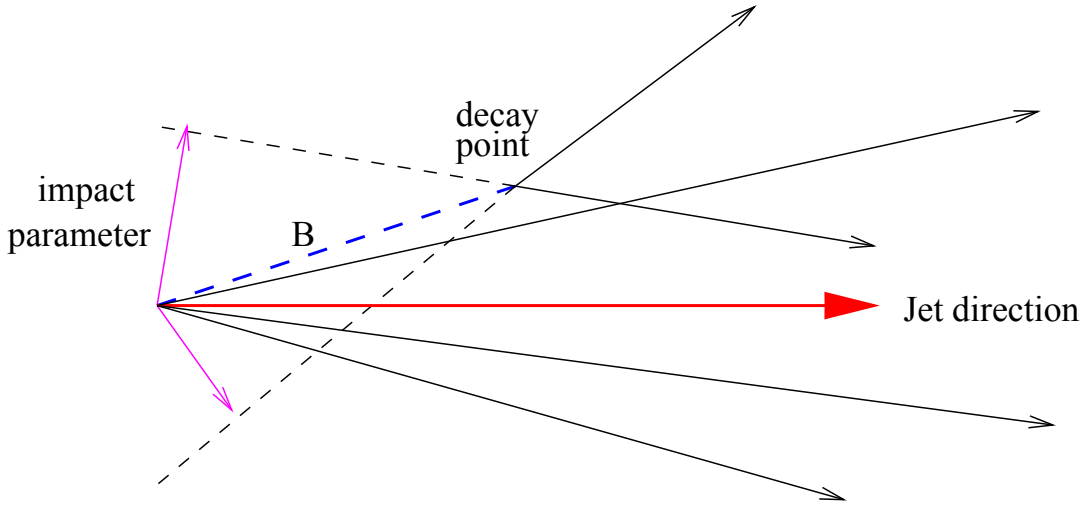


Figure 6.6: An event with a secondary vertex [116] and its impact parameter.

Figure 6.6 shows a schematic sketch of an event with a secondary vertex. The primary vertex is on the left side of the plot, several tracks are originating from it. Also an electrically neutral B hadron, shown with the blue dashed line, is produced in this example – it thus does not produce hits in the tracking system. In this example, it decays into two particles with a reconstructed track. By extending the track one can finally calculate the distance of the track to the primary vertex, a value which is called impact parameter.

The *Track Counting* algorithm is based on the impact parameters significance of all good tracks in an jet which are then ordered by decreasing significance. Two variants of the algorithm are available: a *High Efficiency* variant requiring the second most significant impact parameter to be above a certain threshold, and a *High Purity* variant

where the third most significant impact parameter is considered.

In CMS, a *loose*, a *medium*, and a *tight* working point is defined for all b tagging algorithms. The working points are defined by their mistag rates of 10 %, 1 %, and 0.1 % respectively. For this analysis the Track Counting High Efficiency (TCHE) algorithm at the loose working point is used to maximise the selection efficiency. It corresponds to an impact parameter significance cut of 1.7.

Despite being rather simple, the track counting algorithm was working well in the analysed dataset. Other algorithms with different properties, for example the Simple Secondary Vertex (SSV) algorithm, worked equally fine. SSV first tries to find a secondary vertex and then relates it to the primary vertex. The choice of TCHE was based on previous dilepton cross section measurements, e.g. [91, 117].

Meanwhile a sophisticated Combined Secondary Vertex tagging algorithm is available which takes into account the impact parameter significance, the secondary vertex, and jet kinematics. It is not only slightly more efficient then TCHE or SSV but also more robust in events with high pile-up, it is used in the update of the analysis (chapter 10).

Chapter 7

Event selection

The event selection used to measure differential cross sections is described in this section. At first, the data source is selected and badly reconstructed events are rejected. Then it is made sure that the correct triggers have fired. Cuts on special properties of top quark events are finally applied to end up with a high purity of events with a top quark pair which decays in the dilepton channel.

This event selection is based on [91] ($\mu\mu$ channel) and [117] ($e\mu$ channel). The analysis code of these two analyses has been combined, simplified and extended to also cover the ee channel. While many of the selection cuts are similar to those presented in [91, 117], many details from the pile-up treatment over the used triggers to the leptons selection have changed (for example PF2PAT and top projections, see section 6.7.1, have replaced a $\Delta R(\ell, j)$ lepton veto).

7.1 Data source

The data produced by the CMS detector is recorded after triggering (see section 7.4) and stored in different data streams so that one can already choose events with specific features at a very early stage. For analysis of top quark events decaying into a final state with two leptons, the streams with two leptons can be used. There are three such streams which are used in this work, `/DoubleElectron` for the ee channel, `/DoubleMu` for the $\mu\mu$ channel, and `/MuEG` for the $e\mu$ channel.

7.2 Good runs

From these data streams, certified runs recorded before LHC's technical stop on June 28th, 2011 are selected by applying the certified good-run lists as given in [118] for the re-reconstruction of the first $214.8\,\mathrm{pb}^{-1}$ of data and in [119] for the prompt reconstruction for the rest of the analysed data ($927.9\,\mathrm{pb}^{-1}$). The total integrated luminosity of the complete data sample is thus $1.14\,\mathrm{fb}^{-1}$. The run ranges and exact sample names are shown in table 7.1.

These good-run lists are created centrally by the CMS collaboration in DQM (see 3.6) shifts, they contain a list of all runs in which all components of the CMS detector were switched on and working correctly.

Data samples	Run range
/DoubleMu/Run2011A-May10ReReco-v1/AOD	160431–163869
/DoubleMu/Run2011A-PromptReco-v4/AOD	165088–167913
/DoubleElectron/Run2011A-May10ReReco-v1/AOD	160431–163869
/DoubleElectron/Run2011A-PromptReco-v4/AOD	165088–167913
/MuEG/Run2011A-May10ReReco-v1/AOD	160431–163869
/MuEG/Run2011A-PromptReco-v4/AOD	165088–167913

Table 7.1: Collision data samples and corresponding run ranges used in the analysis. The gap between runs 163869 and 165088 is due to a technical stop.

7.3 Event cleaning

Only clean collision events can be used for data analysis. There are three different filters applied on the data to achieve this.

The first filter removes *beam scraping* events, i.e. events which are recorded due to background from the beam. In the detector, these events have many reconstructed tracks which have a low purity. Thus the filter accepts all events with less than 10 reconstructed tracks regardless of their quality and all events with at least 10 tracks only when at least one quarter of the tracks has a high track quality. In other words, events with at least 10 tracks and at least three quarters of badly reconstructed tracks are rejected.

The second filter requires that there is a *primary vertex* in the event which fulfils some quality criteria. Also it must lie in the centre of the pixel detector, i.e. within $\pm 15\,\mathrm{cm}$ in z direction from the beam spot and within less than $2\,\mathrm{cm}$ in the ρ direction.

The third event cleaning filter removes events in which there is *anomalous noise* in the hadronic calorimeter[1]. This noise is caused by instrumentation issues with the Hybrid Photo Diodes and the Readout Boxes and can be suppressed for example by looking at the timing, the pulse shape of the signal or its correlation with deposits in the electromagnetic calorimeter.

7.4 Trigger

In order to filter out events with top quarks from the large amount of data recorded by the LHC the data are preselected by triggers (see section 3.5). The decay channels

[1] https://twiki.cern.ch/twiki/bin/view/CMS/HcalNoiseInfoLibrary

described in this thesis all contain at least two isolated leptons, which is represented in the triggers. For each channel the trigger is generally chosen such that it is unprescaled and its threshold is well below the cuts applied in further selection steps. During the data taking period the triggers are subject to change, i.e. their thresholds need to increase with an increasing instantaneous luminosity to limit the total output rate. Also triggers might not be fully optimised in the beginning, thus there can be several versions of a trigger with the same thresholds. A list of the triggers used is given in table 7.2 where the "*" is a wildcard for the different versions.

Sample	Channel	Trigger
same for data and simulation	$\mu^{\pm}e^{\mp}$	HLT_Mu8_Ele17_CaloIdL_* or HLT_Mu17_Ele8_CaloIdL_*
simulation and runs 160431–163869	$\mu^{+}\mu^{-}$	HLT_DoubleMu7_*
runs 165088–167913		HLT_Mu13_Mu8_*
same for data and simulation	$e^{+}e^{-}$	Ele17_(CaloId-Iso)_Ele8_(CaloId-Iso) or Ele17_SW_TightCaloEleId_Ele8_HE_L1R_*

Table 7.2: Triggers used in data and in simulation. An asterisk is used as a wildcard to include different versions of the triggers.

For the $e\mu$ channel, the trigger requires a muon with a transverse momentum of at least 8 GeV and an electron with a transverse momentum of at least 17 GeV or vice versa. On top of that the electron needs to fulfil loose identification criteria in the electromagnetic calorimeter.

In the ee channel, the trigger requires two electrons with transverse momenta greater 17 GeV and 8 GeV respectively. Also here some electron identification criteria are applied. Additionally the electrons are required to be loosely isolated in the calorimeter (see section 6.3). A second trigger with the same momentum thresholds but different isolation and identification criteria is used to maximise the efficiency.

In the $\mu\mu$ channel, two muons are required by the trigger. In the simulation and in the first data taking period (first $214.8\,\mathrm{pb}^{-1}$), both muons have to have a transverse momentum of at least 7 GeV. In the later data taking periods ($927.9\,\mathrm{pb}^{-1}$) this trigger was strongly prescaled (3.5.1) and thus replaced by a trigger with higher and asymmetric transverse momentum thresholds of 13 GeV for the first and 7 GeV for the second muon.

7.5 Simulation

The simulation used in this work is based on the `Summer11` simulation production (and `Fall10` where `Summer11` was not available). The signal sample was created using MadGraph interfaced with Pythia for hadronisation and parton fragmentation and is nor-

malised to the measured inclusive $t\bar{t}$ cross section in the dilepton channel [120]:

$$\sigma_{t\bar{t}} = 169.9 \pm 3.9\,(\text{stat.}) \pm 16.3\,(\text{syst.})\ \text{pb} \tag{7.1}$$

As an alternative, the signal process is also simulated with the POWHEG and the MC@NLO generators.

The cross sections assigned to other processes are taken from other calculations [121]. Apart from the signal, W + jets is also simulated using MADGRAPH (also interfaced with PYTHIA). Single top and antitop productions are simulated with POWHEG.

There are several Z/γ^* (Drell-Yan) simulations which are divided into different dilepton mass ranges. In the low mass region between 10 and 20 GeV three dedicated PYTHIA samples are used, one per decay channel. In the region between 20 and 50 GeV also three PYTHIA samples are used. These cover the whole dilepton mass range from 20 GeV to infinity, however they are cut off on generator level in the analysis at 50 GeV. For higher dilepton masses, MADGRAPH (interfaced with PYTHIA) is used because it describes the data better, especially at higher jet multiplicities. In the low mass region no MADGRAPH simulation was available, thus the switch to PYTHIA.

All other samples are simulated using PYTHIA as well. These are the diboson samples (WW, WZ, and ZZ) and the QCD multijet background simulation. The latter is split into one muon enriched sample which is used in the $\mu\mu$ and $e\mu$ channel, and six electron enriched samples which consist of three different p_T ranges split into light and heavy flavour.

Table 7.3 shows all samples and corresponding cross sections.

Unfortunately in the interfacing of MADGRAPH with PYTHIA a bug was present during the production which lead to a violation of energy and momentum conservation. This PYTHIA bug happened when hadronising events produced by an external matrix element program and passing the information via LHE files. The effect was tiny for Q^2 ordered PYTHIA tunes (e.g. D6T, not used here) but significant for p_T ordered showers (e.g. Z2 tune, used in this work). Even there, the violation was small for most of the events, but in about 2 % of the events the effect was larger than 5 GeV and those events had to be rejected for this analysis. The problem was originally discovered in $Z \to \ell\ell$ events where the energy of the final state leptons was 150 GeV larger than the centre-of-mass energy $\sqrt{s} = 7$ TeV [122, 123].

7.6 Pile-up removal

To be less affected by the effects of pile-up, only the hardest primary vertex is used, i.e. the vertex in which the sum of the track momenta coming from this vertex is maximised. All contributions in the detectors which can be assigned to a different vertex are removed from the event. This way, even events with 30 vertices can be analysed as if only one vertex was present. This step is described in detail in section 6.7.

Simulation sample	σ [pb]
TTJets_TuneZ2_7TeV-madgraph-tauola	169.9
TTTo2L2Nu2B_7TeV-powheg-pythia6	169.9
T_TuneZ2_tW-channel-DR_7TeV-powheg-tauola	7.87
Tbar_TuneZ2_tW-channel-DR_7TeV-powheg-tauola	7.87
WWTo2L2Nu_TuneZ2_7TeV_pythia6_tauola	4.51
WZTo3LNu_TuneZ2_7TeV_pythia6_tauola	0.61
ZZ_TuneZ2_7TeV_pythia6_tauola	7.4
WJetsToLNu_TuneZ2_7TeV-madgraph-tauola	31314
DYToEE_M-10To20_TuneZ2_7TeV-pythia6	3457
DYToMuMu_M-10To20_TuneZ2_7TeV-pythia6	3457
DYToTauTau_M-10To20_TuneZ2_7TeV-pythia6-tauola	3457
DYToEE_M-20_TuneZ2_7TeV-pythia6	1666
DYToMuMu_M-20_TuneZ2_7TeV-pythia6	1666
DYToTauTau_M-20_TuneZ2_7TeV-pythia6-tauola	1666
DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola	3048
QCD_Pt-20_MuEnrichedPt-15_TuneZ2_7TeV-pythia6	84679
QCD_Pt-20to30_EMEnriched_TuneZ2_7TeV-pythia6	2.50×10^6
QCD_Pt-30to80_EMEnriched_TuneZ2_7TeV-pythia	3.63×10^6
QCD_Pt-80to170_EMEnriched_TuneZ2_7TeV-pythia6	143×10^3
QCD_Pt-20to30_BCtoE_TuneZ2_7TeV-pythia6	139×10^3
QCD_Pt-30to80_BCtoE_TuneZ2_7TeV-pythia6	144×10^3
QCD_Pt-80to170_BCtoE_TuneZ2_7TeV-pythia6	9431

Table 7.3: Summary of simulated data samples used in this analysis. Also the cross section of the process is given.

7.7 Lepton selection

After all the mainly technical cuts and implementation details described in the previous sections, we can now look at the special properties of top quark events.

In this thesis only those top quark pairs are considered as signal where both W bosons decay into final states with a muon or electron (intermediate taus are allowed). All other events with a top quark pair are called *other* $t\bar{t}$.

It seems natural to start with the selection of the two isolated leptons because these have the most distinctive power. In addition, the datasets used already require two leptons on trigger level. To be in the high efficiency region of the trigger (see section 8.3 for trigger scale factors), the thresholds applied on analysis level need to be higher than the ones on trigger level. Thus a minimum transverse momentum of 20 GeV is requested and the leptons have be be reconstructed within $|\eta| < 2.4$. These cuts are of course applied in addition to the quality cuts described in section 6.1 and 6.2.

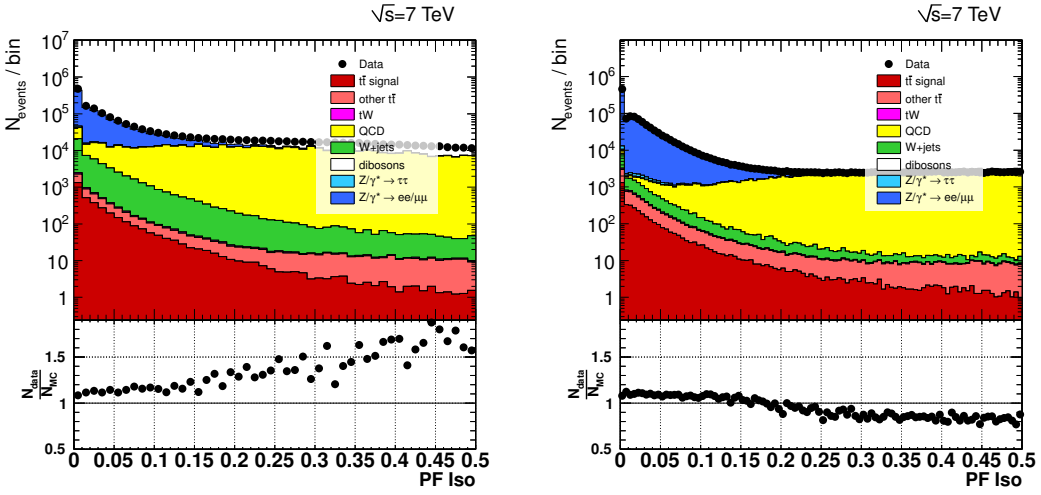


Figure 7.1: Isolation of the leptons: electrons in the ee channel on the left, muons in the $\mu\mu$ channel on the right. The applied cuts are $I_{rel}^{PF} < 0.17$ for electrons and $I_{rel}^{PF} < 0.20$ for muons.

As the leptons from the W decays are isolated, this is also requested for the reconstructed leptons. Figure 7.1 shows the isolation I_{rel}^{PF} for the electrons (left) and the muons (right), calculated using a cone size of $\Delta R = 0.3$. One can see that leptons from QCD multijet background have high isolation values while the top signal but also the leptons from Drell-Yan events have low values. To reject the QCD events, only those electrons (muons) where the isolation is smaller than 0.17 (0.20) are kept.

Top projections are used in the particle flow algorithm, i.e. leptons fulfilling all mentioned cuts including the isolation are now projected out of the event before the jets are reconstructed (see also section 6.7.1).

7.7.1 Disambiguation of events with more than two leptons

In events with more than two reconstructed isolated leptons it must be decided which two should be considered the leptons from the top decay. As an example, an event might have a μ^+ with 60 GeV, a μ^- with 30 GeV and an e^- with 40 GeV transverse momentum. Thus we have three combinations for the lepton pair in the event: $\mu^+\mu^-$, μ^+e^-, and μ^-e^-. The μ^-e^- combination can be rejected because the leptons need to have opposite charge. From the remaining dilepton combinations, the pair with the largest sum of their transverse momenta is chosen, in this example the 60 GeV μ^+ and the 40 GeV e^-. This way of disambiguation also avoids double counting of events which appear in data streams (see 7.1) for more than one channel. This is because an event is only accepted if it originates from that data stream which corresponds to the identified lepton pair. This means that the given example event would be counted only in the

MuEG stream but be rejected in the DoubleMu (and DoubleElectron) stream.

While the case of three isolated leptons in an event is rare, the decision to choose the highest sum of transverse momenta is motivated by the origin of the leptons. Leptons from the W boson from a top quark tend to have higher energies than leptons from background processes.

7.8 Dilepton mass veto

A large number of dilepton events come from leptonically decaying low mass resonances such as J/ψ or the Y. These events have a low invariant lepton pair mass while this isn't the case for top events. For this reason, events with a dilepton mass of less than 12 GeV are rejected.

Figure 7.2 (top) shows the spectrum of the dilepton pair mass in the ee channel and $\mu\mu$ channel after the low mass cut. Clearly visible is the peak around 91 GeV coming from the Z boson decay products. The background events from this resonance need to be rejected. An exclusion window of ± 15 GeV around the Z mass, i.e. 76 to 106 GeV was chosen in this analysis.

In addition it can be seen that the agreement between data and simulation is not good below ≈ 40 GeV. This is caused by the different simulation (Pythia vs. MadGraph) and covered in a dedicated systematic uncertainty (see section 8.5.2).

In the $e\mu$ channel the contribution of Drell-Yan events is much lower because Z/γ cannot decay directly into two different lepton flavours. The only way to get an electron and a muon is the decay into two τ leptons where one of them decays further into an electron while the other decays into a muon. Because the decay of the taus into muons or electrons each has two additional neutrinos, the invariant mass of the $e\mu$ pair is smaller than the Z mass—it is in the range between 40 and 85 GeV. Figure 7.2 (bottom) shows this behaviour. In this channel, no cut on the dilepton mass is applied to exclude these events.

7.9 Jet selection

The top quark events also contain at least the two jets from the b quarks from the top decay. In Drell-Yan to $\ell\ell$ events, there are to first order only leptons an no jets in the final state, thus requiring two jets provides a good distinction. Of course, additional jets can be present in the events due to initial state and final state radiation (ISR and FSR), however they are usually softer than the ones from the b quarks from top decays.

Figure 7.3 shows the multiplicity of all jets after the previously described dilepton selection cuts. In this analysis, jets are required to have a transverse momentum p_T greater than 30 GeV, reconstructed with the particle flow algorithm using the anti-k_T jet clustering algorithm with a cone of $\Delta R = 0.5$ (see 6.4) and jet energy corrections

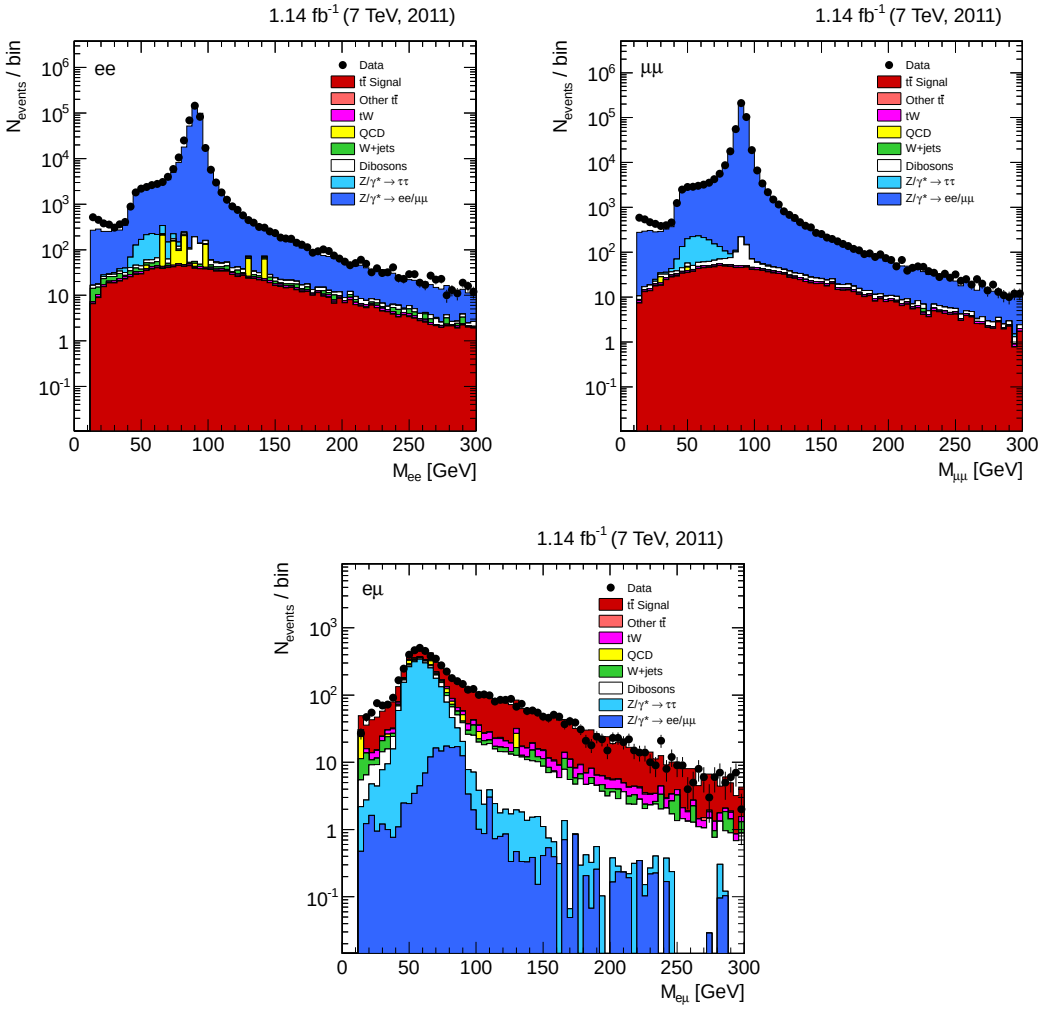


Figure 7.2: Invariant dilepton mass with the Z peak around 91 GeV in the ee channel (top left) and $\mu\mu$ channel (top right). The bottom plot shows the invariant dilepton mass in the $e\mu$ channel. The main background comes from a Z boson decaying into two τ leptons which further decay into an electron and a muon. In the ee and $\mu\mu$ plots the order of the entries of the stacked histogram has been reversed with respect to the legend because of the large number of Drell-Yan events.

applied (see 6.4.5). In addition they have to be detected in the central detector ($|\eta| < 2.4$) and fulfil "loose" jet identification cuts. These require that a jet consists of at least two constituents and that the energy deposited in the electromagnetic and hadronic calorimeter is each less than 99 % of the total jet energy.

As expected, one can see that the 0 and 1 jet bins are dominated by Drell-Yan background, especially in the ee and $\mu\mu$ channel while the top quark signal is only visible beginning in the 2 jet bin. In the $e\mu$ channel the background is so low that even the 1 jet bin contains about 43 % signal events (see table 7.5). In this bin, one of the two b jets was not reconstructed. This is the reason that nevertheless only those events with at least two jets are selected because two reconstructed jets need to be available for the kinematic event reconstruction (see 7.12).

What one can also see is that the agreement of data and simulation is still good up to high jet multiplicities. Also the QCD and W + jets backgrounds have been removed almost completely.

7.10 Missing energy selection

As the top quarks decay into W bosons and these further decay into a lepton and a neutrino, there is missing transverse energy in the event because the neutrinos cannot be detected in the CMS detector. This is a further distinction to Drell-Yan events because those do not have neutrinos and thus no missing transverse energy ($Z \to \nu\bar{\nu}$ events of course have missing energy, but they are lacking the two leptons).

The distribution of the missing energy is shown in figure 7.4. At this point already 88 % of the $e\mu$ events are from the signal process, so no cut is needed in this channel.

In the ee and $\mu\mu$ channel the situation is different. Only 35 % of the events come from top quarks pairs, the majority are still Drell-Yan events. One can see that those events have low missing energy, while the top events have larger missing energy. Thus a cut requiring $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV is introduced after which 78 % of the remaining events are from the signal.

Another effect is that the simulation doesn't describe the data well in the low $E_{\mathrm{T}}^{\mathrm{miss}}$ region. By cutting this region away, not only Drell-Yan background events are removed but also the overall agreement between data and simulation is improved.

7.11 b tag selection

The next selection step exploits that the jets are coming from b quarks. As described in section 6.8, these jets can be distinguished from light quark jets using the long lifetime of the B hadrons. For this analysis the Track Counting High Efficiency algorithm is used at the loose (10 % mistag rate) working point.

The reason to allow this rate of mistags is that the sample is already signal dominated

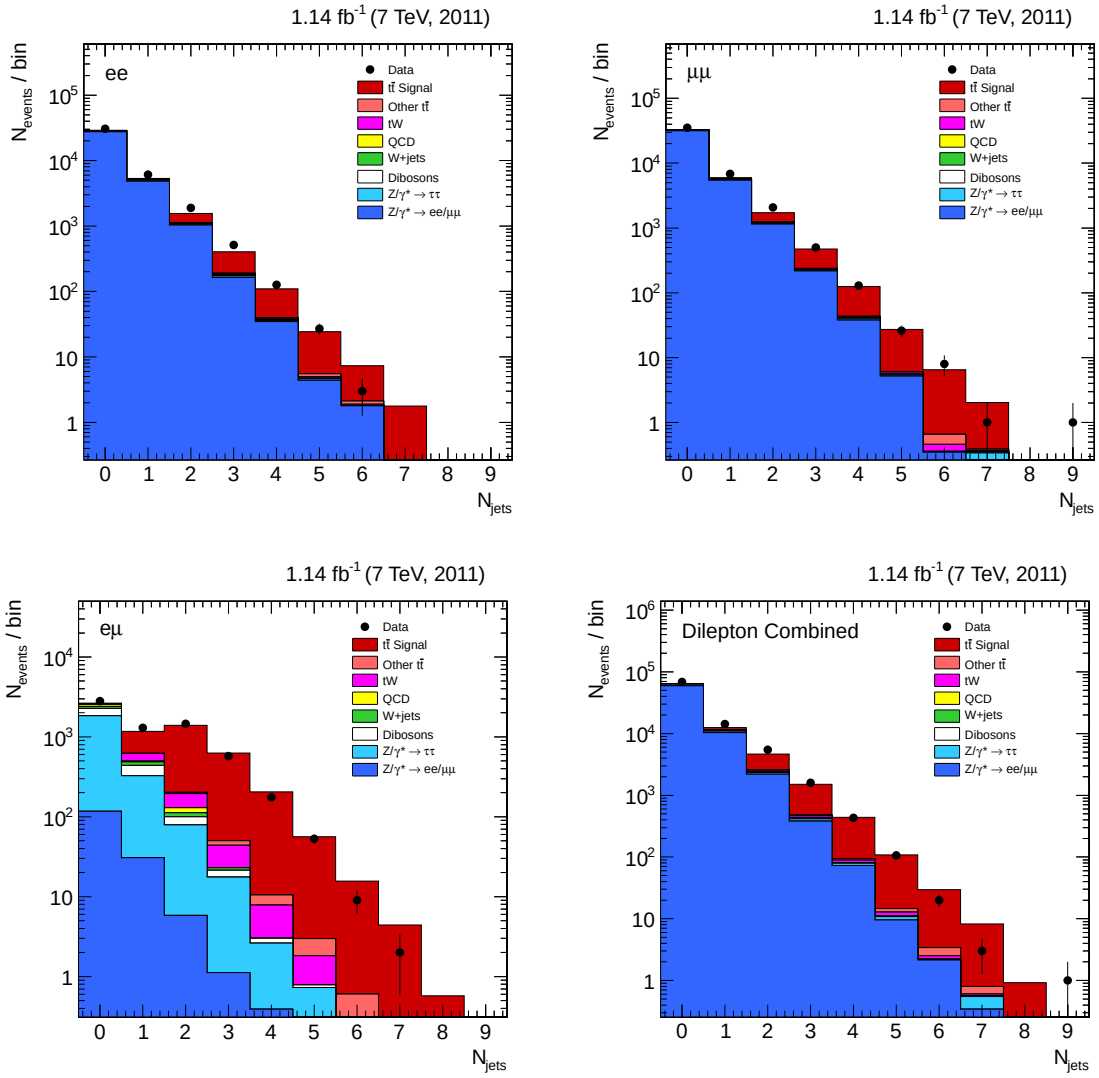


Figure 7.3: Jet multiplicity after dilepton selection. The top row shows the ee channel and $\mu\mu$ channel, the bottom row shows the $e\mu$ channel and the combination of all channels. Lepton correction factors (see section 8.4) have been applied.

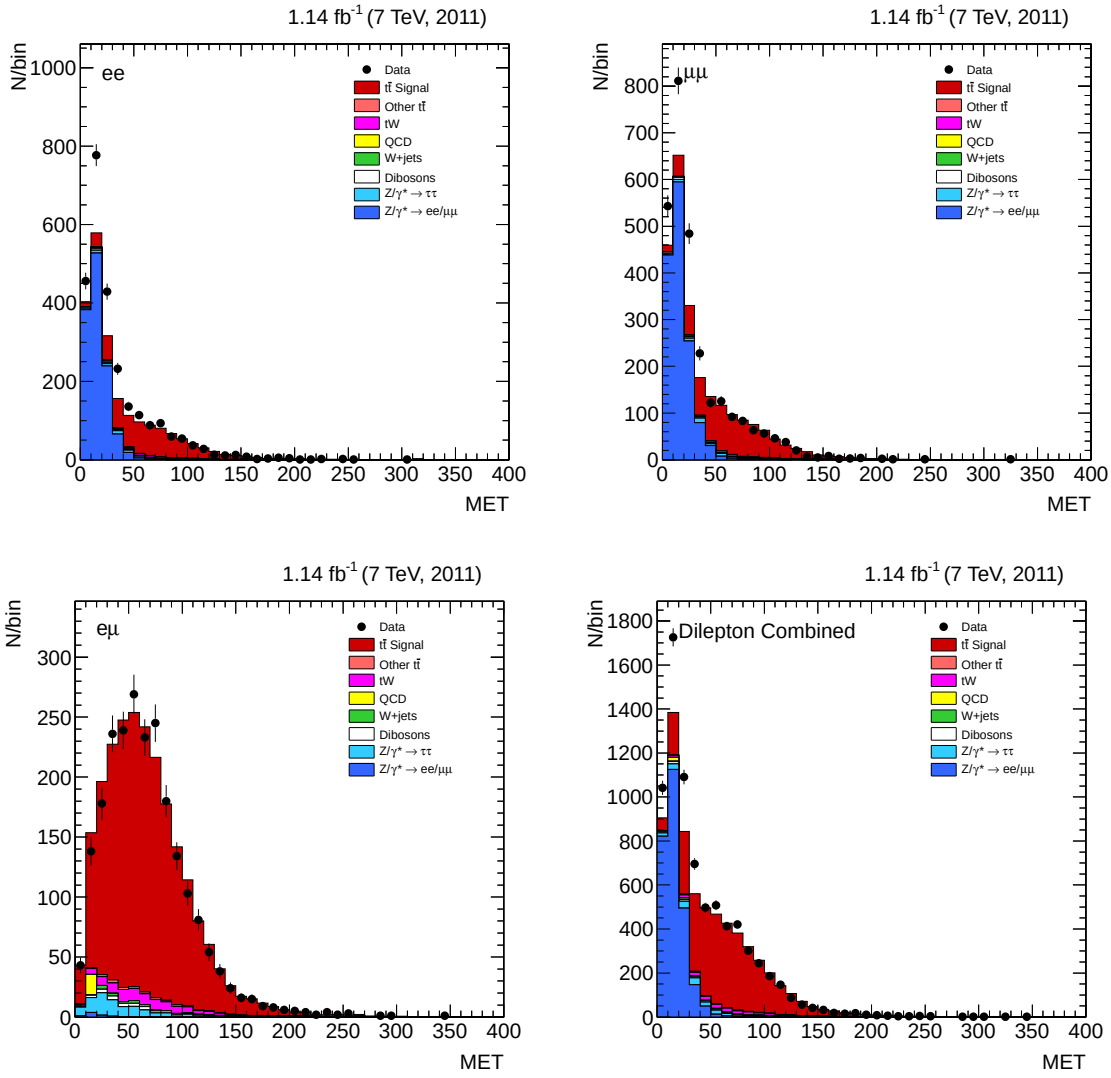


Figure 7.4: Missing transverse energy after selection of two jets. The top row shows the ee and $\mu\mu$ channel, the bottom row shows the $e\mu$ channel and the combination of all channels.

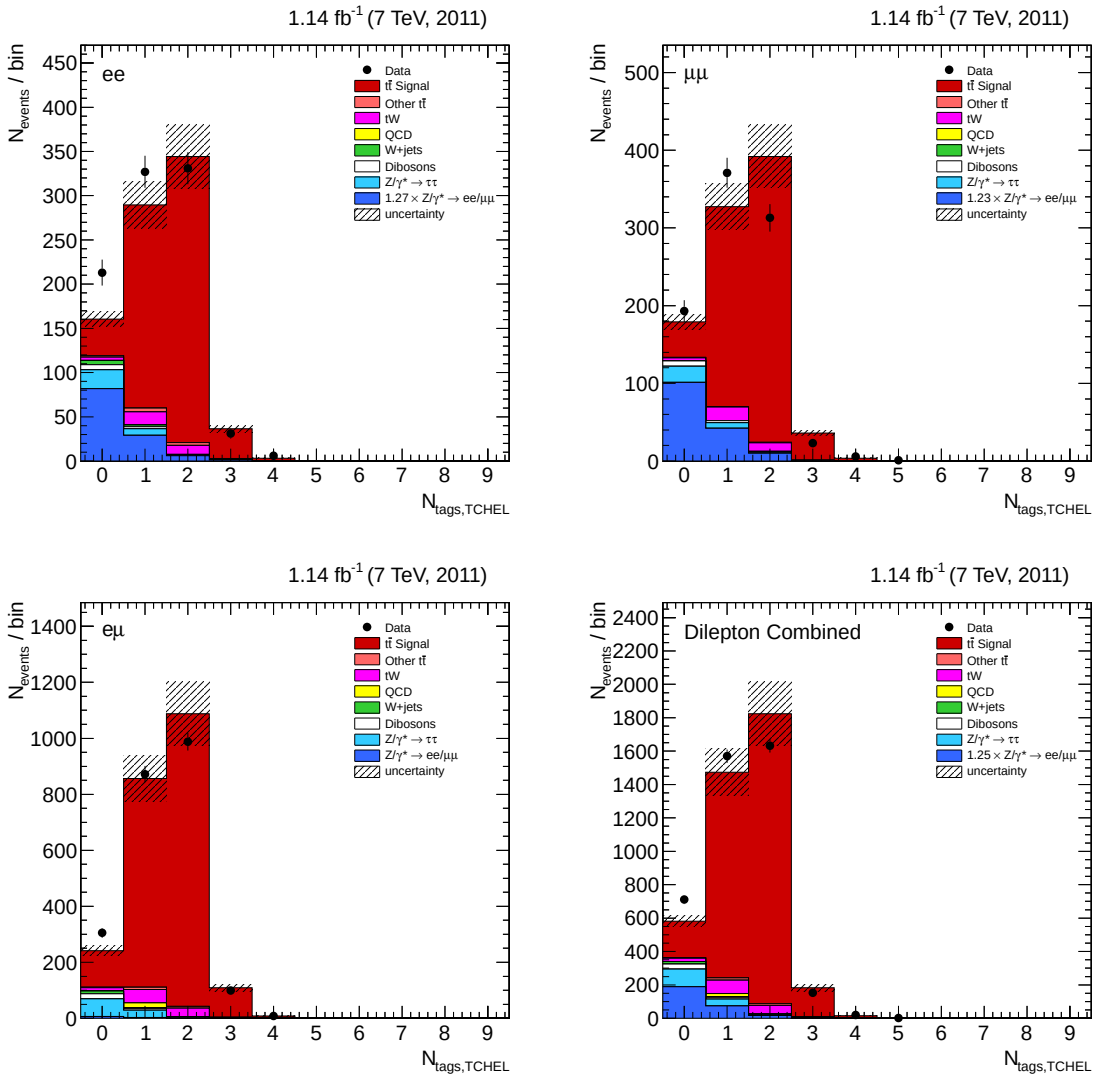


Figure 7.5: Multiplicity of the b tags. The top row shows the ee and $\mu\mu$ channel, the bottom row shows the $e\mu$ channel and the combination of all channels.

or in the $e\mu$ channel almost background free and thus a tighter tagging would just remove too many signal events. As one can see in figure 7.5 the bin with 0 tagged jets contains most of the Drell-Yan background but not many signal events.

In principle there should be two b tagged jets in event. However there are many signal events where only one jet was tagged while the other wasn't, for example due to inefficiencies in the tagging algorithm. As the rate of background events in the one tag bin is only ~ 20 % ($\sim$10 % in $e\mu$), it is beneficial for this measurement to request only one b tag.

In the $e\mu$ channel this requirement is not needed at all, but it is included anyway to have a consistent b tag selection in the three channels. This is used later on in the kinematic event reconstruction.

7.12 Kinematic event reconstruction

7.12.1 Event reconstruction as a cut

The last selection step is the kinematic event reconstruction. In contrast to the previous steps, its main purpose is not to increase the signal purity but to reconstruct the full kinematics of the top quarks themselves before the decay. The reconstruction does not work in all cases and thus as a side effect, the amount of rejected events from the background and the signal are different.

For the inclusive $\mathrm{t\bar{t}}$ cross section measurement, studies have been made to check if the event reconstruction can be used for background rejection [91, 117]. In the $\mu\mu$ channel the event reconstruction was used as an alternative to b tagging which was not well understood at that time and thus lead to a larger systematic uncertainty. Without b tagging, the kinematic event reconstruction rejects twice more Drell-Yan events than signal events. In the $e\mu$ channel on the other hand the significance of the signal was even decreased by the event reconstruction. Also this study was done without b tagging.

After b tagging the reconstruction decreases the significance in both channels and is thus not used in recent inclusive cross section measurements [120].

7.12.2 Reconstruction

In dilepton events, the final states (4-momenta) of the two leptons and the two b tagged jets can be measured but the two neutrinos escape undetected and thus only the sum of their x and y momenta is known. Additional constraints are the masses of the W

bosons and the equality of top and anti-top quark masses.

$$\vec{E}^{\text{miss}} = \begin{pmatrix} E_x^{\text{miss}} \\ E_y^{\text{miss}} \\ \text{not measured} \end{pmatrix} = \begin{pmatrix} p_x(\nu) + p_x(\bar{\nu}) \\ p_y(\nu) + p_y(\bar{\nu}) \\ p_z(\nu) + p_z(\bar{\nu}) \end{pmatrix} \tag{7.2}$$

$$m(W^+) = m(W^-) = 80.4 \text{ GeV} \tag{7.3}$$

$$m(t) = m(\bar{t}) \tag{7.4}$$

As the neutrinos have a very low mass of less than 1 eV [12] while their momenta are in the GeV range, one can treat them as massless:

$$E^2(\nu) = p_x^2(\nu) + p_y^2(\nu) + p_z^2(\nu) \tag{7.5}$$

$$E^2(\bar{\nu}) = p_x^2(\bar{\nu}) + p_y^2(\bar{\nu}) + p_z^2(\bar{\nu}) \tag{7.6}$$

Furthermore we know that the W^+ decays into an anti-lepton and a neutrino. Using the 4-momentum p we get:

$$m^2(W^+) = p^2(W^+) = \big(p(\bar{\ell}) + p(\nu)\big)^2 \tag{7.7}$$

$$\text{and} \quad m^2(W^-) = p^2(W^-) = \big(p(\ell) + p(\bar{\nu})\big)^2 \tag{7.8}$$

Finally we know that the top quark decays into a W boson and a b quark. Again one can calculate the mass from the decay products:

$$m^2(t) = \big(p(b) + p(W^+)\big)^2 = \big(p(b) + p(\bar{\ell}) + p(\nu)\big)^2 \tag{7.9}$$

$$\text{and} \quad m^2(\bar{t}) = \big(p(\bar{b}) + p(W^-)\big)^2 = \big(p(\bar{b}) + p(\ell) + p(\bar{\nu})\big)^2 \tag{7.10}$$

The latter two equations (7.9 and 7.10) are the ones which need to be solved. The problem is that the six neutrino momenta are unknown while there are only 5 boundary conditions (each 2 in equation 7.2 and 7.3 and 1 in 7.4).

To simplify this system of equations, approximations based on the large mass differences between leptons, b quarks and top quarks are made:

$$m^2(\ell) \ll m^2(b) \tag{7.11}$$

$$\text{and} \quad m^2(b) \ll m^2(t) \tag{7.12}$$

Finally one can transform these equations into a polynomial equation of fourth order in any one of the neutrino components, for example in $p_x(\nu)$. It depends on m_t and the measured quantities $p_{l^+}, p_{l^-}, p_b, p_{\bar{b}}$ [91]:

$$0 = \sum_{i=0}^{4} c_i(m_t, p_{l^+}, p_{l^-}, p_b, p_{\bar{b}}) p_x^i(\nu) \tag{7.13}$$

Assuming the top quark mass is known, m_t can be fixed to this value and the equation can be solved analytically. As it is of fourth order, there can be from zero up to four meaningful solutions. For this work, a scan over the top quark mass is performed by varying m_t in 1 GeV steps from 100 to 300 GeV. This is done even though the top quark mass is known to be 173.5 GeV [13] to account for mismeasurements and increase the number of events which have at least one meaningful solution (see also 8.7).

The described method is similar to the MWT method described in [124].

7.12.3 Selection of the solution

The different possible solutions are not the only place where an ambiguity is introduced. There is also a problem with the jets in the event. By b tagging one can only identify jets from a b *or* a $\bar{b}$ quark but it doesn't distinguish between them. Thus equation 7.13 has to be solved for both permutations of the jets.

It gets even worse because only one jet is required to be b tagged to improve the total number of signal events (see section 7.11). If a third jet (or even more) exists in the event, all of these can originate from the second b quark. Thus there is even more combinatorics for the jets.

In the end, this method can thus end up with a larger number of solutions for the neutrino momentum and it has to be decided how to proceed. To do this, in [91] PYTHIA has been used to simulate 10000 dilepton top pair events. The resulting neutrino spectrum was fitted using an unbinned maximum likelihood fit with a two-dimensional Landau distribution as shown in figure 7.6.

The weights assigned to each solution represent the probability calculated from the Landau distribution, i.e. a solution with $E_\nu = E_{\bar{\nu}} = 58$ GeV will be preferred over a solution further away from the maximum.

The assigned weight is only the second criterion, the first one is the b tag information. If solutions with two b tagged jets exist, these are preferred over solutions with only one b tagged jet regardless of the weight. From all solutions with two tagged jets, the one with the highest weight is taken. If no solutions with two tags exist, the solution with one tag and highest weight is taken.

7.13 Selection summary

After the complete event selection the original goal to end with a highly signal dominated dataset has been reached successfully. The signal fraction is 92.8 % in the $e\mu$, 90.6 % in the $\mu\mu$ and 90.4 % in the ee channel. The cuts used to achieve this high purity are summarised in table 7.4 and the resulting event yields after several different selection steps can be found in table 7.5.

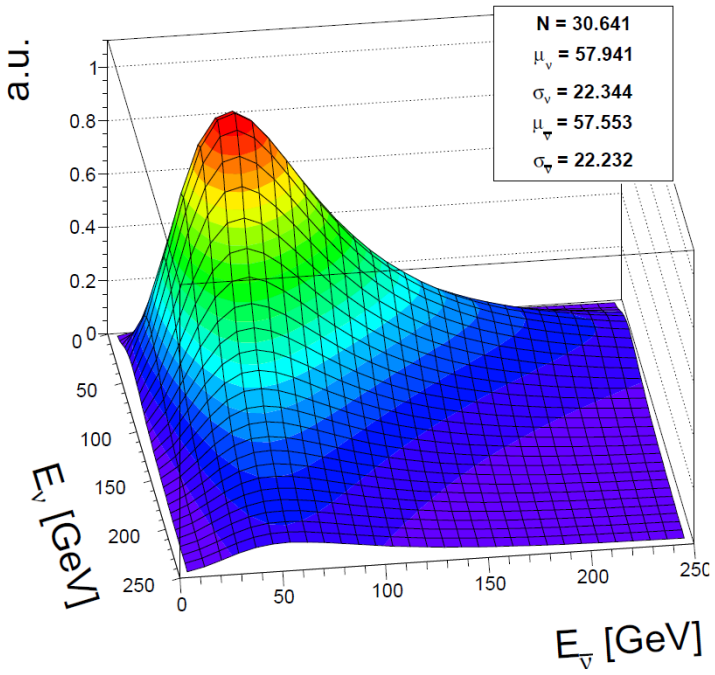


Figure 7.6: Fit to a simulated neutrino spectrum generated with PYTHIA [91]. The most probable neutrino energies, $\mu_{\nu/\bar{\nu}}$, are ~ 58 GeV for both neutrinos; the width σ is ~ 22 GeV, and N is a normalisation factor irrelevant for this study.

Channel	ee	$e\mu$	$\mu\mu$
Trigger	$p_T^{e1(e2)} > 17(8)$ GeV isolation & id	$p_T^{\ell 1(\ell 2)} > 17(8)$ GeV electron id	$p_T^{\mu 1(\mu 2)} > 7(7)$ GeV or $p_T^{\mu 1(\mu 2)} > 13(8)$ GeV
Cleaning	no HCAL noise, no beam scraping, good primary vertex		
Lepton selection	quality cuts conv. rejection $I_{rel}^{PF}(\Delta R = 0.3) < 0.17$ $p_T > 20$ GeV $\lvert\eta\rvert < 2.4$	one muon (see $\mu\mu$ channel) and one electron (see ee channel)	quality cuts global muon $I_{rel}^{PF}(\Delta R = 0.3) < 0.20$ $p_T > 20$ GeV $\lvert\eta\rvert < 2.4$
Dilepton inv. mass	$m_{ee} > 12$ GeV $\lvert m_Z - m_{ee}\rvert > 15$ GeV	$m_{e\mu} > 12$ GeV	$m_{\mu\mu} > 12$ GeV $\lvert m_Z - m_{\mu\mu}\rvert > 15$ GeV
Jets	≥ 2 particle flow jets, anti-k_T with $R = 0.5$, loose jet id $p_T > 30$ GeV, $\lvert\eta\rvert < 2.4$		
Neutrinos	$E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV	no cut	$E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV
b tag	one b tagged jet, impact parameter based tagger with 10 % misidentification rate		
Kin. reco.	at least one solution found		

Table 7.4: Summary of all cuts applied

$\mu^+\mu^-$ sample	2 leptons	2 jets	$E_{\mathrm{T}}^{\mathrm{miss}}$	b-tag	kin. fit
$\mathrm{t\bar{t}}$ signal	1088.8	833.1	710.1	664.4	588.2
other $\mathrm{t\bar{t}}$	3.5	2.9	2.2	1.6	1.5
tW	102.5	39.3	33.4	29.5	22.2
diboson	267.5	20.0	10.2	3.5	2.3
W	9.0	0.7	0.0	0.0	0.0
DY$\rightarrow \tau\tau$	1056.3	44.5	29.4	8.6	8.0
DY$\rightarrow \ell^+\ell^-$	38565.7	1412.0	124.9	42.5	27.2
QCD	16.3	0.0	0.0	0.0	0.0
Sum MC	41109.6	2352.4	910.2	750.1	649.4
Data	44883.0	2745.0	907.0	714.0	612.0

$\mu^\pm e^\mp$ sample	2 leptons	2 jets	$E_{\mathrm{T}}^{\mathrm{miss}}$	b-tag	kin. fit
$\mathrm{t\bar{t}}$ signal	2655.0	2036.9	2036.9	1906.2	1706.8
other $\mathrm{t\bar{t}}$	21.2	17.2	17.2	15.5	14.0
tW	242.0	92.6	92.6	81.1	62.0
diboson	571.1	25.1	25.1	7.4	5.8
W	280.1	20.3	20.3	5.5	4.1
DY$\rightarrow \tau\tau$	2116.1	92.5	92.5	29.3	27.2
DY$\rightarrow \ell^+\ell^-$	154.9	7.4	7.4	2.2	1.8
QCD (see below)	168.4	17.3	17.3	17.3	17.2
Sum MC	6208.7	2309.2	2309.2	2064.5	1838.8
Data	6368.0	2272.0	2272.0	1967.0	1764.0

e^+e^- sample	2 leptons	2 jets	$E_{\mathrm{T}}^{\mathrm{miss}}$	b-tag	kin. fit
$\mathrm{t\bar{t}}$ signal	979.3	743.3	632.0	590.6	518.2
other $\mathrm{t\bar{t}}$	13.5	11.3	9.7	8.2	7.3
tW	92.4	34.7	29.7	25.9	19.4
diboson	234.1	17.7	8.6	2.8	1.8
W	245.9	17.3	9.6	2.7	1.3
DY$\rightarrow \tau\tau$	980.8	47.3	30.0	8.6	7.9
DY$\rightarrow \ell^+\ell^-$	33606.1	1241.6	92.6	28.4	17.6
QCD	358.6	0.0	0.0	0.0	0.0
Sum MC	36510.7	2113.2	812.3	667.3	573.5
Data	39317.0	2570.0	908.0	695.0	597.0

Table 7.5: Event selection yields in the different channels after each selection cut. The expected number of events in simulation corresponds to the luminosity of the data, and correction factors (described in chapter 8) have been applied. Only one event with a large weight remains in the QCD selection for the $e\mu$ channel (originating from the `QCD_Pt-80to170_BCtoE_TuneZ2_7TeV-pythia6` sample).

Chapter 8

Data-driven background determination and systematic uncertainties

In this chapter the different data-driven background estimates taken into account in this analysis are discussed. They result in scaling factors for their respective area; these factors are applied to the Monte Carlo samples to make the simulation agree better with data.

In addition to scaling factors which correct for shortcomings or simplifications in the simulation, for example a not 100 % correctly modelled detector geometry or a cut-off in a calculation at leading order, another type of event weighting exists for quantities in data which change during the data taking, i.e. the integrated luminosity and pile-up distribution.

In addition the systematic uncertainties associated to the scaling factors and weightings are described. While all uncertainties need to be taken into account for the measurement of the total inclusive cross section, many uncertainties cancel out at least partly when measuring normalised differential cross sections. The part that remains in normalised cross sections is the difference in the *shape* of a distribution. Therefore the uncertainties on the normalised results are much lower than for non-normalised results. In this work, all systematic uncertainties are determined separately in each bin of each measurement, so that the values given in this chapter are only typical values, defined as the average uncertainty of all bins in the lepton η distribution. Especially some edge bins with low statistics have larger uncertainties due to statistical effects. This can for example be seen in symmetric distributions (η or y) where the systematic uncertainties in the same bins on opposite sides of the distribution varies strongly.

8.1 Luminosity

When simulating events of a process, of course only a finite number N of events can be produced. This number should be as large as possible to get a more precise prediction of the process. On the other hand a given process only has a certain limited cross section

σ. This cross section has either been measured in previous experiments and/or analyses, or it has been calculated up to a certain order. Using these values, one can assign an equivalent MC-luminosity L_{MC} to each sample using equation 3.3: $L_{MC} = N_{MC}/\sigma$.

To compare the process with data, the equivalent MC-luminosity needs to be scaled to the luminosity L_{data} delivered by the experiment. Thus the whole simulated sample has to be weighted with the following luminosity weighting factor c_{lumi}:

$$c_{lumi} = \frac{L_{data} \cdot \sigma}{N_{MC}} \tag{8.1}$$

This weighting is calculated and applied individually for each simulated process.

The value for the uncertainty on the data luminosity was determined to be 4.5 % at the time this analysis was done, while the current measurement is more precise and only has an uncertainty of 2.5 % [125]. No differences of the shapes are introduced by a different luminosity, so the luminosity uncertainty is irrelevant for the normalised results.

8.2 Pile-up

8.2.1 Reweighting of the simulation

The Monte Carlo event simulation (see chapter 5) is done assuming a certain LHC setup. However, with time the setup has changed so that for example bunches contain a larger number of protons and beams are better collimated. Thus there are more proton-proton interactions in a single bunch crossing. During production of the Monte Carlo events, it was not clear how the LHC would perform and especially how the distribution of the number of primary vertices in an event (pile-up) would look like. To be able to use the simulation for different scenarios, the vertex multiplicity was supposed to be generated with a flat shape up to 10 vertices and then to fall off steeply in a tail up to multiplicities of about 20. The blue curve in figure 8.1 shows the ideal behaviour. Due to the simulation of out-of-time pile-up, i.e. vertices from preceding or succeeding bunch crossings, the observed distribution (red curve) effectively corresponds to a convolution of the ideal curve with a Poisson distribution at every bin, leading to a softer tail [126]. The first bin (one vertex) is an exception: due to a bug with the random number generation seed too many events were produced there.

In order to use the simulated events for a given scenario, they need to be reweighted to match the real distribution in the LHC. This works well as long as the real distribution peaks in the area where the simulated distribution is flat—otherwise large weights will be given to just a few events in the tail, effectively reducing the statistics. Thus this simulation can only be used for the first half of 2011 LHC operation, later data contains too much pile-up.

In figure 8.2 the vertex multiplicity is shown in events containing one muon and one electron. On the top left plot, it is clearly visible that the simulated multiplicity does

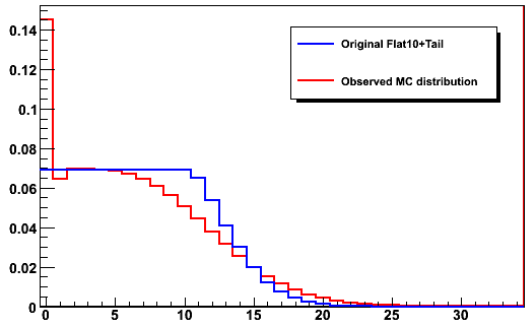


Figure 8.1: Ideal (blue curve) and observed (red curve) primary vertex distribution for simulated samples [126], including the unintended large one vertex bin.

not match the data. On the right, the plot is shown after the reweighting. Both plots are created after the selection of two isolated leptons.

The reweighting itself is done using recommendations from the CMS pile-up reweighting group [126]. This means that only in-time pile-up is taken into account, i.e. multiple interactions which occur in the same bunch crossing. Neglecting out-of-time pile-up, one can reweight the observed vertex distribution in the simulation to the one in the data. Remaining differences between data and simulation are covered by the pile-up systematic uncertainty.

8.2.2 Uncertainty

The uncertainty caused by pile-up is estimated by varying the mean of the vertex multiplicity up and down by 0.6 events, corresponding to the CMS recommendation [127]. These variations are shown in figure 8.3.

When comparing with the nominal pile-up (figure 8.2), one can clearly see the shift to the left and right away from the data, i.e. the pile-up variation is conservative. It seems that the data prefers a slightly higher pile-up than the nominal, but lower than the systematic upward shift. With both shifts all cross sections are recalculated to estimate the influence of the final result on pile-up.

Due to the pile-up removal procedure (see 6.7) and the very clean signature due to the two leptons, the result is almost independent of pile-up, and the effect is usually less than 0.5 % in all bins of the normalised measurement despite the conservative estimation process.

8.3 Trigger

The trigger efficiency for the used triggers has been determined in [128]. It is calculated in events coming from a missing energy data stream. Those triggers are only weakly cor-

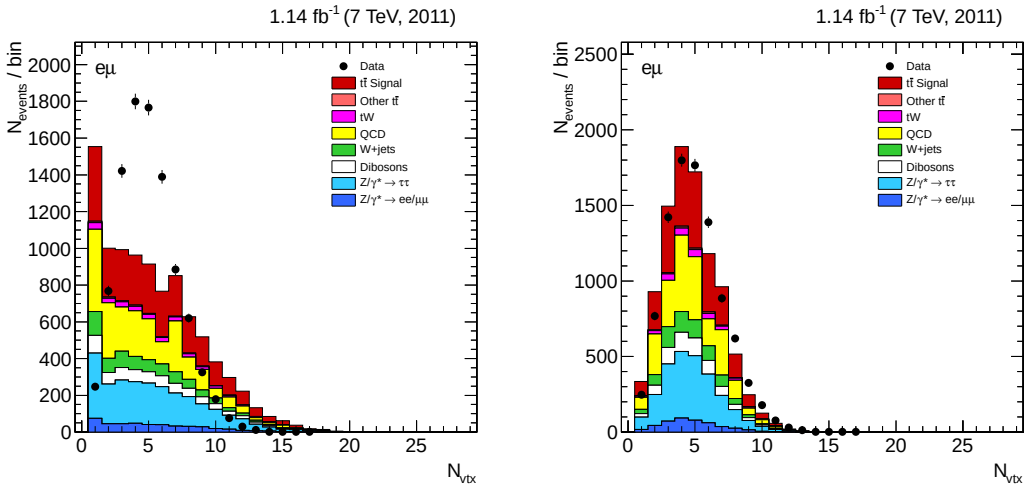


Figure 8.2: Vertex multiplicity in the $e\mu$ channel before (left) and after (right) pile-up reweighting. These plots are created after the selection of one isolated muon and one isolated electron without any further selection.

related to the dilepton triggers because they use completely different objects to trigger: they require $E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV, $E_{\mathrm{T}}^{\mathrm{miss}} > 65$ GeV and a central jet with $p_T > 80$ GeV, or $E_{\mathrm{T}}^{\mathrm{miss}} > 45$ GeV and two jets with $p_T > 60$ GeV. The number of those events which contain two leptons in the offline reconstruction is then compared to the number of events in which also the corresponding dilepton trigger has fired.

The events with large $E_{\mathrm{T}}^{\mathrm{miss}}$, a jet and two leptons are signal like, so the efficiencies in data are compared to the $\mathrm{t\bar{t}}$ MadGraph simulation. The scale factors obtained can be found in table 8.1. Also the corresponding uncertainties are listed. They were determined for an inclusive cross section measurement, i.e. they contain contributions from the absolute scale. For the normalised measurement the uncertainties are smaller, but haven't been determined. Therefore they are also used as a conservative estimate for the normalised cross section.

Channel	ee	$e\mu$	$\mu\mu$
Scale factor	0.977 ± 0.025	0.987 ± 0.020	0.961 ± 0.015
Uncertainty	2.5 %	2.0 %	1.5 %

Table 8.1: Trigger scaling factors as found in [128].

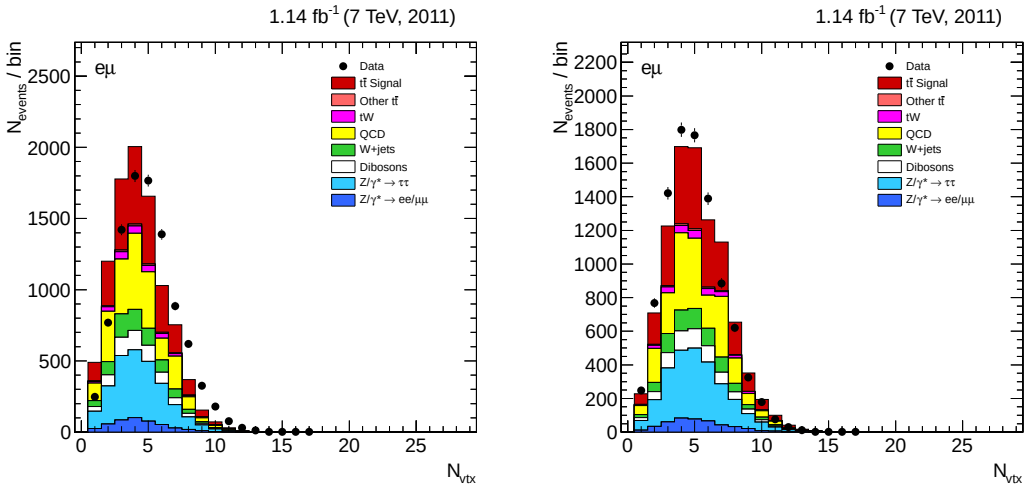


Figure 8.3: Vertex multiplicity in the $e\mu$ channel with systematic downward (left) and upward (right) fluctuation in the pile-up reweighting. These plots are created after the selection of one isolated muon and one isolated electron without any further selection.

8.4 Lepton scaling factor

The reconstruction of leptons depends on a number of factors, for example on the reconstruction algorithm but also on the detector performance. A tag and probe method on the Z^0 peak has been used to measure the identification and isolation efficiencies [129] with the result that they are above 80 % for electrons and above 90 % for muons in all bins of the measurement. The simulation describes these efficiencies correctly, so that the remaining differences to data are smaller than 2 % for the identification and almost negligible (< 0.5 %) for the isolation efficiency.

To account for possible differences of the leptons coming from Z events and those from $t\bar{t}$ events (which have additional hadronic activity), another 2 % uncertainty are added (determined in [128]), resulting in an estimate of 4 % for the total cross section. For the differential cross section it is assumed that only 50 % of this leads to a different normalisation and the rest causes different shapes, thus the estimate for the uncertainty on normalised cross sections is 2 %. Taking these numbers from other analyses leads to a rather large uncertainty, but it is sufficient for a first differential cross section measurement.

The combined identification and isolation scale factors have been measured in four different bins of the lepton transverse momentum and two respectively three pseudorapidity bins for electrons and muons. The values can be found in table 8.2.

Deviations in the lepton energy scale are smaller than 1 % and neglected.

Electrons	$\lvert\eta\rvert < 1.5$	$\lvert\eta\rvert \geq 1.5$
20 GeV $\leq p_T <$ 30 GeV	1.0094	1.0537
30 GeV $\leq p_T <$ 40 GeV	1.0027	1.0324
40 GeV $\leq p_T <$ 50 GeV	0.9981	1.0110
50 GeV $\leq p_T$	0.9990	1.0056

Muons	$\lvert\eta\rvert < 1$	$1 \leq \lvert\eta\rvert < 2.1$	$\lvert\eta\rvert \geq 2.1$
20 GeV $\leq p_T <$ 30 GeV	1.0030	0.9984	0.9870
30 GeV $\leq p_T <$ 40 GeV	0.9945	0.9921	0.9811
40 GeV $\leq p_T <$ 50 GeV	0.9988	0.9901	0.9924
50 GeV $\leq p_T$	1.0022	0.9896	0.9947

Table 8.2: Scale factors for electrons (top) and muons (bottom table) including isolation and identification [129]. They are calculated in two (three) pseudorapidity ranges for electrons (muons) and in four different p_T areas.

8.5 Drell-Yan

The most important background in the ee and $\mu\mu$ channels is the contribution by Drell-Yan events. Therefore this analysis does not purely rely on the simulation for this process. Instead, the normalisation is determined from the data by scaling the DY contribution to the measured Z mass peak [128, 130].

For this method to work, the events which were cut out by the dilepton invariant mass cut ($|m_Z - m_{\ell\ell}| > 15$ GeV, see section 7.8) are used. The further requirements of two jets and missing transverse energy are also applied to these events. Then the number of Drell-Yan events in the $|m_Z - m_{\ell\ell}| \leq 15$ GeV region is determined in data and the simulation. The simulation is then scaled to the data such that the number of events inside the Z mass peak is the same for data and simulation. The result can be seen in figure 8.4.

To calculate the number of observed DY events outside the Z veto $N_{out}^{\ell^+\ell^-,obs}$, the following formula is used (for $\ell\ell = ee$ or $\mu\mu$):

$$N_{out}^{\ell^+\ell^-,obs} = R_{out/in}^{\ell^+\ell^-,MC} \cdot \left(N_{in}^{\ell^+\ell^-} - \frac{1}{2} N_{in}^{e\mu} k_{\ell\ell} \right) \tag{8.2}$$

The ratio $R_{out/in}^{\ell^+\ell^-,MC}$ of events outside the Z mass region N_{DYMC}^{out} over the number of events inside the Z peak N_{DYMC}^{in} is taken from the the DY simulation:

$$R_{out/in}^{\ell^+\ell^-,MC} = \frac{N_{DYMC}^{\ell^+\ell^-,out}}{N_{DYMC}^{\ell^+\ell^-,in}} \tag{8.3}$$

To account for the non-DY events in the number of measured events inside the DY peak region $N_{in}^{\ell^+\ell^-}$, the number of events from other processes has to be subtracted. This

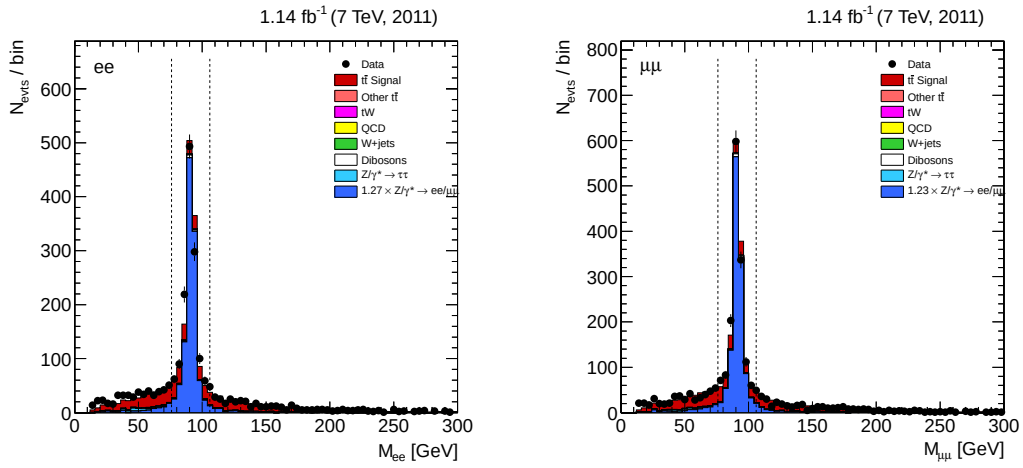


Figure 8.4: Scaling of the Drell-Yan sample in the ee (left) and $\mu\mu$ channels (right). The DY contribution is scaled so that the number of events inside the Z mass veto region (indicated with dashed lines) matches the number of data events in this region. This is done after the missing energy cut.

number, $N_{in}^{e\mu}$, is determined in the $e\mu$ channel which does not contain DY events. Due to the higher combinatorics for the $e\mu$ channel, a factor of 1/2 has to be applied here. In addition, the differences between the selection efficiencies in the ee and $\mu\mu$ channels need to be taken into account. This is done using the $k_{\ell\ell}$ factors:

$$k_{ee} = \sqrt{\frac{N_{in,loose}^{e^+e^-}}{N_{in,loose}^{\mu^+\mu^-}}} = 0.96 \pm 0.01 \tag{8.4}$$

$$k_{\mu\mu} = \sqrt{\frac{N_{in,loose}^{\mu^+\mu^-}}{N_{in,loose}^{e^+e^-}}} = 1.04 \pm 0.01 \tag{8.5}$$

These factors are determined with a looser selection, i.e. without the missing energy cut applied to keep a larger amount of events. This is justified because the selection efficiency for muons and electrons should not depend on the missing energy in the event. On the other hand the $k_{\ell\ell}$ factors need to be calculated from events as close to the final selection as possible, therefore removing further cuts introduces larger uncertainties.

The final scaling factor c_{DY} for the simulation can be calculated by dividing the contributions from data by the ones from simulation:

$$c_{DY} = DY_{data}/DY_{MC} \tag{8.6}$$

The resulting values can be found in table 8.3.

Channel	e^+e^-	$\mu^+\mu^-$
DY MC	91.8 ± 4.2	121.8 ± 5.0
Estimate from data	116.8 ± 6.8	149.8 ± 0.8
$R_{out/in}$	0.11 ± 0.01	0.12 ± 0.01
c_{DY}	1.27 ± 0.11	1.23 ± 0.09

Table 8.3: Data-driven Drell-Yan background estimation in the $\mu\mu$ and ee channels after the missing transverse energy cut.

8.5.1 Scale factor at different selection steps

In previous analyses, e.g. [91], it was found that the scale factor strongly depends on the selection steps. While it was consistent with 1 up to the requirement of only one jet, it went up to 2.9 after the requirement of missing energy in the $\mu\mu$ channel.

With the newer simulation used in this work, the agreement between data and simulation got much better and the scale factor is consistent with 1 before the missing energy cut. Table 8.4 shows the scale factors determined at the later steps.

Selection step	Missing energy	b tag	kin. reco.
e^+e^-	1.27 ± 0.11	1.37 ± 0.19	1.47 ± 0.24
$\mu^+\mu^-$	1.23 ± 0.09	1.27 ± 0.16	1.30 ± 0.21

Table 8.4: Drell-Yan scale factors after different selection steps.

One can still see increasing scale factors when tightening the cuts. This is probably caused by mismeasured leptons in DY events which migrate out of the Z peak. Note that these events must have initial and/or final state radiation to survive the jet cuts as well as missing energy, i.e. already a mismeasured transverse momentum balance. This imbalance might be caused by a mismeasured lepton, for example. It seems that this effect is a bit larger in the data than in the simulation, and the ee channel is slightly more affected than the $\mu\mu$ channel. However the scale factors determined after the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, after b tagging and after the kinematic reconstruction are all compatible within their uncertainties. It was chosen to apply the scale factor determined after the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, i.e. when it first becomes different from 1, and take the differences in the further steps into account for the uncertainty.

8.5.2 Uncertainty

Because of the different scaling factors in the different steps, a rather large variation by ±50 % is applied to the Drell-Yan samples (DY $\rightarrow ee/\mu\mu$ only) to determine the uncertainty on the cross section. This value is a conservative choice made to account for the differences in the calculated scaling factors at different selection steps. In the ee channel,

the factor ranges between 1 at loose selection steps and ≈ 1.5 at the full selection, a hint that the simulation might have problems to describe the data correctly. And the scaling itself does not necessarily help because it only scales the absolute DY contribution but assumes that the shape description, i.e. the $R_{out/in}^{\ell^+\ell^-,MC}$ variable is described correctly.

After applying this ± 50 % variation, the scale factor is recalculated before the cross sections are determined again.

In addition, the data/simulation disagreement at low invariant dilepton masses (see 7.8) needs to be taken into account. Therefore also the cut on $m_{\ell\ell}$ is raised from 12 to 30 GeV as a variation. Both uncertainties are added in quadrature and the result is taken as final DY uncertainty.

For the normalised results, this total uncertainty is in the order of 1.5 % for the ee and $\mu\mu$ channels and about 0.5 % in the $e\mu$ channel.

8.6 Non-Drell-Yan backgrounds

All backgrounds except the the DY $\rightarrow ee/\mu\mu$ simulation are scaled up and down by 30 % simultaneously to determine their influence. This is a rather large estimate also used in other analyses [91, 131]. As it is not the dominant uncertainty, no attempts to reduce this value were made.

The important backgrounds are the single top production, top quark pairs which do not decay into dileptons, and DY events decaying into a $\tau^+\tau^-$ pair. As the selection is very efficient, all remaining backgrounds together (see table 7.5) only make up about 5 % of the selected events, the resulting variation is very small: in all three channels the variation typically leads to 0.3 % uncertainty on the normalised cross section.

8.7 Kinematic event reconstruction

As explained in section 7.12, the kinematic event reconstruction is only needed to calculate the properties of the top quark. The cross section of the top quark production does not need these quantities as input, and even differential cross sections (see section 9.2) can be calculated as long as they are differential in a quantity that can be measured directly.

To estimate if the kinematic event reconstruction modifies the resulting cross section, figure 8.5 shows the lepton pseudorapidity and transverse momentum before (red) and after (blue) the kinematic event reconstruction for data (points) and simulation (histograms). The ratio plot below shows the efficiency, i.e. the percentage of events which pass the reconstruction (again for data (points) and simulation (lines)).

For the pseudorapidity on the left, the efficiency in data is in good agreement with the simulation. In addition, the efficiency is flat in η—only at large $|\eta|$ values the it seems to be a bit lower, but that is well covered by the statistical uncertainties. Therefore

it can be concluded that the reconstruction does not depend on the lepton η and no corrections need to be applied.

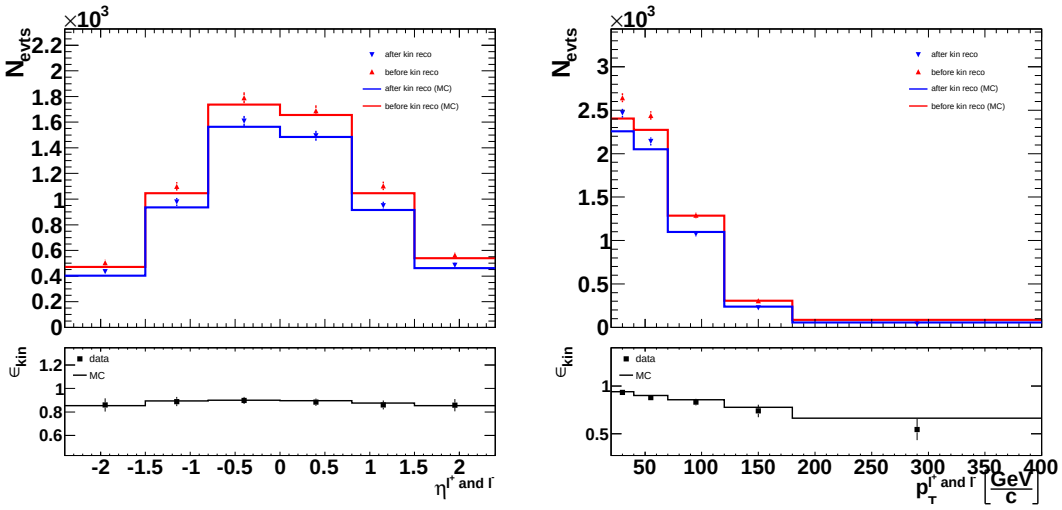


Figure 8.5: Differences between the number of events before (red) and after (blue) the kinematic reconstruction. The left plot shows the behaviour in the pseudo-rapidity and the right plot in the transverse momentum of the leptons.

This is different for the transverse momentum of the leptons, a clear p_T dependence is visible. The kinematic reconstruction is obviously less efficient for larger transverse momenta of the leptons, both for data and the simulation. And it is also visible that the data and simulation efficiencies behave slightly different. Therefore a scale factor for the kinematic event reconstruction is calculated for each bin in the lepton p_T, the values can be found in table 8.5.

p_T range (GeV)	20 – 40	40 – 70	70 – 120	120 – 180	> 180
scale factor	0.9940	0.9736	0.9728	0.9499	0.8201

Table 8.5: Scale factors per lepton determined for the kinematic event reconstruction in five different p_T ranges.

As there are two leptons in the event and both have to be taken into account, the geometrical mean is calculated as the final scaling factor c_{reco} from the individual scaling factors for for the first ($\ell 1$) and second ($\ell 2$) lepton:

$$c_{reco} = \sqrt{c_{reco}^{\ell 1} \cdot c_{reco}^{\ell 2}} \tag{8.7}$$

Further studies on the event reconstruction have been performed with the top mass fixed to 173 GeV and/or using only the two leading jets regardless of b tagging [132].

The scale factor is a measure for how well the kinematic event reconstruction describes the events in different regions of the detector. Therefore the analysis is repeated with and without applying the scale factor and the resulting differences are taken as systematic uncertainty. The effect is 2.0 % in same flavour lepton channels and 1.7 % in the $e\mu$ channel for absolute cross sections and typically 1 % for the normalised cross sections (up to 4 % at large top quark momenta).

8.8 Jet energy scale

As explained in section 6.4, several corrections factors are applied to jets. The uncertainty on the jet energy scale depends on the detector region (η) in which a jet is reconstructed and on its measured momentum. The exact values can be found in [63, 133]. In this analysis the jet energy scales of all jets from all simulated samples, i.e. the respective 4-momenta, are varied up and down within their uncertainty, then the complete analysis is repeated from the beginning.

The resulting cross sections with upward and downward jet energy scale variation are then compared to the result without any variation (nominal result). The absolute values of the relative differences to the nominal result are averaged for both the up- and downward variations in each bin of all measurements to determine the uncertainty.

The difference for the total inclusive cross section is between 2.3 % ($\mu\mu$) and 2.4 % (ee and $e\mu$ channel), the normalised results only change by typically 0.2 % ($e\mu$ channel) to 0.5 % (ee channel). The *typical* number quoted here is defined as the average uncertainty of all bins in the lepton η distribution. Detailed uncertainties for all bins in all distributions can be found in figure 8.7 on page 102.

8.9 Jet energy resolution

The jet energy resolution (JER) is defined as the standard deviation of a Gaussian fitted to the jet response. It turns out to be slightly different in data and simulation, so that the resolution in the simulation is increased by default by 10 % to fit the data best [134].

The JER uncertainties are determined by varying the resolution depending on the jet pseudorapidity: in the central detector the variations are smaller than in forward directions. The values for the variation [134] are given in table 8.6.

The influence on the results has been determined by applying the variation, rerunning the complete analysis and comparing with the nominal result, just as for the JES. Again, upward and downward fluctuations are averaged and the typical value given for the normalised results is the average of the lepton η uncertainties.

The result is that the JER uncertainties are smaller than the JES uncertainties, the total inclusive cross section only changes by 1.0 % for ee, 0.8 % for $\mu\mu$ and 1.3 % for the $e\mu$ channel. The typical effect on normalised differential cross section is below 1 %.

	$\lvert\eta\rvert < 1.5$	$1.5 \leq \lvert\eta\rvert < 2.0$	$\lvert\eta\rvert \geq 2.0$
JER down	1.0	0.95	0.9
Nominal	1.1	1.1	1.1
JER up	1.2	1.25	1.3

Table 8.6: Jet energy resolution modification factors [134] for the nominal setup (+10 % everywhere) and systematic variations in three different η bins.

The uncertainties for all bins in all distributions can be found in figure 8.7 on page 102.

8.10 Hadronisation

As described in chapter 5, the event simulation contains a hadronisation step which is based on a phenomenological description and tuning of the event generator to the data. To determine an uncertainty on the hadronisation two different programs are used for the hadronisation step. As the MadGraph simulation was only available with Pythia for hadronisation, two Powheg samples were selected for the signal simulation (`TTTo2L2Nu2B_7TeV-powheg-herwig6` and `TTTo2L2Nu2B_7TeV-powheg-pythia6`). One of them uses Pythia, the other herwig for hadronisation. The analysis is then redone once for each of these new signal simulations while keeping all all other simulated samples constant. The relative differences between the results are finally applied to the results with the MadGraph simulation.

The uncertainties determined this way are about 5 % (4.9 % for ee, 4.3 % for $\mu\mu$ and 5.4 % for $e\mu$) and thus dominant for the total inclusive cross section measurements. The effect on differential measurements is much smaller, typically 1 % (ee channel) or below. Again, the typical value is taken from the average of the lepton η. The statistics in these samples are small, this leads to large hadronisation uncertainties for example in highest bin of the $p_T^{\ell\ell}$ distribution where the uncertainty is ~ 16 % (see figure 8.7).

8.11 Mass of the top quark

Similar to the hadronisation, also for the top quark mass uncertainty determination only the signal sample needs to be varied. The default top quark mass in the MadGraph simulation is 172.5 GeV and only simulated mass variations with a difference of at least ± 6 GeV, i.e. 166.5 GeV and 178.5 GeV were available (`TTJets_TuneZ2_mass166_5_7TeV-madgraph-tauola` and `TTJets_TuneZ2_-mass178_5_7TeV-madgraph-tauola`). The top quark mass is however known to a higher precision (172.0 ± 1.6 GeV [12] when the samples were created, now 173.5 ± 1.0 GeV [13]), so that the variations are too large. Thus, the uncertainties determined from the

large mass variations samples are divided by 5, corresponding to a variation of ± 1.2 GeV.

The variations result in an uncertainty of 1.3 % for the $\mu\mu$ channel and 0.9 % for the ee and $e\mu$ channels for the total inclusive cross section. The typical values for the normalised differential cross sections are below 0.5 % for all channels.

8.12 Q^2, ISR, FSR and matching scale

Uncertainties on the QCD renormalisation and factorisation scales are determined by varying the the $Q^2 = m_t^2 + \sum p_T^2$ scale for hard scattering and the subsequent parton showering simultaneously. The upward variation changes Q by a factor of 2, the downward variation by a factor of 1/2 [28]. This change is done by using dedicated samples for the signal simulation (`TTjets_TuneZ2_matchingup_7TeV-madgraph-tauola` and `TTjets_TuneZ2_matchingdown_7TeV-madgraph-tauola`) and rerunning the analysis.

No additional uncertainty due to initial and/or final state radiation needs to be taken into account, it is already covered by this uncertainty.

The resulting uncertainty on the total inclusive cross section is 1.7 % ($e\mu$ channel) to 3.2 % (ee channel), for the normalised cross sections the typical value is smaller than 1 % for the $e\mu$ channel, 1.3 % in $\mu\mu$ and 2.4 % in ee. These dedicated Monte Carlo samples have small statistics, which is one possible source of the different typical values in the three channels.

Also dedicated $\mathrm{t\bar{t}}$ samples are used for the variation of the matching of parton showers to partons (`TTjets_TuneZ2_scaleup_7TeV-madgraph-tauola` and `TTjets_TuneZ2_scaledown_7TeV-madgraph-tauola`). In the nominal sample, the threshold for interfacing MadGraph matrix elements with Pythia's parton showering is 20 GeV, the variations use 10 and 40 GeV instead [28]. The resulting uncertainty is 2 to 3 % for the total inclusive cross section and, the typical value for the normalised differential cross sections is 0.6 to 1.1 %.

8.13 b tagging

B tagging is one of the important selection steps to separate signal events from background in the ee and $\mu\mu$ channels, and it is also applied in the $e\mu$ channel for consistency and the use of the b tagging information in the kinematic event reconstruction.

The efficiency for a jet being b tagged by the track counting algorithm at the loose working point ϵ_b (i.e. the algorithm used in this work, see also section 7.11) as well as the corresponding scale factor s_b have been determined in [135]:

$$\epsilon_b = 0.76 \pm 0.01 \tag{8.8}$$

$$s_b = 0.95 \pm 0.10 \tag{8.9}$$

As the top quarks events contain two b jets, the scale factor s_b has to be converted to represent that only one out of two b jets need to be matched:

$$s_2 = \frac{1-(1-\epsilon_b)}{1-\left(1-\frac{\epsilon_b}{s_b}\right)} = 0.982 \pm 0.034 \tag{8.10}$$

The uncertainty of 3.4 % is calculated by error propagation. For the differential cross section, half of the value is used, i.e. 1.7 %.

These b tag values were cross checked using the $e\mu$ channel (implemented mainly by [136]). This is done using the following method (similar to [137]): assuming that the two leading jets are the b jets, the following relations are found:

$$N_0 \propto (1-\epsilon)^2 \tag{8.11}$$
$$N_1 \propto \epsilon(1-\epsilon) + (1-\epsilon)\epsilon = 2\epsilon(1-\epsilon) \tag{8.12}$$
$$N_2 \propto \epsilon^2 \tag{8.13}$$

In these equations, N_i is the number of events with exactly i b tags in the two leading jets. The probability to not tag a b jet is $1-\epsilon$, hence the squared value yields N_0. The probability to tag both jets is ϵ^2, and the probability for one jet to be tagged is the combination of a tagged and an untagged jet multiplied with 2 for combinatorial reasons.

The simple assumption that the b jets are always leading jets is however wrong, so corrections have to be applied to the formulas 8.11 to 8.13. This is done by determining the fractions α_j of events with j b jets ($j = 0, 1, 2$) among the two leading jets from the simulation. It turns out that both b jets are the leading jets in only in 64 % of the events ($\alpha_2 = 0.64$), while in 32 % one jet from ISR or FSR is among the leading jets ($\alpha_1 = 0.32$). In only 2 %, none of the leading two jets is from a b quark ($\alpha_0 = 0.02$).

Another simplification in the calculation is that the rate of false positives is neglected, i.e. b tagged jets which do not originate from a b jet (light flavour or gluon jets). This results in the following modified formulae:

$$N_0 \propto \alpha_0 + \alpha_1(1-\epsilon) + \alpha_2(1-\epsilon)^2 \tag{8.14}$$
$$N_1 \propto \alpha_1\epsilon + 2\alpha_2\epsilon(1-\epsilon) \tag{8.15}$$
$$N_2 \propto \alpha_2\epsilon^2 \tag{8.16}$$

The efficiency can now be calculated:

$$\epsilon = \frac{(\alpha_1 + 2\alpha_2)N_2}{\alpha_2(N_1+N_2)} \tag{8.17}$$
$$\sigma_\epsilon = \frac{2\alpha_2+\alpha_1}{\alpha_2} \cdot \sqrt{\frac{N_1 N_2}{(N_1+N_2)^3}} \tag{8.18}$$

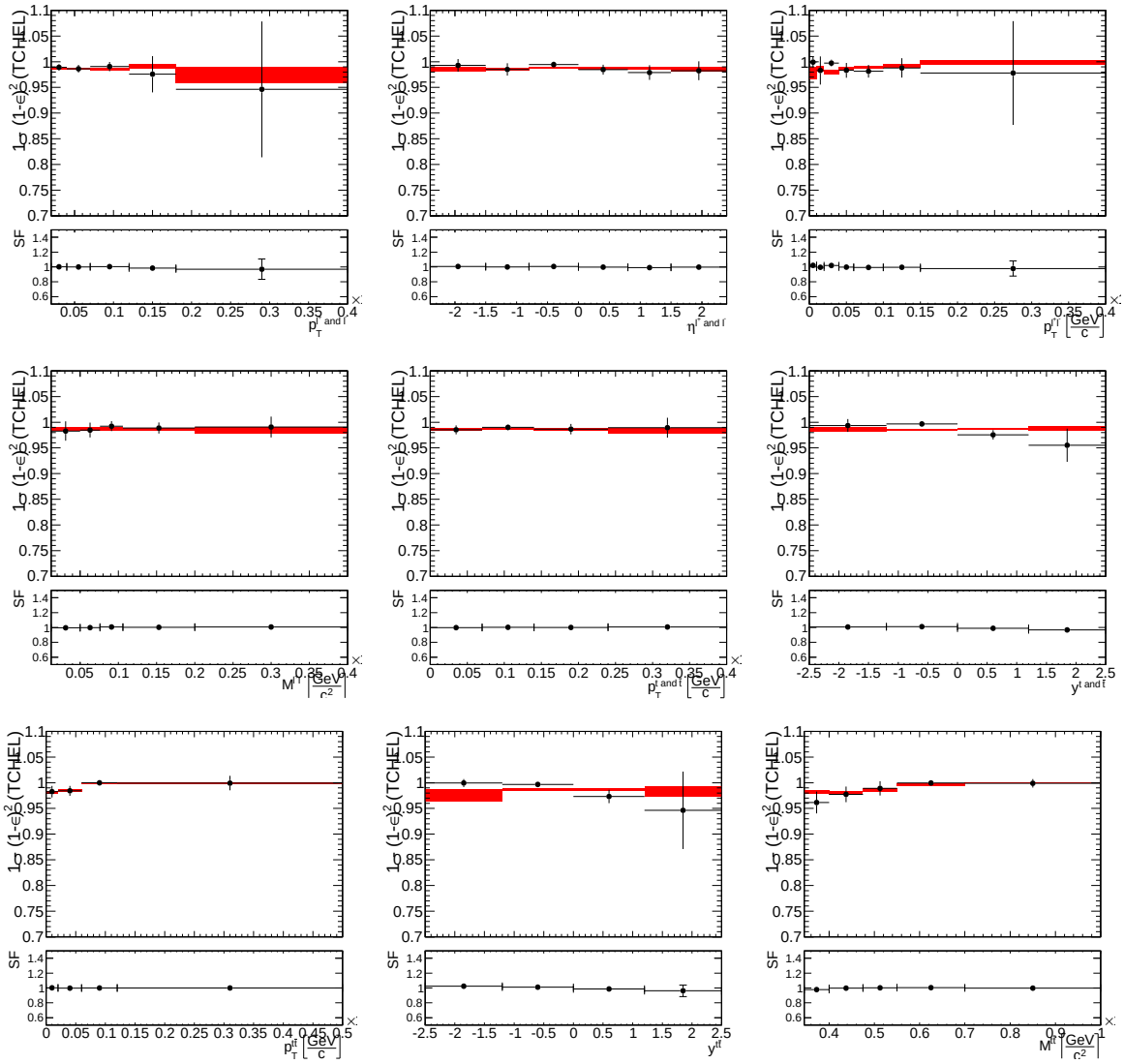


Figure 8.6: Cross check of the b tag efficiency $1-(1-\epsilon)^2$, i.e. tagging at least one out of two jets, and the scale factor SF calculated using $e\mu$ events. The red bands correspond to the simulation (including uncertainty), the black points show the data. The plots show the nine different variables in which the differential cross section are measured.

Cross section type	total			differential		
Channel	$\mu\mu$	ee	$e\mu$	$\mu\mu$	ee	$e\mu$
Trigger efficiency	1.5	2.5	2.0	1.5	2.5	2.0
Lepton selection	4.0	4.0	4.0	2.0	2.0	2.0
Backgrounds (other)	1.9	2.4	2.2	0.3	0.3	0.3
Backgrounds (DY $\rightarrow \mu\mu$ / ee)	3.1	2.1	0.1	1.6	1.4	0.5
Jet energy scale	2.3	2.4	2.4	0.3	0.5	0.2
Jet energy resolution	0.8	1.0	1.3	0.4	0.5	0.5
$E_{\mathrm{T}}^{\mathrm{miss}}$ / pile-up	0.3	0.3	0.1	0.4	0.4	0.3
B-tagging	3.4	3.4	3.4	1.7	1.7	1.7
Kinematic fit	2.0	2.0	1.7	0.5	0.5	0.5
Hadronisation	4.3	4.9	5.4	0.6	1.0	0.3
Top quark mass	1.3	0.9	0.9	0.3	0.5	0.2
Q^2 scale	2.1	3.2	1.7	1.3	2.4	0.6
ME/PS threshold	2.2	2.8	1.2	1.1	0.6	0.6
$\mathcal{BR}\,(t\bar{t} \rightarrow \ell\ell + X)$	1.5	1.5	1.5	–	–	–
Luminosity	4.5	4.5	4.5	–	–	–
Total error	12.2	12.3	11.3	4.0	4.8	3.6

Table 8.7: Summary of the systematic uncertainties (given in %). The left columns correspond to the calculation of the total cross section, the right side shows typical values (the average uncertainty of all bins in the lepton η distribution, see also figure 8.7) for the normalised differential cross section.

The results from this calculation can be found in figure 8.6 for all the observables in which differential cross sections are calculated. One can see that the scale factors are reasonably flat and in agreement with the flat overall factor of (98.2 ± 3.4) % as calculated from [135].

8.14 Summary of the uncertainties

Table 8.7 summarises the uncertainties for the total cross section as well as the normalised differential cross sections. The total error is calculated by summing the individual contributions in quadrature, resulting in approximately 12 % uncertainty for the total cross section, while the typical uncertainty on normalised cross section is only in the order of 4 %.

Detailed values for the differential cross section measurement where the uncertainties are shown for all bins of the measurement separately are given in figure 8.7. The plots show the squares of the all uncertainties in all bins, so one has to be careful when reading off the total value. The squares have been chosen because the uncertainties are added in quadrature and the plot thus allows to read the sum and the individual

contributions. What is apparent is that the uncertainties in the high invariant mass, high p_T or outer (pseudo-)rapidity bins are rather large. This is mainly caused by low statistics in the simulation in these bins. While individual values for the three channels were given in this chapter, the figures show the combination of the three channels which is explained in section 9.1.1, the corresponding plots for the separate channels can be found in appendix B.

Figure 8.8 shows the statistical uncertainties, the sum of the systematic uncertainties, and the total uncertainty in all bins. The systematic uncertainties are further split into experimental uncertainties and assumptions in the modelling (hadronisation, Q^2, matching scale, top quark mass). One can see that the statistical error is still the largest contribution but further analyses with more data will also have to reduce the model uncertainties, especially for the top quark and top quark pair quantities.

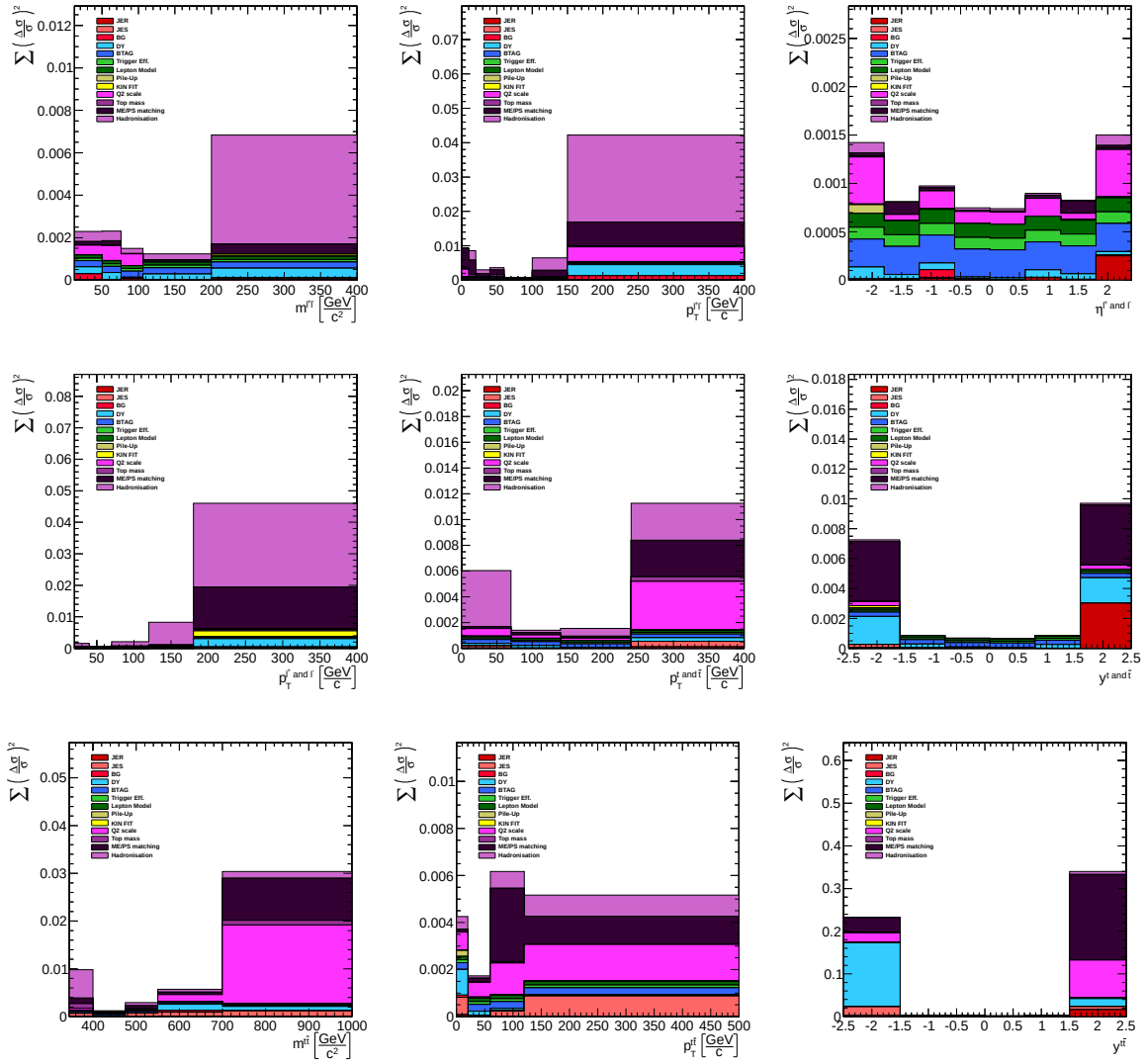


Figure 8.7: The plots show the squares of the systematic uncertainties in all bins of all measured cross section quantities. All three decay channels have been combined (see section 9.1.1) for the result.

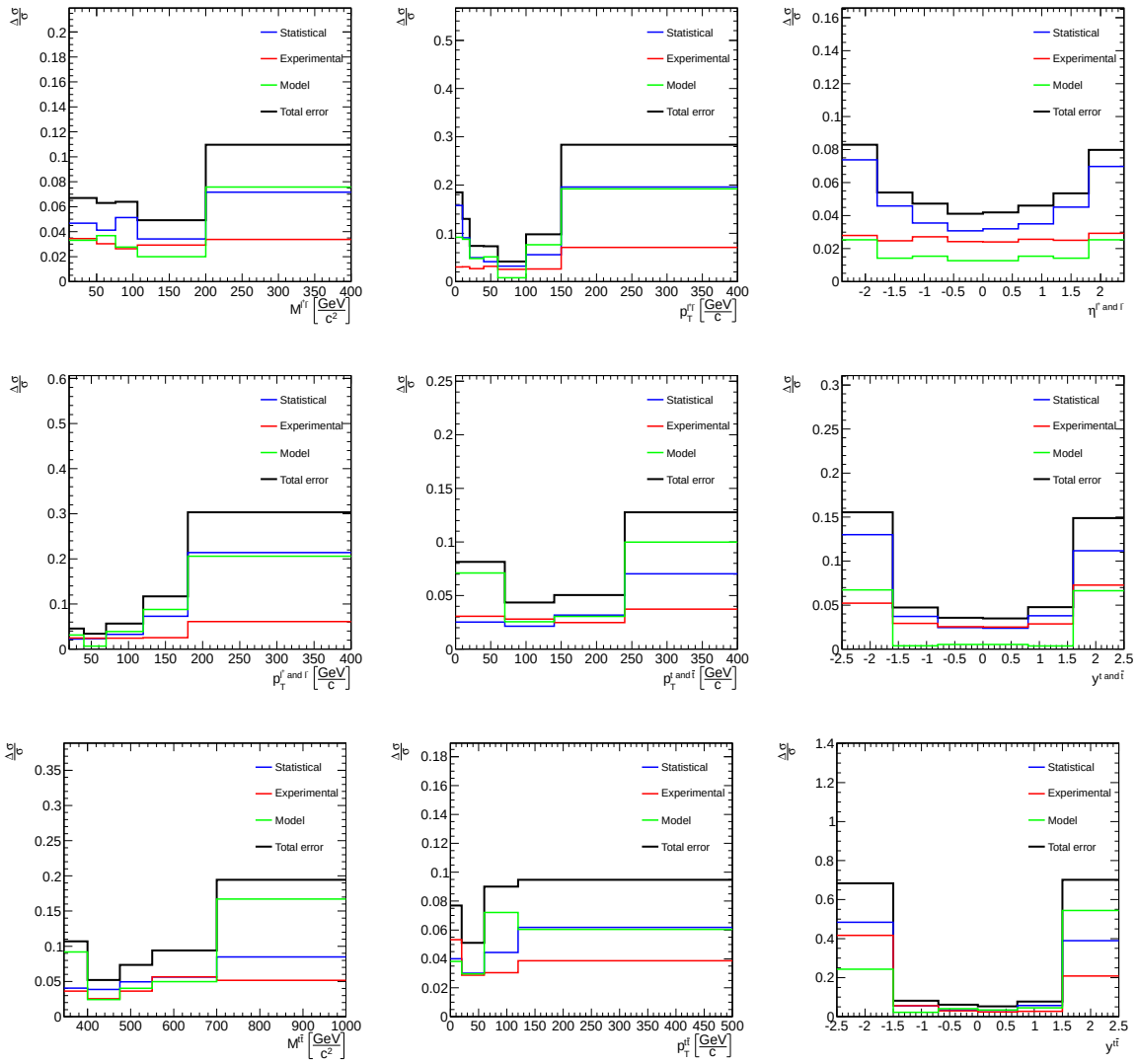


Figure 8.8: Comparison of the relative systematic (model and experimental) and statistical uncertainties in all bins of all measured cross section quantities. All three decay channels have been combined for the result.

Chapter 9

Results

In this chapter the calculation of the cross sections resulting from this work is explained. For the three dilepton decay channels and their combination the results are given. All cross sections are always determined after the full event selection as described in chapter 7 and of course also after applying all corrections and scale factors as described in chapter 8.

The main result of this work are the differential cross sections which are described in section 9.2. Before that, the total inclusive cross section is calculated in section 9.1 as a cross check with other analyses.

9.1 Total inclusive cross section

The total inclusive cross section $\sigma_{\mathrm{t\bar{t}}}$ for the production of top quark pairs is calculated using the following formula:

$$\sigma_{\mathrm{t\bar{t}}} = \frac{N_{\mathrm{Data}} - N_{\mathrm{BG}}}{L \cdot \epsilon \cdot \mathrm{BR}_{\mathrm{t\bar{t}} \to \ell\ell X}} \tag{9.1}$$

In this formula, the nominator contains the number of measured top quark events consisting of the difference between N_{Data}, the number of events measured in data, and N_{BG}, the number of background events estimated from the simulation. The denominator contains the integrated luminosity L of the analysed dataset, the branching ratio $\mathrm{BR}_{\mathrm{t\bar{t}} \to \ell\ell X}$ describing the ratio of top quark pairs decaying into final states with two leptons (electrons or muons) (see equations 2.7 to 2.9), and the signal efficiency ϵ. It is determined from the ratio of simulated signal events which survive the complete selection and includes the simulated detector acceptance. In addition, the corrections for the trigger, the lepton isolation and identification, the b tagging, and the kinematic reconstruction efficiency are applied.

Due to the selection cuts on the lepton and jet momenta and pseudorapidities, the event rate is only measured in a limited phase space. The simulation on the other hand covers the full phase space, so that the efficiency factor here also extrapolates the measured cross section to the full phase space.

Channel	N_{Data}	N_{BG}	ϵ	BR
ee	597	59.7	0.146	0.01668
$e\mu$	1764	131.2	0.245	0.03284
$\mu\mu$	612	67.4	0.166	0.01616

Table 9.1: Summary of the measured number of events, the estimated background and the efficiency (extracted from table 7.5). The branching ratios are slightly different from equations 2.7–2.9 because the MadGraph samples are produced with equal branching ratios for all $W \to l\nu$ processes (but different branching ratios for the τ decays).

The number of signal and background events are summarised in table 9.1. With those values, the total inclusive $\mathrm{t\bar{t}}$ cross sections is determined individually for each channel:

$$\sigma_{\mathrm{t\bar{t}}}^{ee} = 193.5 \pm 8.8\,(\text{stat.}) \pm 21.7\,(\text{syst.}) \pm 8.7\,(\text{lumi.})\ \text{pb} \tag{9.2}$$

$$\sigma_{\mathrm{t\bar{t}}}^{\mu\mu} = 177.2 \pm 8.0\,(\text{stat.}) \pm 19.8\,(\text{syst.}) \pm 8.0\,(\text{lumi.})\ \text{pb} \tag{9.3}$$

$$\sigma_{\mathrm{t\bar{t}}}^{e\mu} = 177.7 \pm 4.6\,(\text{stat.}) \pm 18.2\,(\text{syst.}) \pm 8.0\,(\text{lumi.})\ \text{pb} \tag{9.4}$$

The combinatorics due to the decay of two W bosons ($WW \to e\mu$ and $WW \to \mu e$) lead to twice as many produced $e\mu$ events as ee or $\mu\mu$ events. In addition, also the efficiency is higher because there is no Z mass cut in the $e\mu$ channel. It is therefore the expected result that the cross section measured in the $e\mu$ channel delivers the best result. The statistical error is about half the size of the other channels and also the systematic uncertainty is slightly smaller in this channel. This is mainly due to a smaller background contribution from Drell-Yan events.

However, the general systematic uncertainty is rather large in all channels and the result is not competitive with dedicated cross section measurements, e.g. [138]. As the main focus of this work is the measurement of differential cross sections, no further studies have been made to reduce these uncertainties.

9.1.1 Combination

Even though the result from the $e\mu$ channel is best, the same quantity is measured in all three channels, so one can combine the results. Due to the unambiguous assignment of an event to just one channel (see section 7.7.1), these are statistically independent. Thus, the individual results can be weighted by their statistical uncertainty:

$$\sigma_{\mathrm{t\bar{t}}}^{\text{combined}} = \frac{\sum_i \frac{\sigma_i}{\delta\sigma_i^2}}{\sum_i \frac{1}{\delta\sigma_i^2}} \quad \text{where } i \in ee, e\mu, \mu\mu \tag{9.5}$$

The statistical error of the combination is the inverse quadratic sum:

$$\delta\sigma = \frac{1}{\sqrt{\sum_i \frac{1}{\delta\sigma_i^2}}} \tag{9.6}$$

The systematic error is not calculated by simple error propagation because the uncertainties are correlated in the three channels. Thus the systematic uncertainties of the individual channels are ignored for the combination. Instead, the central value of the combined cross section is calculated for the systematics variations. The difference between the combined result for the nominal setup and the combined result for a systematic variation is then taken as the uncertainty for this variation. In those cases where both upward and downward fluctuation were available, the average of the relative individual uncertainties was taken as uncertainty. This is done for all systematics determined in this work. For the b tagging the same value as for the individual channels is taken for the combination because there is no reason that b tagging could behave differently in the three channels. On the other hand the efficiencies for the triggers and the lepton selection could be different in the three channels. So here the inverse quadratic sum (see equation 9.6) is used to calculate the combined uncertainties (1.1 % for the trigger, 2.3 % for the lepton selection).

Using this combination method, the result is

$$\sigma_{\mathrm{t\bar{t}}}^{\mathrm{combined}} = 180.4 \pm 3.6\,(\mathrm{stat.}) \pm 17.1\,(\mathrm{syst.}) \pm 8.1\,(\mathrm{lumi.})\ \mathrm{pb} \tag{9.7}$$

The latest dedicated inclusive dileptonic $\mathrm{t\bar{t}}$ cross section measurement in CMS is [139], where a result of $\sigma_{\mathrm{t\bar{t}}} = 169.9 \pm 3.9\,(\mathrm{stat.}) \pm 16.3\,(\mathrm{syst.}) \pm 7.6\,(\mathrm{lumi.})$ pb was determined. The measurement presented in this thesis yields a slightly larger cross section, but the difference is covered by only 60 % of the systematic uncertainty. When looking at the separate channels, the ee channel gives about 16 pb larger results than the other channels and is significantly larger than other measurements. The reason for this is currently unknown. Using a newer *global tag*, i.e. more accurate detector geometry and calibration constants lead to a slightly lower cross section in the ee channel [140] but the general trend remains. A hint for the problem in ee was already that the Drell-Yan scaling (see table 8.4) is larger in the ee channel compared to the $\mu\mu$ channel, especially at the later selection steps. However, the total inclusive cross section measurement is not the main goal of this work and the whole measurement is dominated by the $e\mu$ channel anyway. Therefore no further studies were made on the ee channel.

9.2 Differential cross sections

The main result of this work are normalised differential $\mathrm{t\bar{t}}$ cross sections:

$$\frac{1}{\sigma}\frac{d\sigma}{dX} \tag{9.8}$$

A differential cross section $d\sigma/dX$ in the quantity X describes how the the inclusive cross section is distributed over this quantity, e.g. if X is the invariant mass of the $\mathrm{t\bar{t}}$ system, the differential cross sections can tell us the cross section for the $\mathrm{t\bar{t}}$ production at a given invariant $\mathrm{t\bar{t}}$ mass. In this specific distribution, one would expect a zero cross section below $2m_t$ (or almost zero because of the natural width of the top quark mass), a steep rise at $2m_t$ and then a falling distribution. Should for example a new resonance exist that can decay into a $\mathrm{t\bar{t}}$ pair, one would see a peak at the corresponding mass.

Apart from $m^{\mathrm{t\bar{t}}}$, the differential cross sections are measured as a function of the transverse momentum and invariant mass of the dilepton system $p_T^{\ell\ell}$ and $m^{\ell\ell}$, and as function of the transverse momentum and the pseudorapidity of single leptons $p_T^{\ell^+ \,\mathrm{and}\, \ell^-}$ and $\eta^{\ell^+ \,\mathrm{and}\, \ell^-}$. As a result of the event reconstruction, the differential cross sections can also be measured as a function of the the transverse momentum and rapidity of both the top and antitop quarks $p_T^{t \,\mathrm{and}\, \bar{t}}$ and $y^{t \,\mathrm{and}\, \bar{t}}$ as well as for the invariant $\mathrm{t\bar{t}}$ system $p_T^{\mathrm{t\bar{t}}}$ and $y^{\mathrm{t\bar{t}}}$. The many variables test the predictions in different areas, for example for a correct modelling of higher order corrections or the PDF set. A good event generator and/or QCD calculation should be able to predict all distributions correctly.

9.3 Visible phase space

In order to be more independent of the simulation, the differential cross sections presented here are only given for the phase space in which they can be measured (*visible phase space*) and no extrapolation to the full phase space like in section 9.1 is done. To achieve this, the selection efficiency determined from the Monte Carlo simulation is calculated only after applying cuts on generator level corresponding to the visible kinematic range. The cuts applied on generator level are the following (see chapter 7 for reconstruction level cuts):

$$p_T^{\mathrm{gen}}(\ell) > 20 \text{ GeV} \quad \text{and} \quad |\eta^{\mathrm{gen}}(\ell)| < 2.4 \tag{9.9}$$

$$p_T^{\mathrm{gen}}(b \text{ quark}) > 30 \text{ GeV} \quad \text{and} \quad |\eta^{\mathrm{gen}}(b \text{ quark})| < 2.4 \tag{9.10}$$

Here the phase space is defined on parton level before final state QCD radiation (corresponds to status 3 particles in MadGraph). The requirements are also placed on the b quark directly instead of a (generator-level) jet. The difference between these definitions of b jet is the treatment of the hadronisation modelling: only the cuts on the hadron level (generator jet) depend on it while the b quark cut doesn't. This means when correcting the result back to parton level instead of hadron level, one has to take gluon radiation and hadronisation effects into account. This has been done in this work. The parton level cut was also synchronised with the other analyses made public in the preliminary CMS results [141].

For the planned publication of the full 2011 data (shown in the chapter 10), this cut has been changed to hadron level, i.e. the cross section definition has changed. When applying the cut on the b jet one first needs to define unambiguously what a b jet exactly

is. Here, generator level b jets have been defined as those jets which contain a daughter particle from the b quark. If multiple jets exist which contain daughters of the same b quark, only the one with highest p_T is considered to be a b jet.

9.4 Normalisation

Given any of these differential cross sections, one can integrate over the corresponding variable and retrieve the total cross section (in the given phase space). In this work however only *normalised* differential cross sections are presented, i.e. the differential distribution is divided by the measured total cross section (in the visible phase space), so that the integral has value one. In the plots, some distributions (e.g. p_T) are cut off at a certain value so that the integral in the plot range is smaller than one.

The reason for the normalisation is that all uncertainties which only have an influence on the total cross section cancel out. Only uncertainties on the shape of a distribution remain. When comparing data to different predictions, it is thus easier to see if the shapes are modelled correctly. The cancelling of uncertainties is also described in chapter 8.

9.5 Bin-by-bin calculation

As any measurement, this one also doesn't have infinite detector resolution and statistics. Therefore, the cross sections cannot be measured differentially in a mathematical sense. Instead several bins of X are measured, and formula 9.8 is modified:

$$\frac{1}{\sigma}\frac{d\sigma}{dX} \longrightarrow \frac{1}{\sigma}\frac{\Delta\sigma}{\Delta X} \tag{9.11}$$

Combining this with equation 9.1 leads to the following formula to calculate the cross section for a given bin i (with ΔX^i being the bin width of bin i):

$$\frac{1}{\sigma}\frac{\Delta\sigma^i}{\Delta X^i} = \frac{1}{\sigma}\frac{N^i_{\text{Data}} - N^i_{\text{BG}}}{L \cdot \epsilon^i \cdot \Delta X^i} \tag{9.12}$$

Analogous to the total cross section calculation, N^i_{Data} is the number of measured events in bin i, which is corrected by the background estimate in that bin N^i_{BG} from the corrected simulation. The luminosity L appears twice: once explicitly and once inversely in the total cross section σ. Thus it cancels out completely and the results neither depend on the luminosity nor have an uncertainty associated to it. The efficiency ϵ^i corrects for detector acceptance and efficiencies in bin i in the visible phase space region as defined in equations 9.9 and 9.10.

As described in 9.1.1 for the total cross section, the statistical error is based on N^i_{Data}, while the whole analysis is rerun for all the systematic variations. The uncertainty is then determined separately in every bin of every distribution. Also here the three dilepton channels can be combined using their statistical uncertainties, and the systematic uncertainty is then determined from the variations of the combined result.

9.5.1 Purity, stability, and efficiency

While there can be many small bins in X for the simulation, the measurement might only allow a few larger bins. The reason is that the limited detector resolution leads to migration of events from one bin into other bins. For small mismeasurements, an event usually does not migrate at all. If it migrates, then mostly to an adjacent bin, but also migrations into bins further away are possible. Useful values to quantify migrations are the efficiency ϵ^i, the purity p^i, and the stability s^i which are determined from the simulation:

$$\epsilon^i = \frac{N^i_{\text{rec}}}{N^i_{\text{gen,tot}}} \tag{9.13}$$

$$p^i = \frac{N^i_{\text{gen and rec}}}{N^i_{\text{rec}}} \tag{9.14}$$

$$s^i = \frac{N^i_{\text{gen and rec}}}{N^i_{\text{gen}}} \tag{9.15}$$

The efficiency describes the fraction of those events which were reconstructed in the bin i (N^i_{rec}) out of the total number of events which were totally generated in that bin ($N^i_{\text{gen,tot}}$). The total denotes that it is the number before the selection, i.e. ϵ^i contains the detector acceptance and reconstruction efficiency in the visible phase space.

For purity and stability, only events which pass the selection are considered. The purity is the quotient of the number of events which were generated and reconstructed in a bin $N^i_{\text{gen and rec}}$ over the number of events which were reconstructed in that bin N^i_{rec}. It is a measure for the migration into the bin: the maximum of $p^i = 1$ is achieved when there is no migration from events into the bin i, and the minimum of $p^i = 0$ corresponds to the situation where all events reconstructed in the bin i were generated in other bins.

The stability quantifies the migration out of a bin. It is defined as the number of events which are generated and reconstructed in a bin $N^i_{\text{gen and rec}}$ over the number of generated events N^i_{gen}. Thus it ranges from $s^i = 1$, meaning no event migrates out of the bin to $s^i = 0$ where all events generated in bin i have migrated to other bins.

In figure 9.1 the correlation plot for the transverse momentum of the leptons is shown on the left. It is a two-dimensional histogram showing the relation of the reconstructed value on the y axis with respect to its true value on the x axis. A measure for the linear dependence of these two quantities is given by the correlation coefficient ϱ which can take values from -1 (anti-correlated) to 1 (100 % correlated, optimal case). The correlation can be calculated with the following formula (cov is the covariance) [142]:

$$\varrho(X^{\text{reco}}, X^{\text{gen}}) = \frac{\text{cov}(X^{\text{reco}}, X^{\text{gen}})}{\sigma(X^{\text{reco}}) \cdot \sigma(X^{\text{gen}})} \tag{9.16}$$

To calculate purity and stability, the correlation plot is binned, in this case into 6 bins and formulas 9.14 and 9.15 can be applied. The result can then be seen on the right side of figure 9.1. In addition, the efficiency is shown.

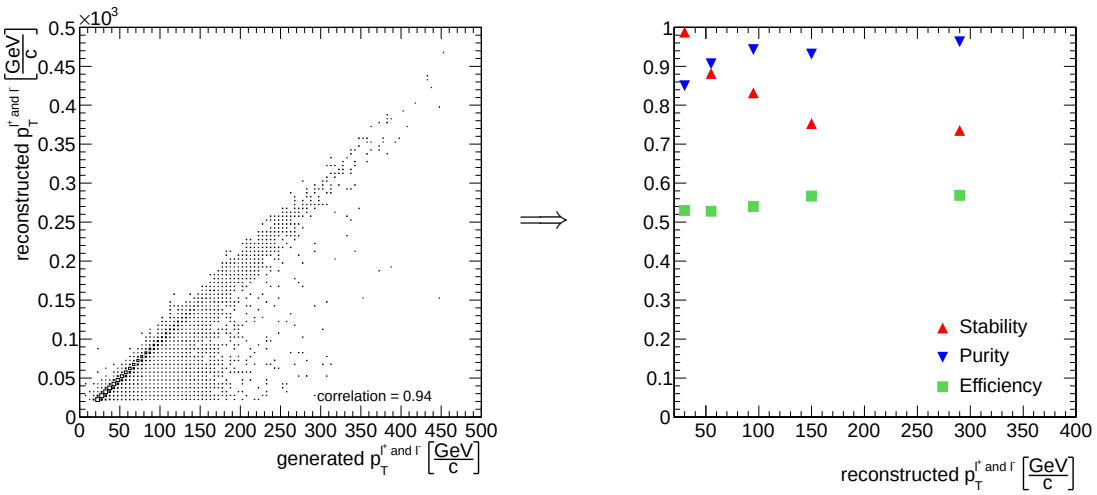


Figure 9.1: Left: correlation plot for the p_T of both leptons. The x axis shows the *true* p_T and the y axis the corresponding value after the reconstruction. The correlation coefficient is 0.94. Right: purity, stability as calculated from the correlation and efficiency are shown. The plots show the combination of the three decay channels (i.e. the correlation plot is the sum and the purity/stability/efficiency plot shows the weighted average).

When choosing the binning, one has to take into account several things. On one hand the binning should be as fine as possible to measure really differentially. On the other hand fine binnings have disadvantages: the migration effects are larger and thus purity and stability get smaller. And in low statistics regions the uncertainty will become large for fine bins. So the bins have finally been chosen such that they have (with few exceptions) better than 60 % purity and stability. Additional considerations were to have an even number of bins in symmetric distributions (η, y) and that the bin boundaries should have "round" numbers, i.e. for the p_T the bin boundaries were placed at integer multiples of 10 GeV, for η and y only values with one digit after the decimal point were considered.

After the binning has been chosen, one can plot the event yield which is simply a count of the number of events which were measured in the specific bins. The result is shown in figure 9.2 (left).

9.5.2 Cross section calculation

Finally equation 9.12 is used to convert the event yield into the differential cross section as shown on the right of figure 9.2. The plot shows the transverse momentum of the leptons on the abscissa and the normalised differential cross section on the ordinate. To compare the results with QCD expectations, three different predictions are shown: the

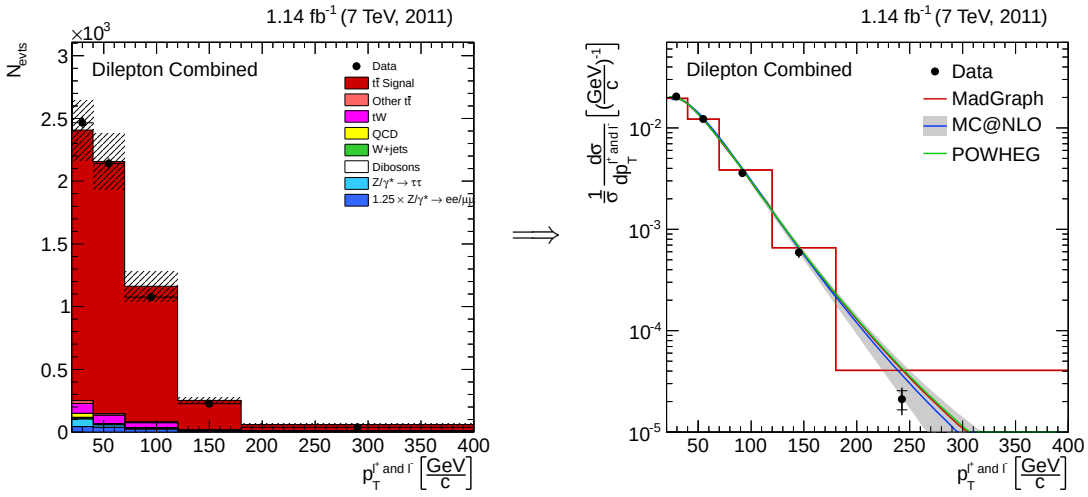


Figure 9.2: Left: yield plot for the p_T of the leptons. The hatched error band corresponds to the uncertainties on the luminosity and the $t\bar{t}$ cross section. Right: the resulting cross section plot with comparison to MadGraph (red histogram and red line), MC@NLO, and Powheg. The error band on MC@NLO includes the uncertainty on the PDF, on m_t and on the Q^2 scale variation. The inner error bars for the data points correspond to the statistical uncertainty, the outer error is the total uncertainty. The plots show the combination of the three decay channels.

green line shows the results from Powheg and the blue line corresponds to MC@NLO. The error band comprises the uncertainty on the PDF determined using PDF4LHC recommendations [143] and the variations on the top quark mass and Q^2 scale. The MadGraph simulation on which the bin by bin calculation is based is shown twice: in a smooth curve and as a binned histogram. The bins of the histogram correspond to the bins of actual measurement which only determines a delta cross section (see equation 9.11), i.e. the bin values quantify the normalised bin-average cross section. The MadGraph curve on the other hand was created using a very fine binning and thus corresponds to the differential cross section.

The measured data is shown using black points with two error bars. The inner error bars correspond to the statistical uncertainty only, the outer error bars visualise the squared sum of the statistical and systematic uncertainty.

9.5.3 Placement of data points

The measurement only provides information about the average cross section in the whole bin but no information about the position inside the bin where the average corresponds to the differential cross section. To avoid putting the point into the middle of the bin

with error bars in x corresponding to the bin size, horizontal bin centre corrections are applied [144]. The bin centre correction places the data point to that position where the bin-average cross section equals the differential cross section according to the Monte Carlo prediction. No uncertainty is assigned to this position.

The bin centres are purely determined from the MadGraph simulation. Therefore on can read off bin centres using the binned and smooth MadGraph curves: the bin centres correspond to those positions where the binned histogram and the smooth curve intersect. This is particularly visible in the two rightmost bins of the $p_T^{\ell^+ \text{ and } \ell^-}$ distribution: the bin centre correction moves the data points to the left.

One could also use other generators for the bin-by-bin efficiency calculation and then later for the bin centre corrections. In this work, Powheg and MC@NLO are shown to compare with the MadGraph generator and of course with data.

The advantage of the bin centre correction is that it makes it easier to compare the measurement to a theory prediction as the measured cross sections can now be assigned an exact position corresponding to a differential measurement. When another measurement is done in a different binning and bin centre corrections are applied as well, one can put all measured points into the same plot and thus compare more easily.

However, one needs to take care of situations in which the smooth curve intersects the binned histogram more than once. In that case one can choose any one of the intersection points. As this can easily lead to confusion, the binning is chosen such that this does not happen. This is also one of the reasons for an even number of bins in symmetric distributions.

9.5.4 Discussion of the result

From the differential cross section shown in figure 9.2, one can learn that the theory describes the data very precisely. The predicted spectrum seems to be just a little bit softer but on the other hand the deviations from the data are within the uncertainties. Also the three generators, MadGraph, MC@NLO, and Powheg predict almost identical distributions and cannot be distinguished from each other. Therefore this quantity is described equally well with a tree level calculation with added parton showering and the next-to-leading order calculations.

9.6 Differential sections in further quantities

In the following the corresponding four plots, i.e. the correlation, the purity, stability and efficiency, the yield, and the differential cross section are shown for the other eight quantities. This chapter only includes the plots in which all three decay channels have been combined because the results in the different channels are similar and the final result is dominated by the $e\mu$ channel anyway. The plots for the separate channels can

be found in appendix A. The combination plots are repeated in the appendix so that a direct comparison is possible.

Furthermore, the results for all lepton quantities are summarised in table 9.2 and all results for top quantities can be found in table 9.3. These tables list the bin centres, the measured cross sections and the uncertainties in all bins individually.

9.6.1 Lepton pseudorapidity

The lepton pseudorapidity has an excellent correlation between reconstructed and generated value, i.e. the direction of the leptons is measured very precisely. Therefore purity and stability are larger than 95 % in all bins, as shown in figure 9.3. A very good agreement between data and the predictions is found and again the three calculations predict similar distributions. MadGraph produces the leptons a bit more centrally than the other generators but the effect is smaller than the statistical uncertainty of the data.

The data is a little bit asymmetric, specifically the bin between $\eta = 0$ and 0.6 seems to have fluctuated downward while the bin between $\eta = 1.2$ and 1.8 has fluctuated upward not only with respect to the simulation but more importantly with respect to the negative η bins where the agreement between data and simulation is very good. Ignoring the uncertainties, it seems that the data is a bit less central than the predictions and thus disfavouring MadGraph compared to MC@NLO or Powheg.

If one considers the individual channels (figure A.2), one can see that this effect is larger in the channels with muons while the *ee* channel is more symmetric. This is probably a statistical effect, but if it remains present with more statistics, it could also give a hint to a problem with the simulation of the muon system because there is no known physics process which could lead to such an asymmetry in one bin. More data is needed here to draw conclusions about this.

The very high purity and stability in this distribution may also be used to measure in even more bins. To keep the statistical uncertainties small, this should only be done once more integrated luminosity is available.

9.6.2 Lepton pair transverse momentum

The lepton pair transverse momentum is an interesting quantity because it is sensitive to higher order QCD effects. The correlation between the generated and reconstructed pairs is good, and most events are in the region below 100 GeV as visible in figure 9.4. Therefore this region was divided into five bins while only two bins cover the larger region up to 400 GeV. Stability and purity are about 80 to 90 %, only the first small bin has a slightly lower purity.

In this plot again a good agreement of the predictions with the data is found. Due to limited statistics the tail of the distribution has been smoothened (see appendix C for an explanation how the smoothing is done). One can see that MadGraph predicts a softer distribution than MC@NLO or Powheg which both lead to almost identical, slightly

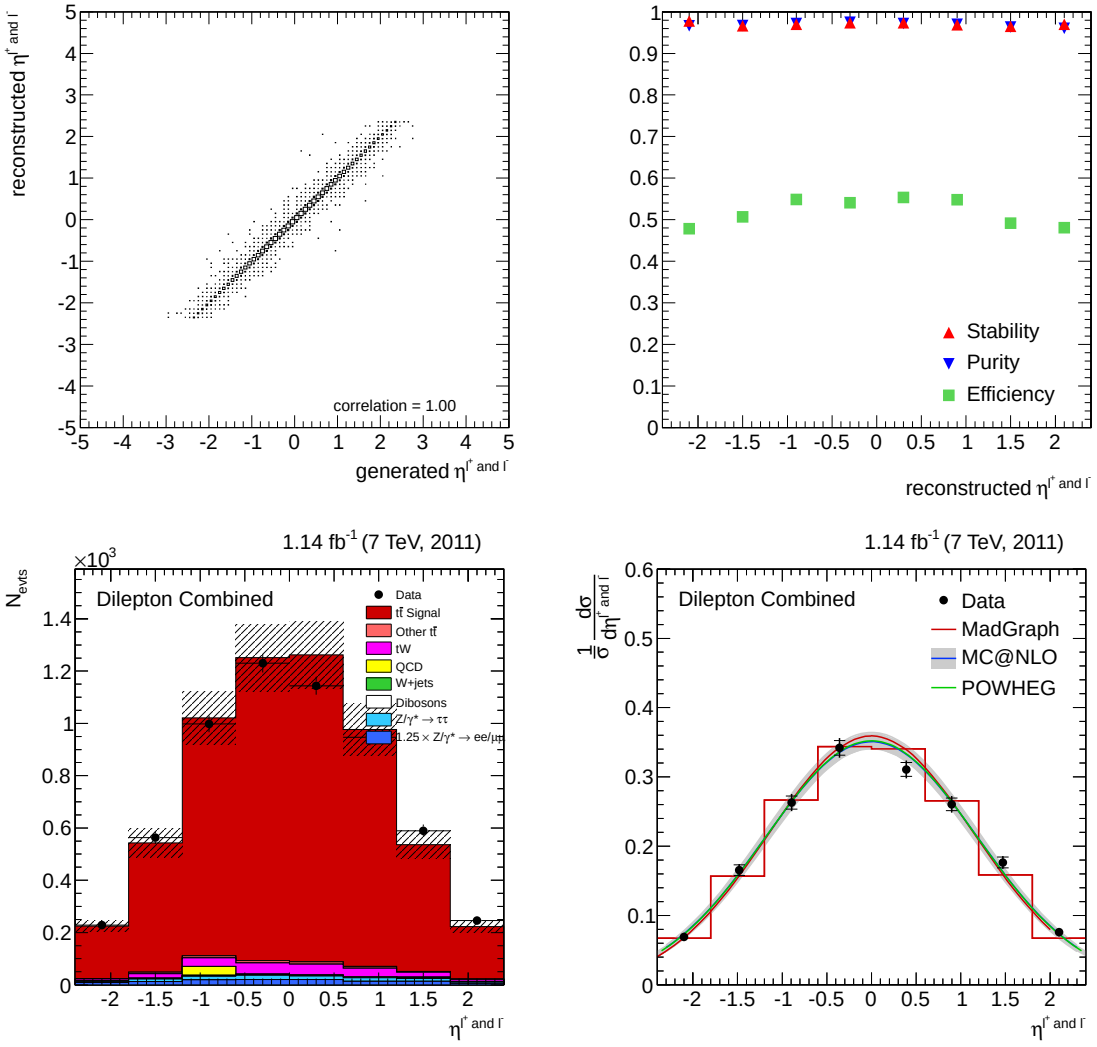


Figure 9.3: Top row: correlation between generated and reconstructed $\eta^{\ell+\,\text{and}\,\ell-}$ and calculated purity, stability and efficiency. Bottom row: event yield on the left and the normalised differential cross section on the right.

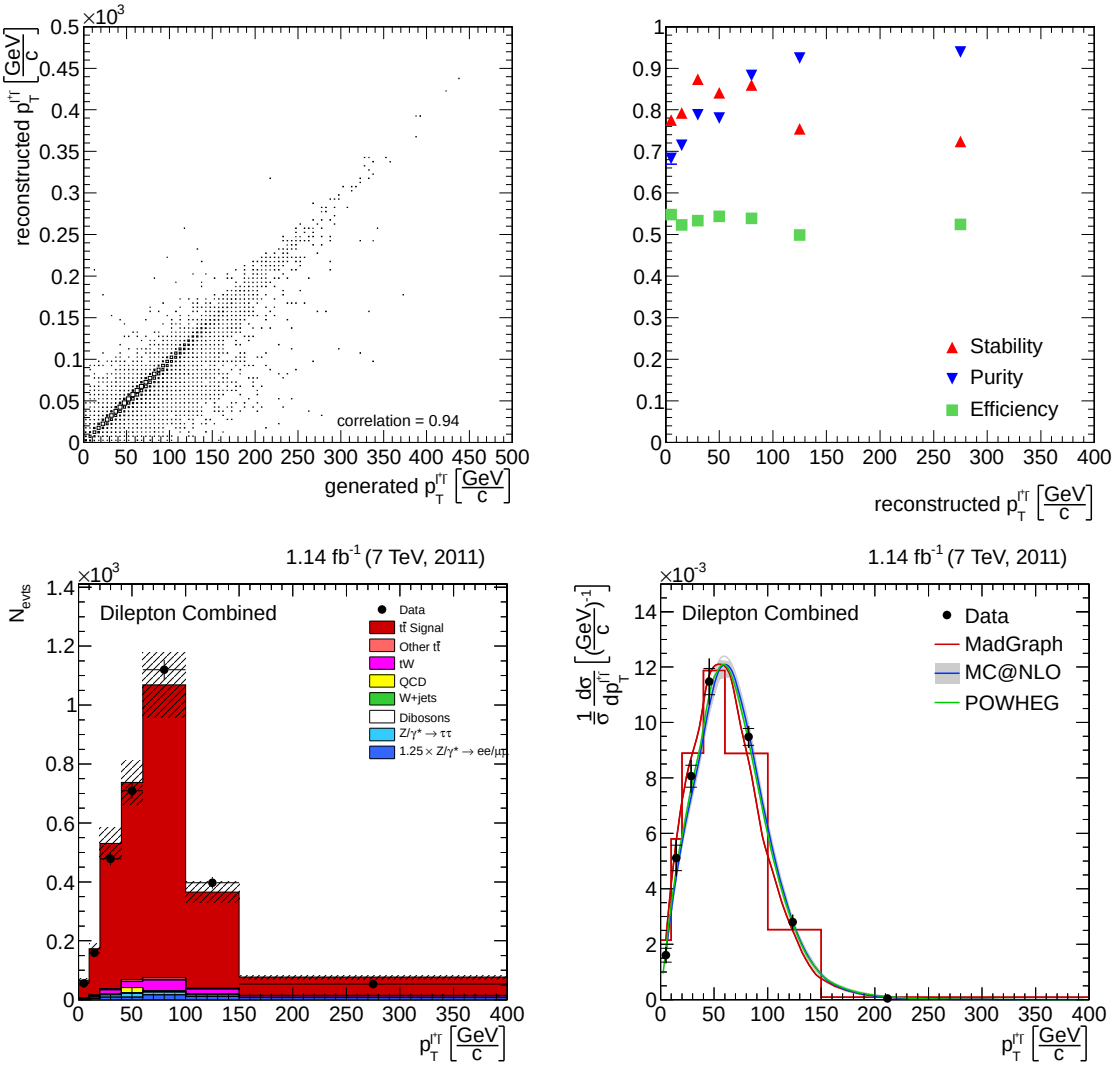


Figure 9.4: Top row: correlation between generated and reconstructed $p_T^{\ell\ell}$ and calculated purity, stability and efficiency. Bottom row: event yield on the left and the normalised differential cross section on the right. In the appendix (figure A.3 on page 150) the cross section is shown in logarithmic scale.

harder distributions. The data seems to prefer a harder spectrum than what MadGraph predicts, but maybe a softer spectrum than the NLO predictions. Nevertheless, the NLO predictions fit the data better.

9.6.3 Lepton pair mass

Like the other lepton quantities, also the lepton pair mass has a high correlation between the generated and reconstructed values and thus a high purity and stability above 80 % as shown in figure 9.5. The binning in lepton pair mass has a constraint due to the cut on the Z mass in the ee and $\mu\mu$ channel: there is one bin from 76 to 106 GeV covering this veto region—filled only by contributions from the $e\mu$ channel. For the rest of the distribution, all three channels are used. This also explains the drop of efficiency and the drop in the event yield in that bin. The regions below 76 and above 106 GeV are then split into two bins each.

For the invariant lepton pair mass the three predictions are again so similar that one can barely distinguish between them and furthermore the data again proves that their quality is very good.

Also in this lepton distribution the number of bins could be increased in future measurements. This could be interesting because new resonances would not only be visible in $m^{\mathrm{t\bar{t}}}$ but also in this lepton-based distribution. While the leptons are measured with high accuracy, the resolution of a possible peak would not be good due to the decay process. One would have to study if this distribution has enough sensitivity nonetheless.

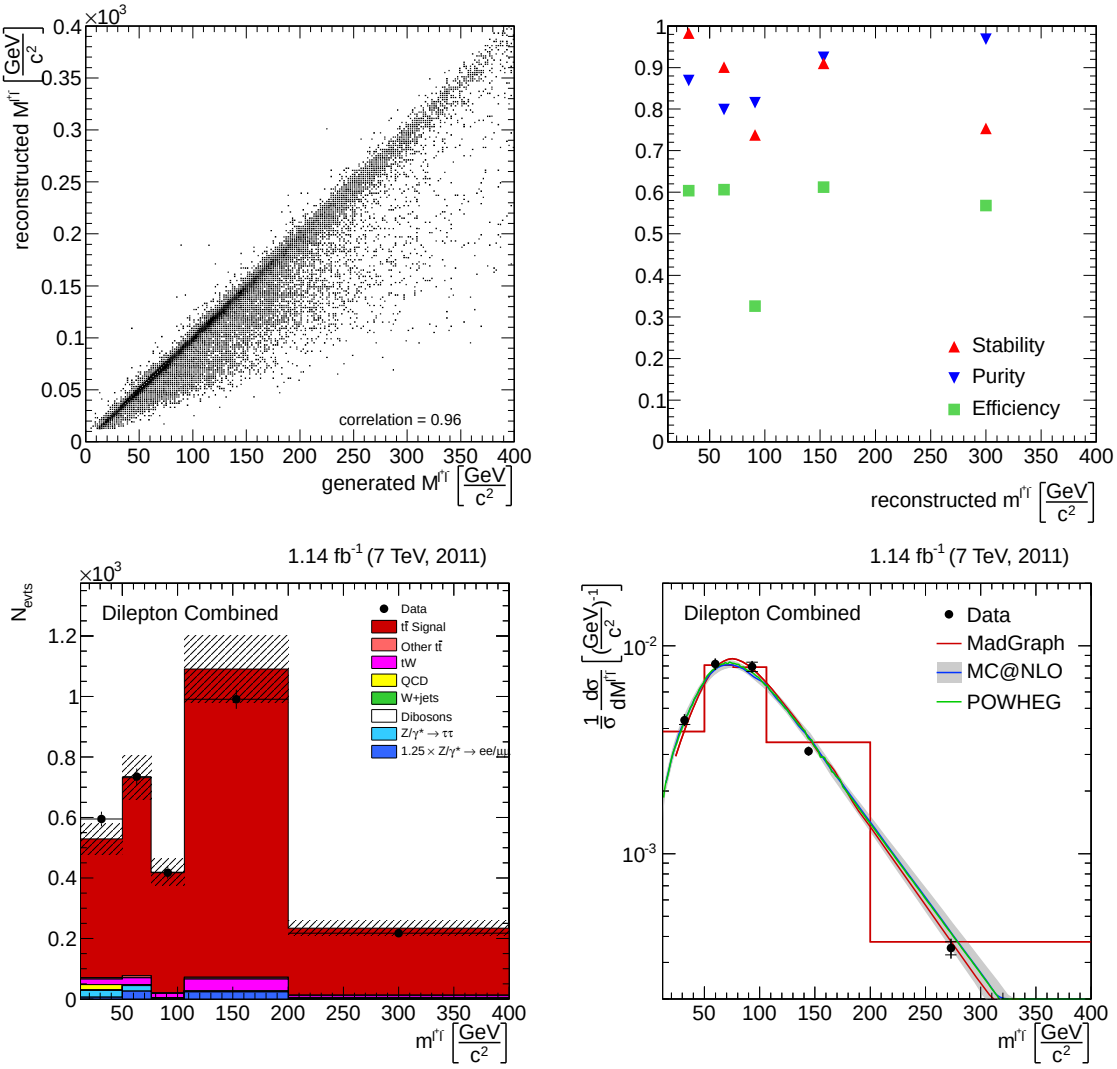


Figure 9.5: Top row: correlation between generated and reconstructed $m^{\ell\ell}$ and calculated purity, stability and efficiency. Bottom row: event yield on the left and the normalised differential cross section on the right.

Lepton transverse momentum

bin centre [GeV]	bin [GeV]	$1/\sigma d\sigma/dp_T^{\ell^+ \text{ and } \ell^-}$	stat. [%]	sys. [%]	total [%]
29	20 to 40	0.02040	2.2	4.0	4.6
55	40 to 70	0.01230	2.3	2.5	3.4
92	70 to 120	0.00359	3.3	4.6	5.6
145	120 to 180	0.00059	7.3	9.1	11.7
243	180 to 400	0.00002	21.5	21.5	30.4

Lepton pseudorapidity

bin centre	bin	$1/\sigma d\sigma/d\eta^{\ell^+ \text{ and } \ell^-}$	stat. [%]	sys. [%]	total [%]
−2.10	−2.4 to −1.8	0.06927	7.4	3.8	8.3
−1.48	−1.8 to −1.2	0.16551	4.6	2.9	5.4
−0.89	−1.2 to −0.6	0.26305	3.5	3.1	4.7
−0.36	−0.6 to 0.0	0.34199	3.1	2.7	4.1
0.39	0.0 to 0.6	0.31084	3.2	2.7	4.2
0.90	0.6 to 1.2	0.26033	3.5	3.0	4.6
1.47	1.2 to 1.8	0.17663	4.5	2.9	5.3
2.10	1.8 to 2.4	0.07596	7.0	3.9	8.0

Lepton pair transverse momentum

bin centre [GeV]	bin [GeV]	$1/\sigma d\sigma/dp_T^{\ell\ell}$	stat. [%]	sys. [%]	total [%]
5	0 to 10	0.00161	15.6	9.7	18.4
15	10 to 20	0.00511	9.0	9.3	12.9
29	20 to 40	0.00807	4.9	5.5	7.4
45	40 to 60	0.01148	4.1	6.0	7.3
82	60 to 100	0.00948	3.2	2.7	4.2
123	100 to 150	0.00280	5.5	8.1	9.8
212	150 to 400	0.00005	19.7	20.5	28.4

Lepton pair mass

bin centre [GeV]	bin [GeV]	$1/\sigma d\sigma/dM^{\ell\ell}$	stat. [%]	sys. [%]	total [%]
32	12 to 50	0.00439	4.7	4.8	6.7
60	50 to 76	0.00817	4.1	4.8	6.3
93	76 to 106	0.00794	5.1	3.9	6.4
144	106 to 200	0.00312	3.4	3.5	4.9
273	200 to 400	0.00035	7.2	8.3	10.9

Table 9.2: Summary of the normalised differential cross section as a function of the lepton and lepton pair observables.

9.6.4 Top quark transverse momentum

After having discussed the lepton quantities in the previous section, now the top quark distributions are shown. As described in section 7.12, the top quark quantities need to be reconstructed with a kinematic event reconstruction. The first top distribution shown here is the p_T distribution for the top quarks. The first apparent difference is that the correlation between the generated and reconstructed $p_T^{t\,\mathrm{and}\,\bar{t}}$ is 0.67 and thus smaller than for all the lepton quantities (see figure 9.6).

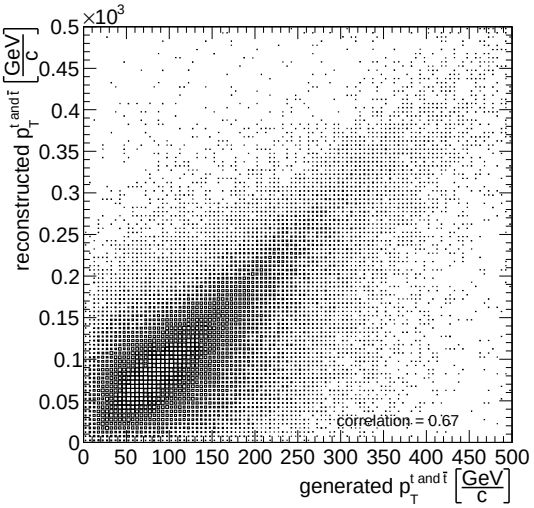


Figure 9.6: Correlation between generated and reconstructed $p_T^{t\,\mathrm{and}\,\bar{t}}$.

This is due to the underconstrained event information due to the unmeasured neutrinos. As a consequence, purity and stability are also smaller and only in the order of 50 %. Therefore tests about the introduced bias have been made, see appendix D. The outcome is that only a small bias is introduced as long as the simulation describes the data reasonably, which is the case here. In figure 9.7, the normalised cross section plot is shown twice, in logarithmic and in linear scale.

The top quark transverse momentum itself is sensitive to higher order corrections and possibly to processes beyond the Standard Model, but again the data agrees well with all three predictions. It could be that the simulation overestimates the transverse momentum, i.e. that the data is softer. However, the uncertainties are too large to prove that.

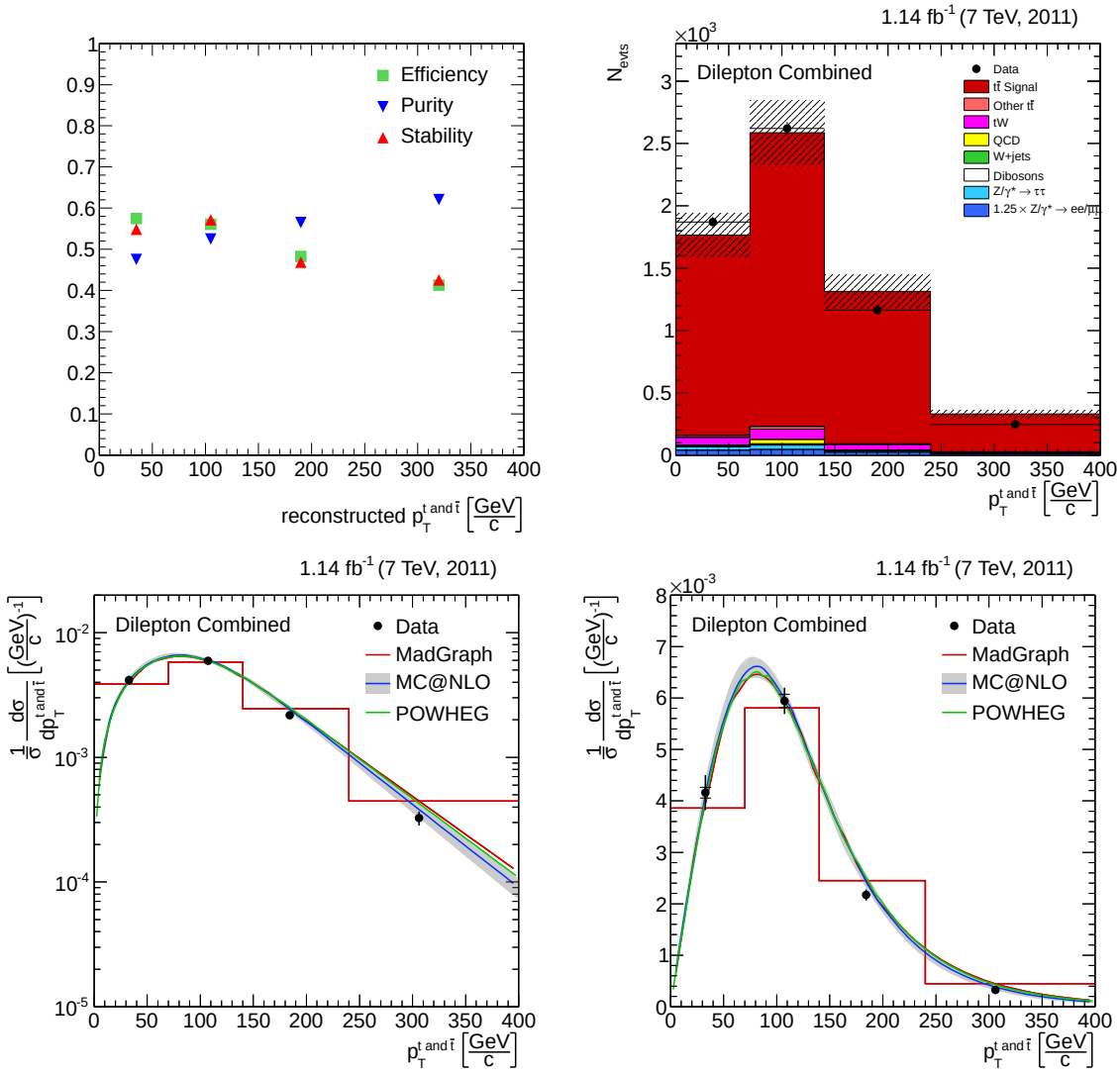


Figure 9.7: Top row: calculated purity, stability and efficiency (left) and event yield (right). Bottom row: the normalised differential cross section in logarithmic (left) and linear (right) scale.

9.6.5 Top quark rapidity

For the rapidity of the top quarks the correlation is 0.80, i.e. larger than for the p_T. In the top right plot of figure 9.8 one can see that the efficiency becomes small for larger rapidities, i.e. the measurement works better for central top quarks. This can also be seen in less than 30 % stability in the outermost bins, i.e. here many events migrate out of these bins and into the central bins. As the number of events in the outer bins is small, the purity is not much affected and stays in the order of 60 % in all bins.

In the event yield and resulting cross section, an excellent agreement between data and the predictions is found. The three generators behave similarly although MadGraph has a slightly more central shape than the NLO calculations MC@NLO and Powheg which lie on top of each other. The measured data however cannot tell them apart. A central bin at $|y| < 0.2$ could help here but on the other hand the difference between MadGraph and the other two generators is fully covered by their uncertainty.

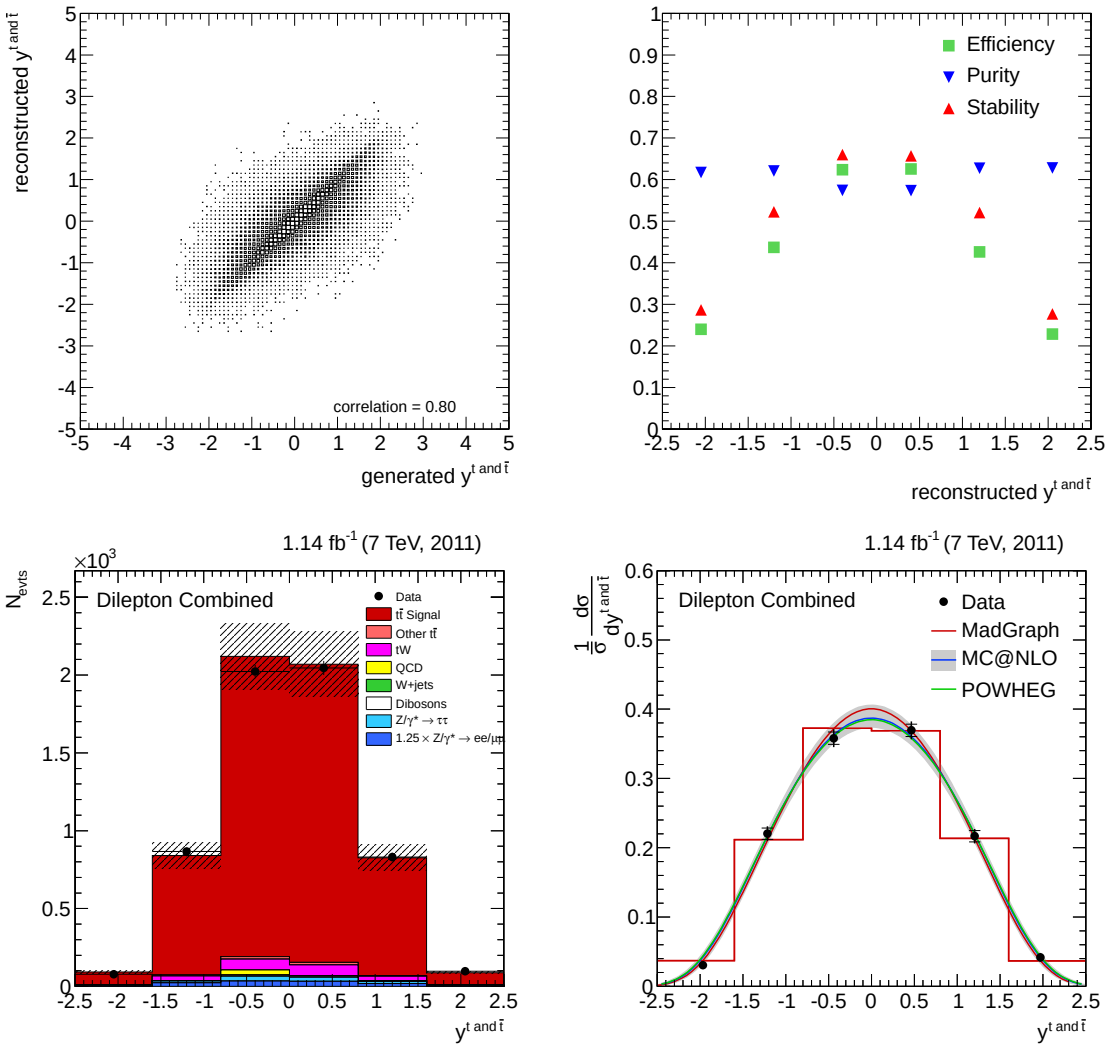


Figure 9.8: Top row: correlation between generated and reconstructed $y^{t\,\mathrm{and}\,\bar{t}}$ (rapidity for top and antitop quarks) and calculated purity, stability and efficiency. Bottom row: event yield on the left and the normalised differential cross section on the right. MC@NLO and POWHEG lie on top of each other.

9.6.6 Top quark pair transverse momentum

The remaining three quantities are the ones determined from the top-antitop pair. The first top pair cross section shown here is its transverse momentum $p_T^{t\bar{t}}$. The correlation in this variable is 0.75, purity and stability are about 60 % for the chosen binning.

Again the measured normalised differential cross section is in excellent agreement with all three predictions, as can be seen in figure 9.9.

Using the $p_T^{t\bar{t}}$ one can check if higher order corrections are simulated correctly. Without the higher orders, the top quark pair would always be produced without any transverse momentum. Obviously this isn't the case and the predictions show that the cross section is largest at approximately $p_T^{t\bar{t}} = 18\,\text{GeV}$ and falls off fast for larger transverse momenta. As the data supports this, it can thus be concluded that all three predictions are sufficiently precise.

9.6.7 Top quark pair rapidity

Similar remarks as made for the top quark rapidity also apply for the rapidity of the top quark pair. Not only the correlation for the rapidity, here 0.89, is again larger than the one for the transverse momentum, but also the purities, stabilities and efficiencies look similar (see figure 9.10). While the purity is about 70 % in all bins, the stability and efficiency go down to only 20 % in the two outermost bins.

The top quark pair rapidity is sensitive to the gluon PDF at high x, but once again the agreement between data and the predictions is excellent, and all the three predictions are exactly on top of each other. No hints for problems in the PDFs or other parts of the calculations are found in this distribution.

9.6.8 Top quark pair invariant mass

The last quantity in which the normalised differential cross sections are measured is the invariant mass of the $t\bar{t}$ system. As described already in section 9.2, this quantity is sensitive to new resonances which could decay into a top and antitop quark pair.

The correlation between the generated and reconstructed invariant $t\bar{t}$ mass is only 0.62 and therefore also the purity and stability are rather small, i.e. in the order of only 40 %, see figure 9.11.

Also in this distribution, the agreement between data and the predictions is very good and nothing unexpected is found.

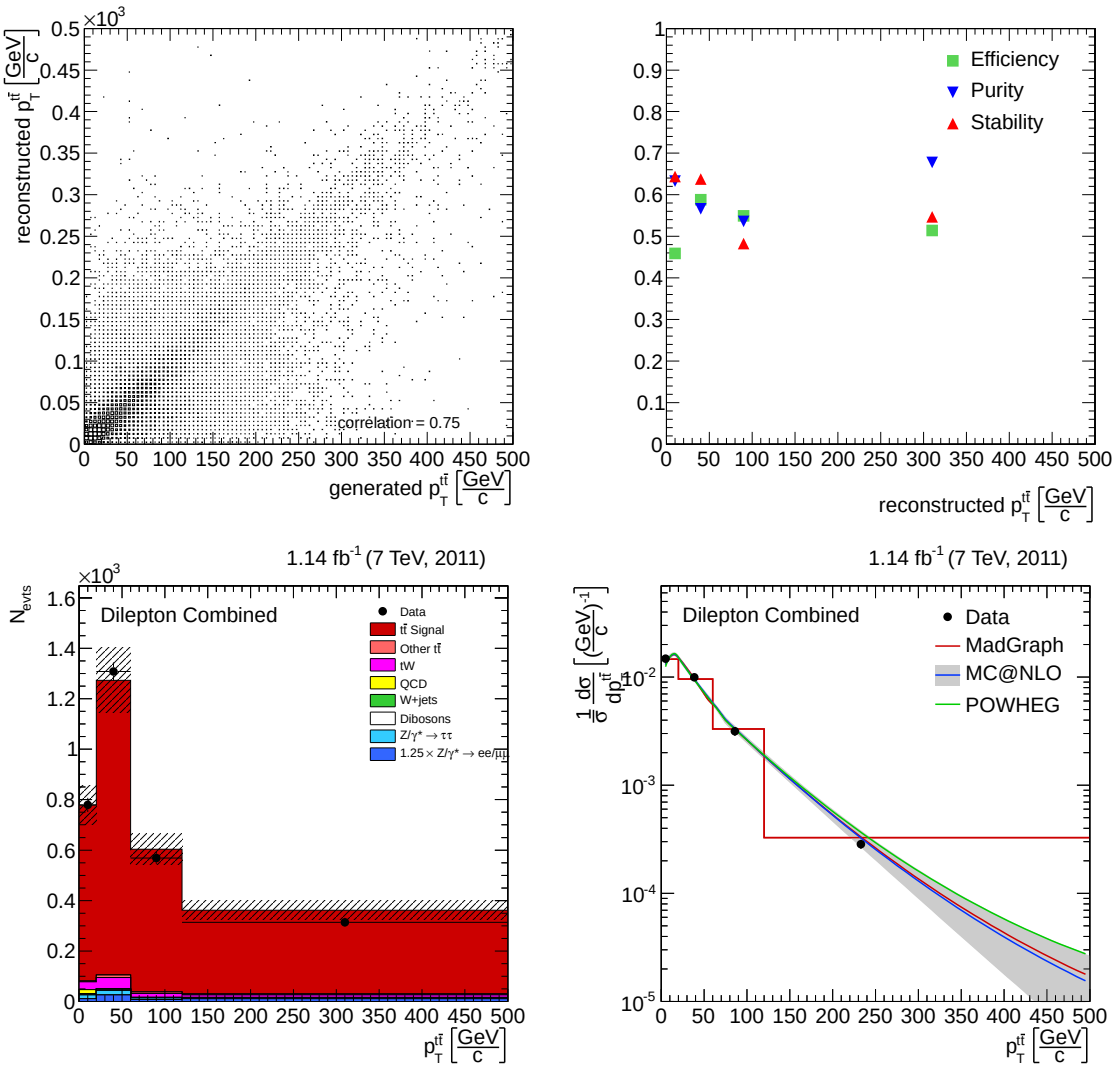


Figure 9.9: Top row: correlation between generated and reconstructed $p_T^{t\bar{t}}$ (transverse momentum for top and antitop quarks) and calculated purity, stability and efficiency. Bottom row: event yield on the left and the normalised differential cross section on the right.

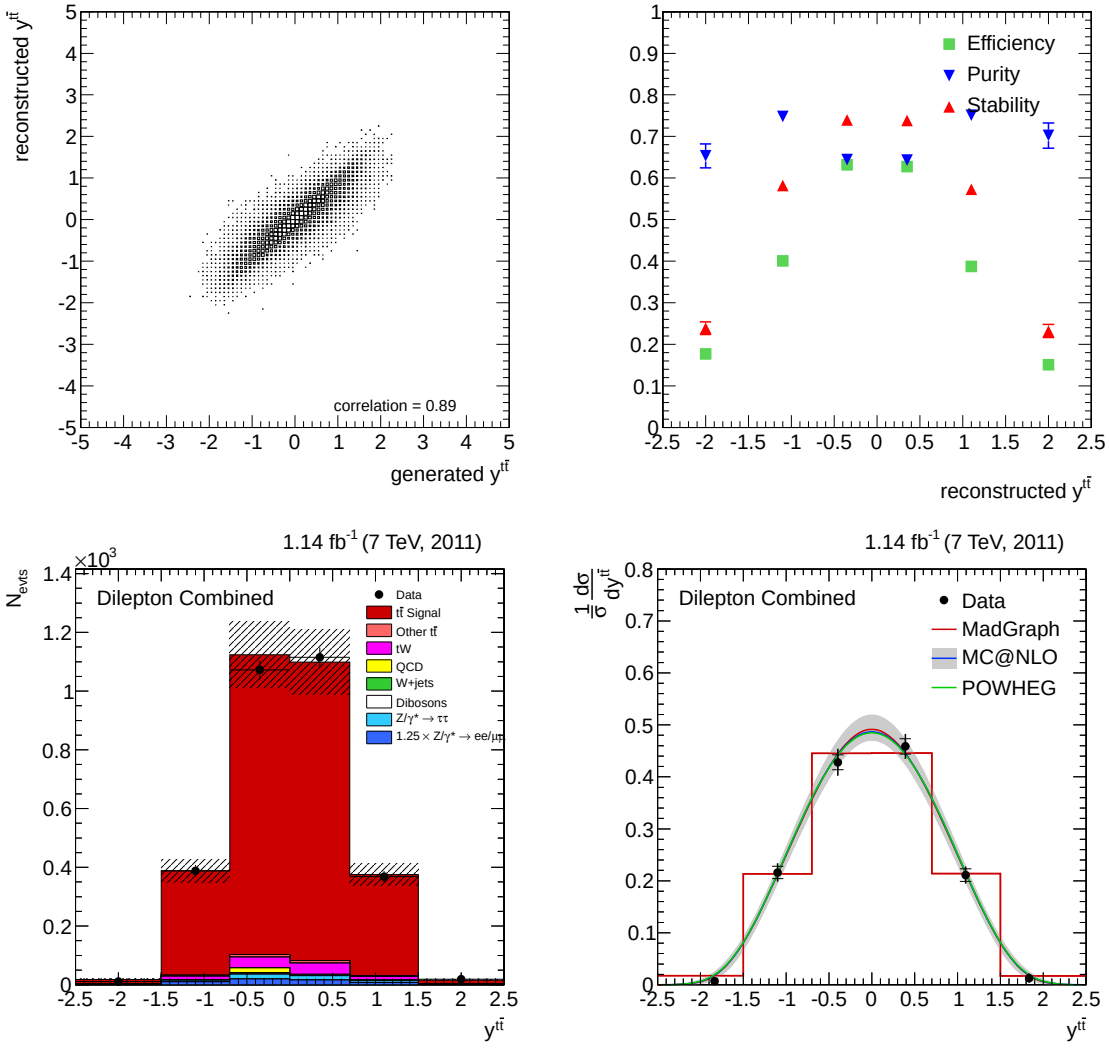


Figure 9.10: Top row: correlation between generated and reconstructed $y^{t\bar{t}}$ and calculated purity, stability and efficiency. Bottom row: event yield on the left and the normalised differential cross section on the right.

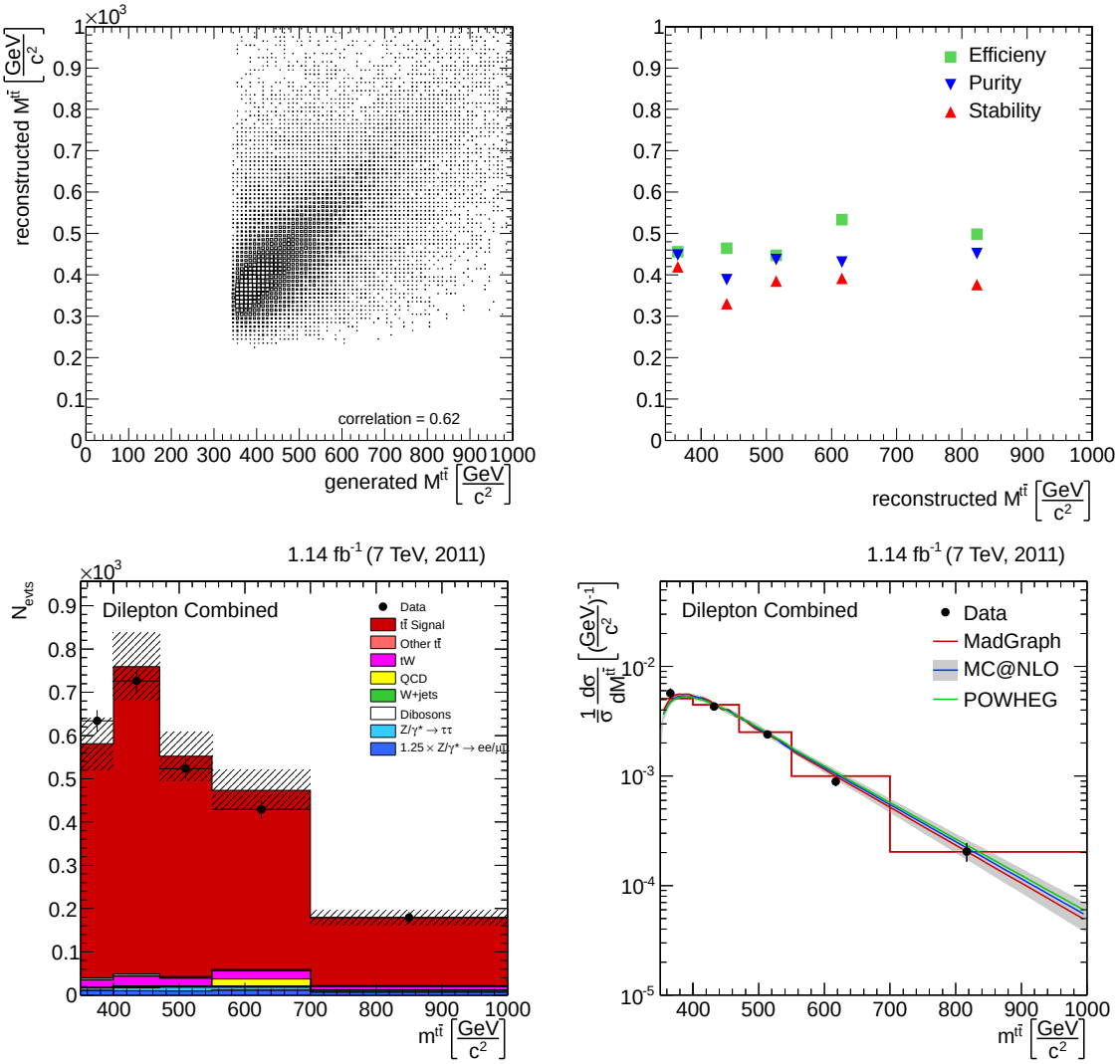


Figure 9.11: Top row: correlation between generated and reconstructed $m^{t\bar{t}}$ and calculated purity, stability and efficiency. Bottom row: event yield on the left and the normalised differential cross section on the right.

Top quark transverse momentum

bin centre [GeV]	bin [GeV]	$1/\sigma d\sigma/p_T^{t \text{ or } \bar{t}}$	stat. [%]	sys. [%]	total [%]
33	0 to 70	0.00416	2.5	7.8	8.2
107	70 to 140	0.00595	2.1	3.8	4.3
184	140 to 240	0.00217	3.2	3.9	5.0
306	240 to 400	0.00033	7.1	10.6	12.8

Top quark rapidity

bin centre	bin	$1/\sigma d\sigma/y^{t \text{ or } \bar{t}}$	stat. [%]	sys. [%]	total [%]
−1.97	−2.5 to −1.6	0.03062	12.9	8.5	15.5
−1.21	−1.6 to −0.8	0.22050	3.7	2.9	4.7
−0.44	−0.8 to 0.0	0.35821	2.4	2.6	3.6
0.46	0.0 to 0.8	0.36954	2.4	2.6	3.5
1.20	0.8 to 1.6	0.21684	3.8	2.9	4.8
1.97	1.6 to 2.5	0.04196	11.2	9.8	14.9

Top quark pair transverse momentum

bin centre [GeV]	bin [GeV]	$1/\sigma d\sigma/dp_T^{t\bar{t}}$	stat. [%]	sys. [%]	total [%]
5	0 to 20	0.01480	4.0	6.5	7.7
39	20 to 60	0.00997	3.0	4.2	5.1
86	60 to 120	0.00316	4.5	7.9	9.0
233	120 to 500	0.00029	6.2	7.2	9.5

Top quark pair rapidity

bin centre	bin	$1/\sigma d\sigma/y^{t\bar{t}}$	stat. [%]	sys. [%]	total [%]
−1.83	−2.5 to −1.5	0.00697	46.4	48.2	66.9
−1.10	−1.5 to −0.7	0.21616	5.5	6.1	8.2
−0.40	−0.7 to 0.0	0.42810	3.4	5.1	6.1
0.39	0.0 to 0.7	0.45884	3.2	4.1	5.3
1.09	0.7 to 1.5	0.21104	5.7	5.3	7.8
1.84	1.5 to 2.5	0.01290	38.1	58.3	69.6

Top quark pair mass

bin centre [GeV]	bin [GeV]	$1/\sigma d\sigma/dM^{t\bar{t}}$	stat. [%]	sys. [%]	total [%]
364	345 to 400	0.00542	4.0	9.9	10.7
438	400 to 475	0.00425	3.8	3.5	5.2
514	475 to 550	0.00233	5.0	5.4	7.3
617	550 to 700	0.00089	5.6	7.6	9.4
817	700 to 1000	0.00020	8.5	17.4	19.4

Table 9.3: Summary of the normalised differential cross section as a function of top quark and top quark pair observables.

9.7 Summary

After having shown nine different distributions, one can say that all three different predictions describe the data very well. The two NLO calculations MC@NLO and POWHEG which differ in the added parton showering predict almost exactly the same distributions and thus cannot be distinguished from each other in any of the cross section plots. The MADGRAPH prediction produces slightly more central lepton and top quark rapidity distributions. For the lepton rapidity the NLO predictions agree better with the data, for the top quark rapidity no preference can be seen with the current level of precision. Differences are also visible in the transverse momentum distributions for the $\ell\bar{\ell}$ system, where MADGRAPH is softer than the NLO predictions. Here, the data is found in between and does not have a clear preference.

In summary, all predictions describe the data well, but there seems to be a tiny preference for the NLO calculations in some distributions.

The differential cross sections calculated in this work (except for $d\sigma/dm^{\mathrm{t\bar{t}}}$) were released to the public in February 2012 as a CMS preliminary result [141]. In the same week, the cross sections were already shown at the LC Forum [145]. Since then, they have also been shown at various other conferences, for example at the PLHC 2012 [146] (presented by myself) and the ICHEP 2012 [147].

9.8 Application in the $Z/\gamma + b$ measurement

The measurement of differential cross sections can not only be used to compare the results with different simulations or to search for new resonances, but it is also useful for the background determination in other analyses.

The measurement of the $Z/\gamma^* + b$ jet cross section [148] can serve as an example. In those events where the Z decays into ee or $\mu\mu$, the event topology of the $Z + b$ is similar to the dileptonic top quark decay: In both cases two oppositely charged isolated leptons and a b jet are found. The main difference is of course the transverse momentum imbalance which is present only in top quark events. However, due to mismeasurements also $Z + b$ events can have non-vanishing $E_{\mathrm{T}}^{\mathrm{miss}}$. Therefore the $Z + b$ events are an important background for this top quark measurement and the invariant mass cut from 76 to 106 GeV has been introduced. However, one can also look at it the the other way round: the dileptonic top quark events are the main background to the $Z + b$ cross section measurement!

Figure 9.12 shows the event yield after the final selection in the $Z + b \to \mu\mu X$ analysis [149]. In that measurement, the Z window is defined in a larger window from 60 to 120 GeV, corresponding to the first three bins.

To estimate the contamination of $\mathrm{t\bar{t}}$ events in these three bins, a ratio of $\mathrm{t\bar{t}}$ events outside (i.e. in the upper sideband, $M_{\mu\mu} > 120$ GeV) and inside the Z window was calculated. Then the ratio can then be corrected using the normalised differential cross

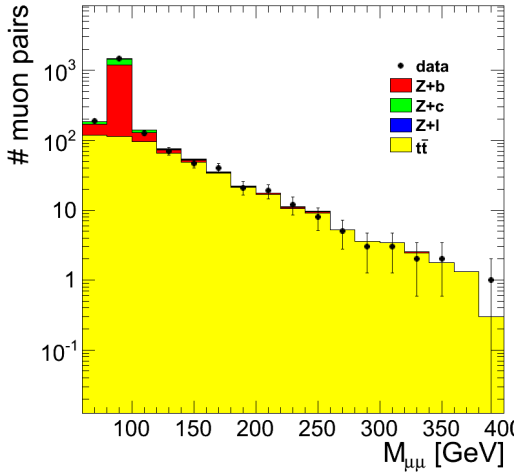


Figure 9.12:
Event yield of the invariant dilepton mass the $Z/\gamma + b$ measurement [149]. The colours used in this plot are different from the rest of this work: the $t\bar{t}$ contribution is shown in yellow. The Z and $t\bar{t}$ simulations are the same as in this work.

section in the invariant dilepton mass in the $e\mu$ channel: also there the cross section inside the Z window and in the upper sideband are calculated, see figure 9.13.

The different selection cuts in the $t\bar{t}$ and the $Z+b$ analysis prevent using the yields or cross sections directly. Instead, as the same $t\bar{t}$ MadGraph simulation is used in both analyses, a ratio is calculated which corrects the simulated differential cross section to the measured cross section. This ratio is then used to correct the Z/upper sideband ratio to finally estimate the $t\bar{t}$ content in the Z region.

A complete description of the method can be found in Johannes Hauk's thesis [149]. His results, using the differential cross sections calculated in this work, have also been included into the latest $Z/\gamma^* + b$ jet cross section publication by CMS [148].

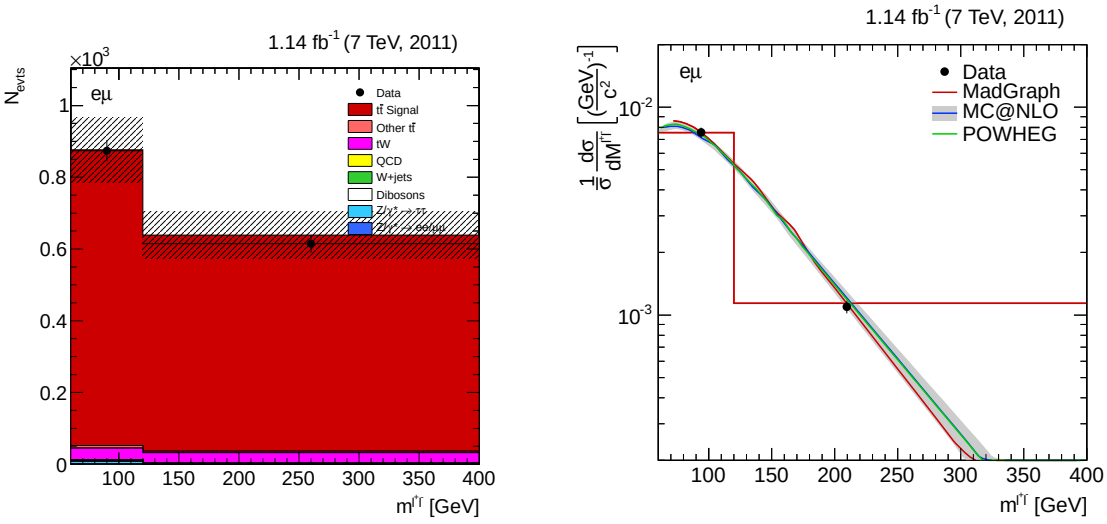


Figure 9.13: Differential cross section in the $e\mu$ channel in a different binning, covering the Z window and the upper sideband. The MC cross sections (in units of $10^3\,\mathrm{GeV}^{-1}$) in the two bins are 7.55 and 1.14, the measured ones are $7.55 \pm 0.27 \pm 0.35$ and $1.09 \pm 0.05 \pm 0.07$.

Chapter 10

Update to 5 fb^{-1}

In the year 2011 a total of 5 fb^{-1} of luminosity have been collected by the CMS detector. In chapter 9 only the first 1.14 fb^{-1} were analysed, the dataset used for the preliminary CMS result [141]. Of course the analysis has not stopped at that point: this work was taken as basis for the analysis of the complete 2011 dataset. The main contributors for the improved analysis were other members of the DESY CMS group: Tyler Dorland (general analysis), David J. Fischer (unfolding), Ivan Asin and Benjamin Lutz (b jets on hadron level), Maria Aldaya (coordination and paper writing), and Andreas Meyer (general supervision) and the results can be found in [150]. My own contribution to the update started with many significant code commits for the creation of a new NTuple and also includes a very first hadron level b jet implementation, many discussions about how things are implemented and how to improve the analysis, help with computing problems, tracking down a strange behaviour in the kinematic event reconstruction, and running of jobs for the energy scale systematics.

10.1 Changes in the update

The update of the analysis is not only an update in the luminosity, but there were also other changes and improvements. The first change is the treatment of τ leptons. While the decay of W bosons to e/μ leptons via intermediate τ leptons is treated as signal in chapter 9, no intermediate τ leptons are allowed in the updated analysis. The main reason for this is that leptons from intermediate τ leptons are softer due to the two additional neutrinos and thus produce a different lepton spectrum. In theory calculations, the inclusion of the extra decay is additional work that is independent of and thus not interesting for the top quark calculations. Therefore comparisons with theories are made easier. On the other hand, the dilepton events via intermediate τ lepton(s), when considered background, need to be subtracted from the number of measured dilepton events, introducing a slightly larger dependence on the simulation in this step.

The second big change was the application of a regularised unfolding procedure [151, 152] which takes migrations into other bins into account. It has been proven that repeated binwise calculation and reweighting of the prediction leads to the same central

values as the full unfolding [153]. The improvements of the full unfolding are that the bias introduced by the MadGraph prediction is small even for larger deviations of data and prediction (see also appendix D) and that the statistical uncertainties are described more precisely, i.e. the bin-by-bin calculation can lead to too small statistical errors.

The third main change is that the definition of the visible phase space has been changed from parton level to hadron level. How this is done and the implications have already been discussed in section 9.3. With the new b jet definition, two additional distributions have been added: the b jet transverse momentum and rapidity.

Other improvements include the addition of latest approximate NNLO calculations by Kidonakis [22, 154] into the top quark transverse momentum $p_T^{t\,\mathrm{and}\,\bar{t}}$ and rapidity $y^{t\,\mathrm{and}\,\bar{t}}$ distributions, a revised binning optimised for the new unfolding procedure, and the use of the Combined Secondary Vertex b tagging algorithm (see section 6.8).

Finally the Monte Carlo simulations were replaced with newer ones which provide more statistics for higher pile-up scenarios. Also the Pythia Drell-Yan sample below 50 GeV invariant mass has been substituted with a MadGraph/Pythia simulation.

A paper including these updates is in preparation and currently under CMS collaboration wide review [155].

10.2 Control plots

Figure 10.1 shows the primary vertex distribution for the updated analysis. One apparent change is the primary vertex spectrum which is now shifted to higher multiplicities when comparing to figure 8.2 where the maximum in the N_{vtx} distribution was below 5 vertices.

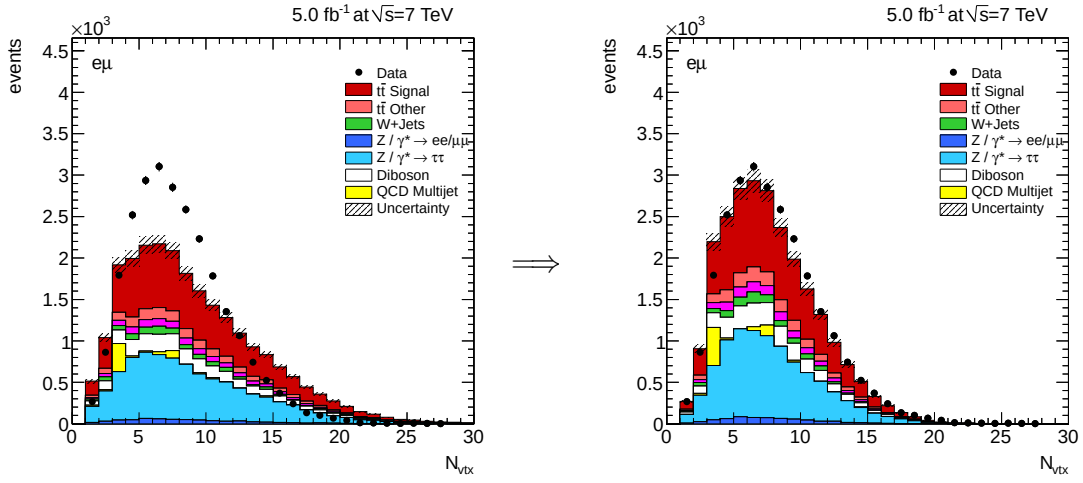


Figure 10.1: Vertex multiplicity in the $e\mu$ channel after dilepton selection before (left) and after (right) reweighting in 5 fb^{-1} [150].

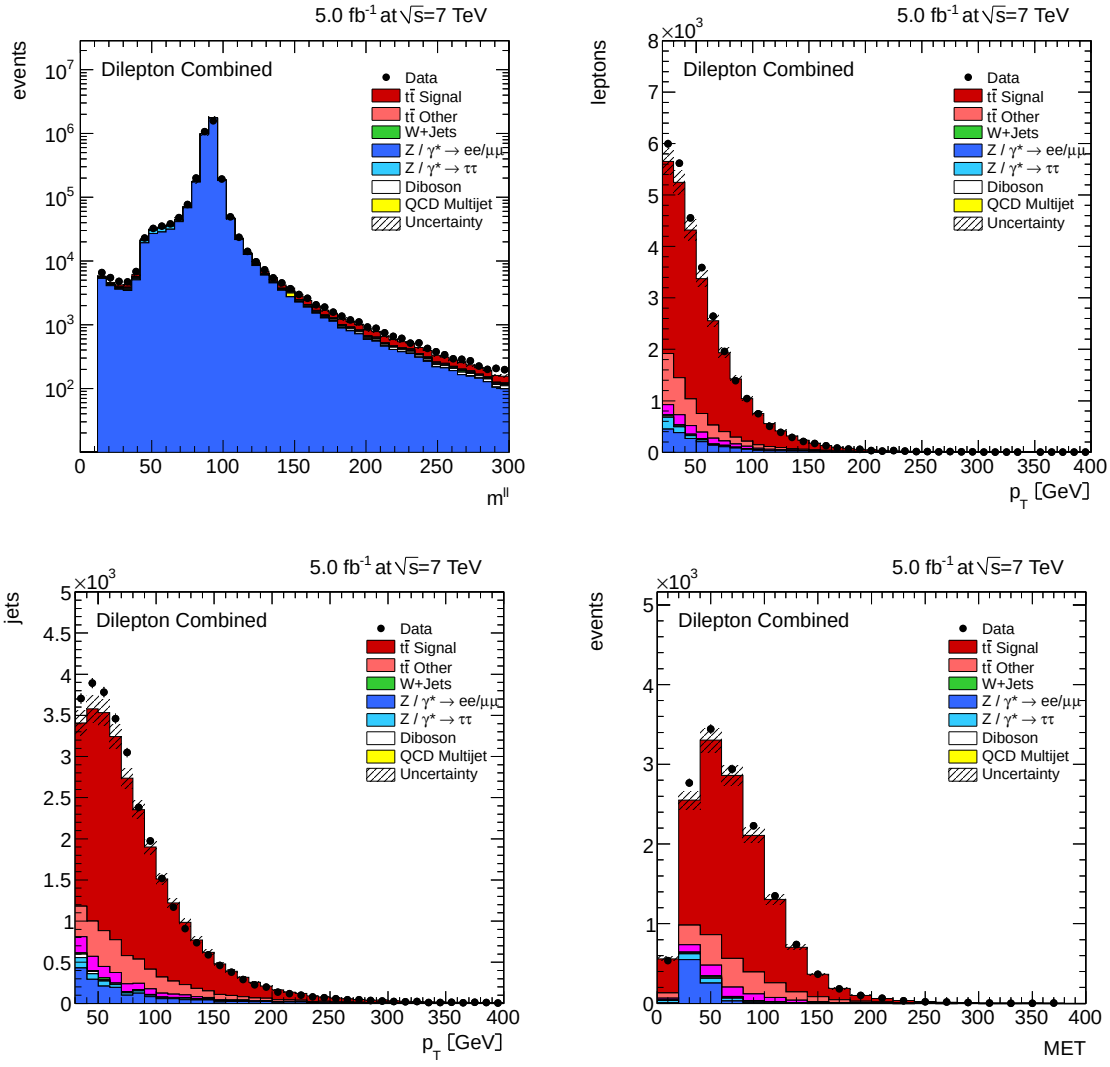


Figure 10.2: Control plots for 5 fb^{-1} [140,150]. The top row shows the invariant dilepton mass on the left (before the Z mass cut). The other plots are made after full selection: the top right plot shows the lepton transverse momentum. The bottom row shows the jet p_T and the missing transverse energy.

In figure 10.2, one can see that the invariant dilepton mass spectrum is now in much better agreement with the data than before, especially in the low mass region due to the described change of the event generator.

The remaining control plots after full selection also show an excellent agreement of data and Monte Carlo. The main difference to the control plots shown in chapter 7 is that the fraction of $t\bar{t}$ other events is larger due to the modified treatment of τ leptons.

10.3 Lepton, lepton pair and jet cross sections

In the following the resulting cross sections are shown for all eleven different quantities. The first variable shown is the lepton transverse momentum. In figure 10.3, the top left plot shows the purities, stabilities and efficiencies. In general the plot is similar to what was shown in the previous chapter. Due to the different treatment of tau leptons however the efficiencies are a bit lower in the updated analysis. On the other hand the events with only two neutrinos are better reconstructed than the events with two additional neutrinos from the tau decay. Therefore the purities and stabilities are higher when intermediate taus are not allowed. This can be seen when comparing the new purities and stabilities to figure 9.1 on page 111.

The plots on the bottom show the event yields and the resulting normalised differential cross section, similar to figure 9.2. The visible difference in the event yield plot is again due to the tau treatment, i.e. the $t\bar{t}$ other contribution is now larger. Also the resulting differential cross section plot is different w.r.t. to chapter 9 in that the cross section was calculated using a full unfolding procedure (see [150] for details) and that the predictions from MC@NLO and POWHEG are shown as bin-average cross sections only. This is useful because otherwise it is often hard to distinguish between the different predictions.

The plot on the top right of figure 10.3 compares the binwise cross section calculation (BBB) used in this thesis with the regularised unfolding (SVD) used in the extended analysis. The τ value given in the plot is a measure for the regularisation, details are given in [150]. The fact that the ratio is close to 1 in all bins is another proof that the binwise calculation done in this thesis works well for this quantity.

The conclusions made from the $p_T^{\ell^+ \text{ and } \ell^-}$ cross section measurement do not change: all three predictions are similar and the data agrees with all of them.

For the other distributions, mostly only the event yields and the resulting normalised differential cross sections are shown. All plots comparing SVD and binwise unfolding can be found in appendix D.2. The purities and stabilities are all similar to the ones shown in the previous chapter, with the difference that the efficiencies are a bit lower and the purities and stabilities a bit higher. The corresponding plots can be found in [150].

In the $\eta^{\ell^+ \text{ and } \ell^-}$ cross section (figure 10.4), the asymmetry seen in the fifth bin in figure 9.3 is (almost) gone. With the larger statistics, one can also see that the data prefers the NLO predictions over MADGRAPH.

The lepton pair $p_T^{\ell\ell}$ cross section with 5 fb^{-1}, shown in figure 10.5, now also allows

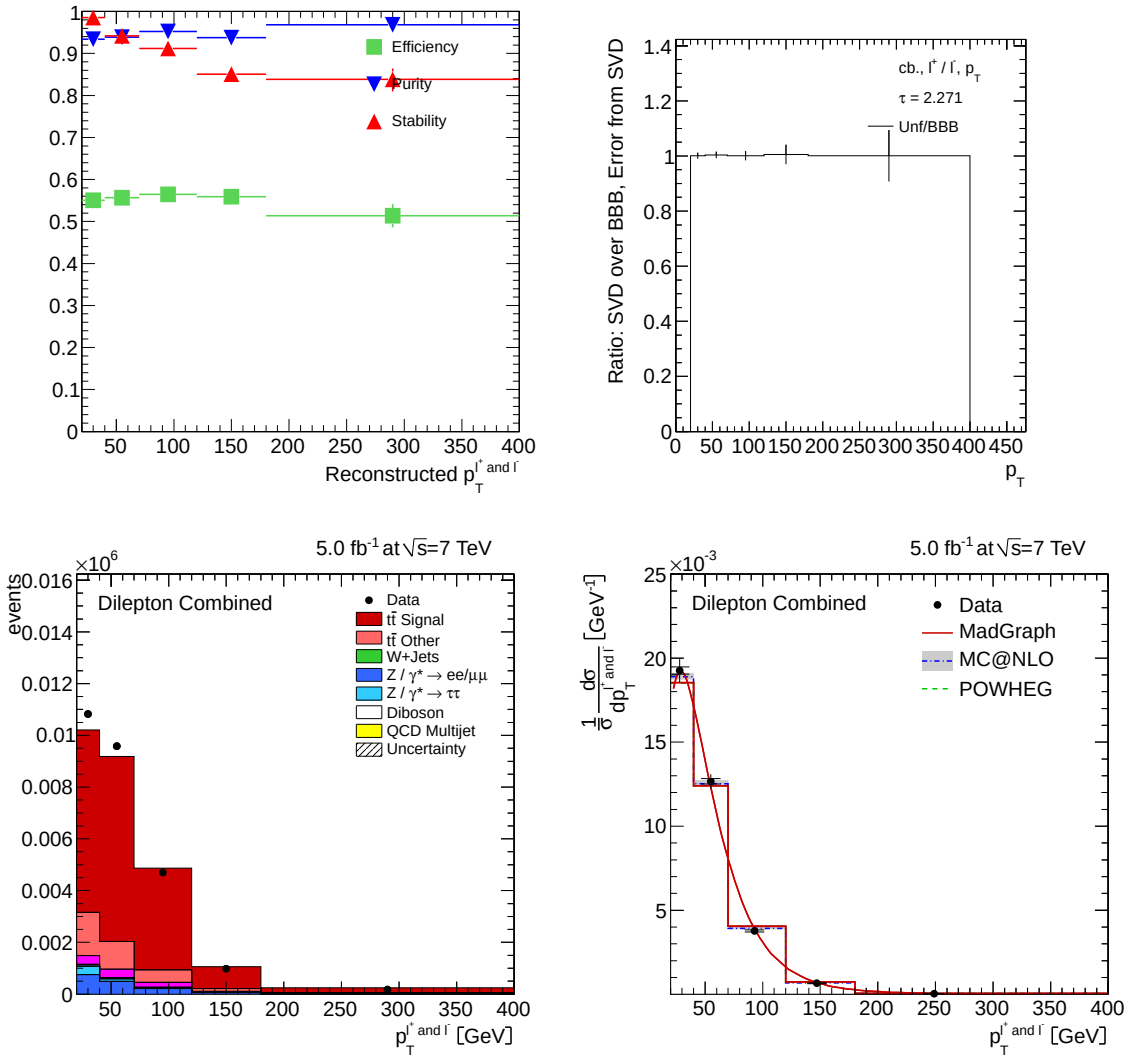


Figure 10.3: Normalised lepton $p_T^{\ell^+ \text{ and } \ell^-}$ cross section for 5 fb^{-1} [140, 150]. The first row shows the purities, stabilities and efficiencies and the ratio between the results from SVD and binwise unfolding [153]. The second row shows the event yield and the normalised differential cross section.

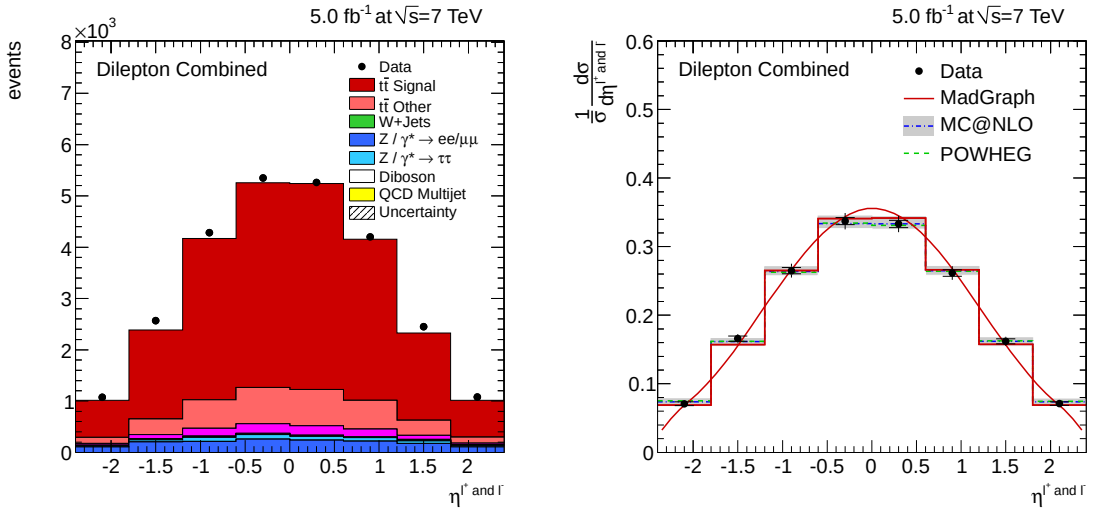


Figure 10.4: Event yield and normalised lepton $\eta^{\ell^+ \text{and} \ell^-}$ cross section for 5 fb^{-1} [140, 150].

conclusions. While it looked like the data was somewhere in between MadGraph and the NLO predictions in 1 fb^{-1} (figure 9.4), with a slight preference to the NLO predictions, the updated analysis clearly shows that NLO fits better in all bins but the fourth.

Figure 10.6 shows the mass spectrum of the lepton pair $m^{\ell\ell}$. While no trend was visible in figure 9.5, now a preference for the NLO calculations is visible especially in the first bin.

Figures 10.7 and 10.8 show the b jet differential cross sections which have not been measured in chapter 9. The $p_T^{b \text{and} \bar{b}}$ distribution (figure 10.7) shows agreement among the predictions but a softer spectrum in data, however with large statistical uncertainties.

The $\eta^{b \text{and} \bar{b}}$ plot (figure 10.8) shows what was also seen in the lepton η distribution: the data is less central than the predictions, corresponding to a preference of the NLO predictions. On the other hand MC@NLO and Powheg are not perfect, both are still too central when compared to the data.

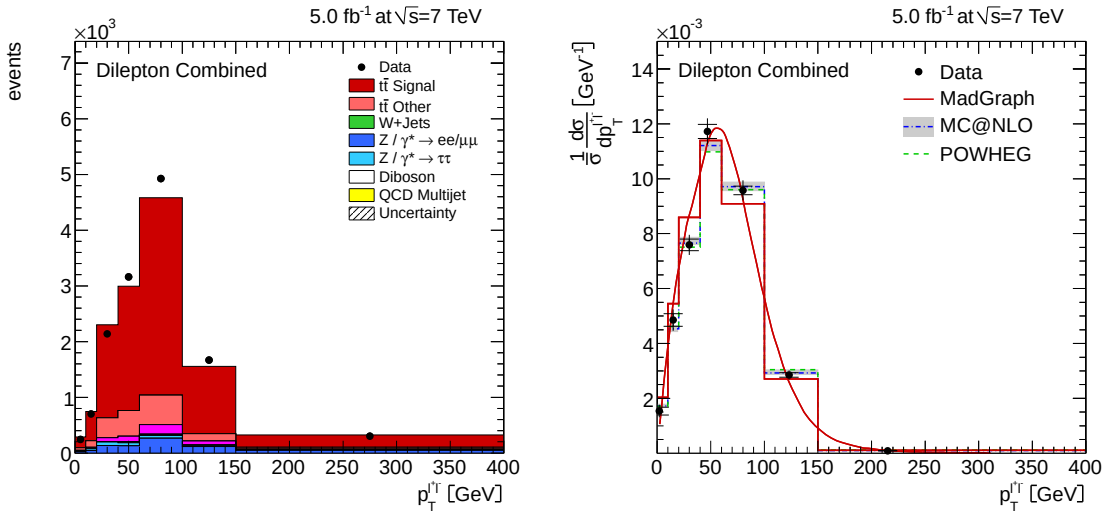


Figure 10.5: Event yield and normalised lepton pair $p_T^{\ell\ell}$ cross section for 5 fb^{-1} [140,150].

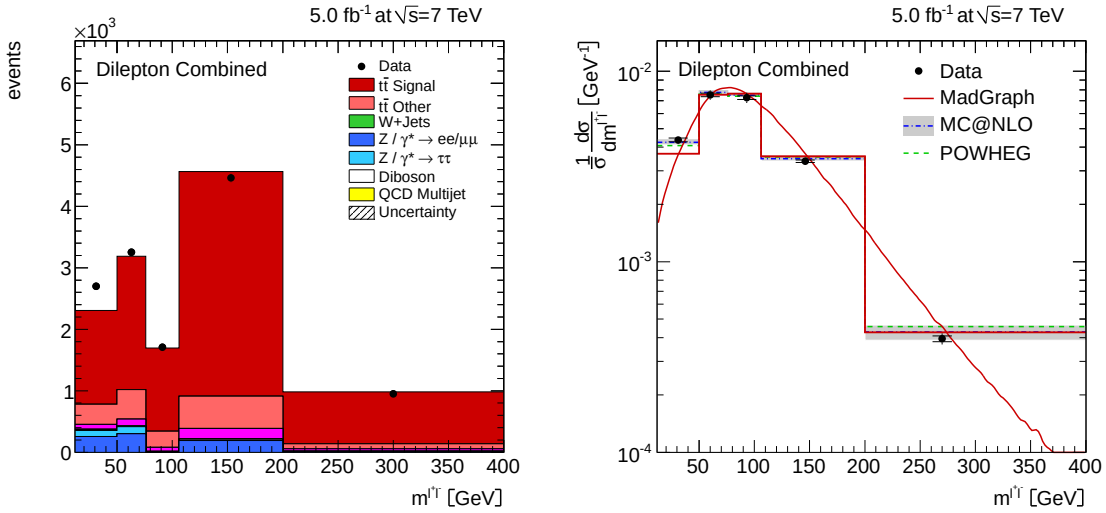


Figure 10.6: Event yield and normalised lepton pair $m^{\ell\ell}$ cross section for 5 fb^{-1} [140,150].

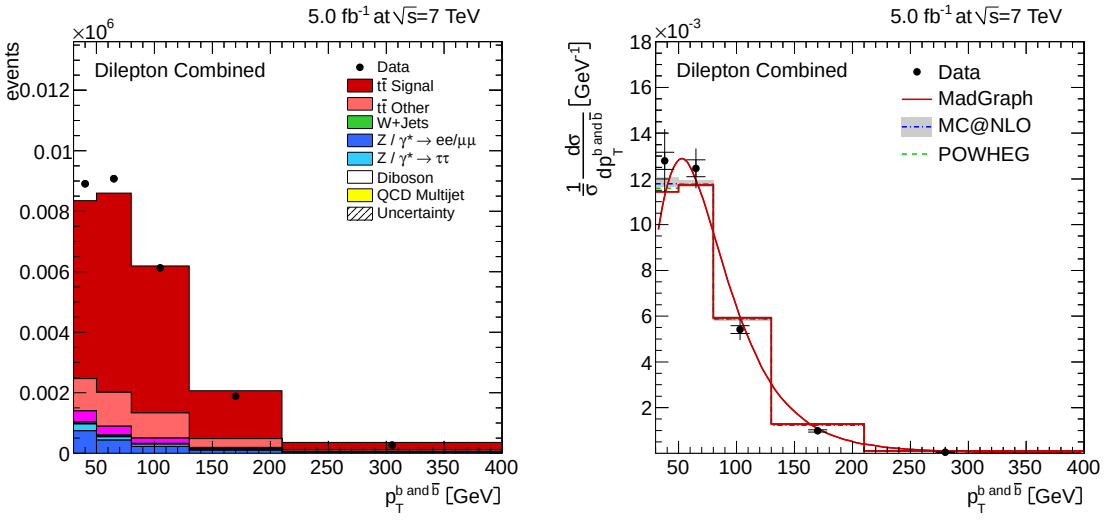


Figure 10.7: Event yield and normalised b jet $p_T^{b\,\mathrm{and}\,\bar{b}}$ cross section for 5 fb^{-1} [140, 150].

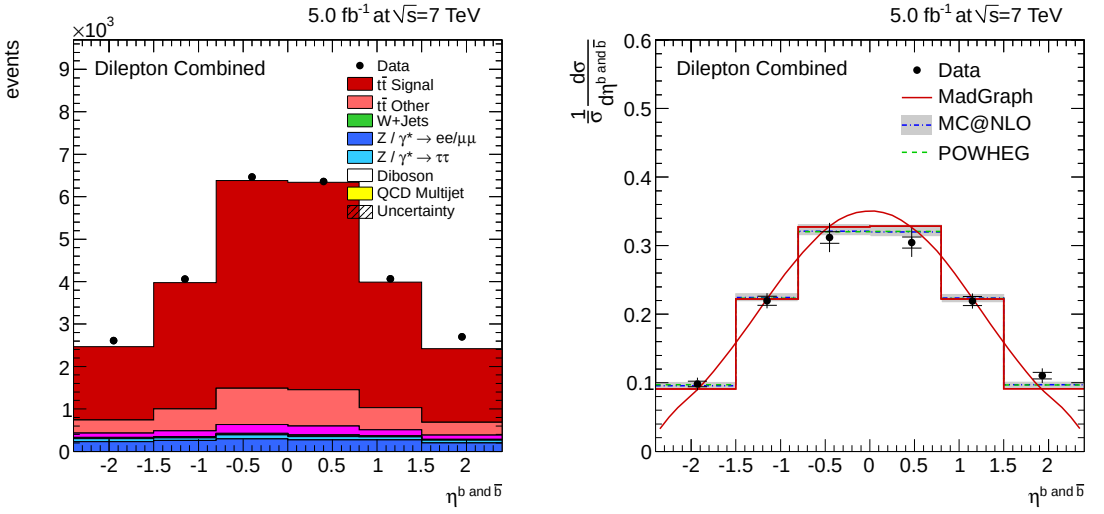


Figure 10.8: Event yield and normalised b jet $\eta^{b\,\mathrm{and}\,\bar{b}}$ cross section for 5 fb^{-1} [140, 150].

10.4 Top quark and top quark pair cross sections

When the first top quark plot was introduced in section 9.6.4, it was noted that the purities and stabilities were lower than in the lepton quantities. In the update of the top quark transverse momentum (figure 10.9), one bin was added w.r.t. to the result presented in figure 9.7. Even though the binning is finer, SVD and BBB unfolding differ only by 5 % or less (shown in figure D.6(e) on page 170). This reconfirms that the bias introduced in chapter 9 is small also for top quark cross sections. In this section, all top and top quark pair quantities are extrapolated to the full phase space to make comparisons to other experiments and calculations easier. The top transverse momentum is also one of the distributions where approximate NNLO calculations [22] are available. While MadGraph and the two NLO calculations produce a spectrum which is harder than the data, the approximate NNLO result is in excellent agreement with data! Obviously it is not sufficient to use NLO calculations to predict the top p_T distribution correctly.

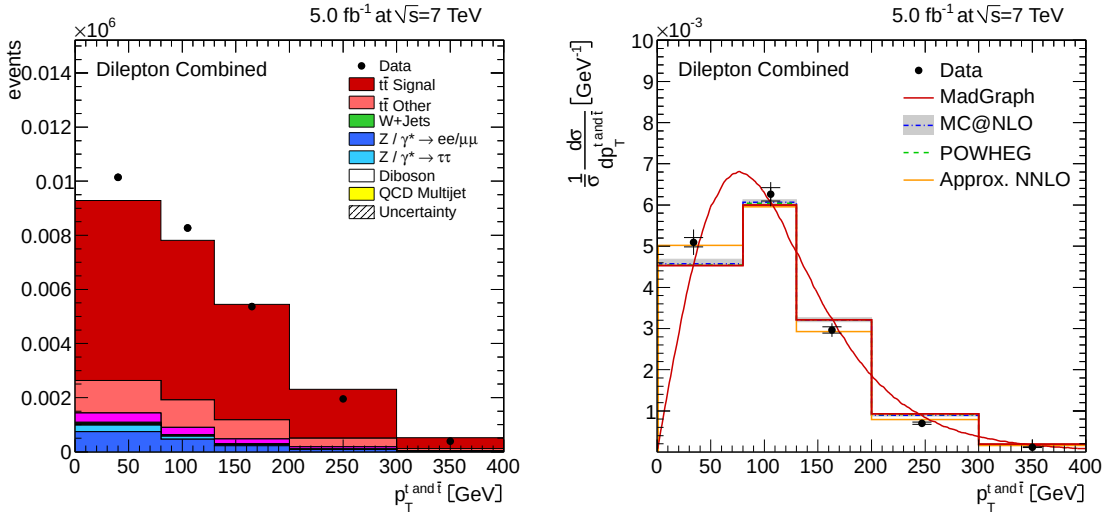


Figure 10.9: Event yield and normalised top quark $p_T^{t\,\mathrm{and}\,\bar{t}}$ cross section for 5 fb^{-1} [140, 150].

An approximate NNLO calculation is also available for the top quark rapidity [154], the distribution is shown in figure 10.10. The NNLO calculation agrees almost completely with MadGraph, while Powheg and MC@NLO predict a very slightly wider spectrum. The differences are only very small and the data is compatible with all distributions because the uncertainties are larger than the differences in the predictions.

Also note that in this distribution even two bins have been added compared to figure 9.8, yet the SVD/BBB ratio is still compatible with 1 in all bins (see figure D.6(f)).

The update of the transverse momentum distribution of the $t\bar{t}$ system, shown in figure

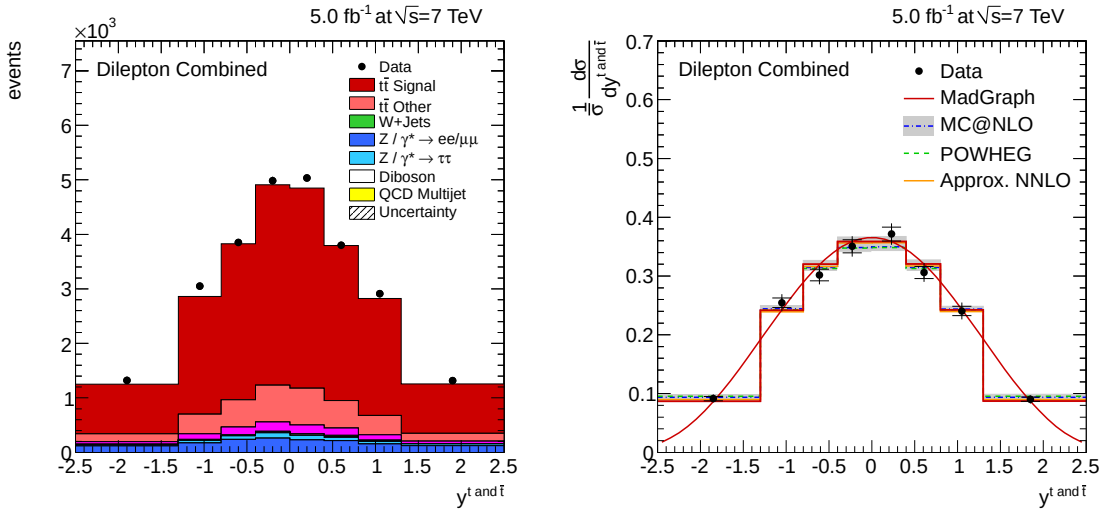


Figure 10.10: Event yield and normalised top quark rapidity $y^{t\,\mathrm{and}\,\bar{t}}$ cross section for 5 fb^{-1} [140, 150].

10.11, does not give any additional insights which have not already been given in section 9.6.6: MadGraph and both NLO predictions show the same behaviour and the data agrees well.

Figure 10.12 shows the $t\bar{t}$ rapidity distribution. The differences between the predictions and also the data were barely visible in figure 9.10—due to the new plotting style in the update one can see that MadGraph is a tiny bit more central than the NLO distributions. The data is even more central and thus closer to the MadGraph bin-average cross sections in all bins. On can thus conclude that the data points agree better with MadGraph than with the NLO calculations, even though it has to be reminded that the measurement is also compatible with MC@NLO and Powheg.

The same tiny preference for MadGraph is also visible in the $m^{t\bar{t}}$ cross section in figure 10.13, otherwise there are no new insights.

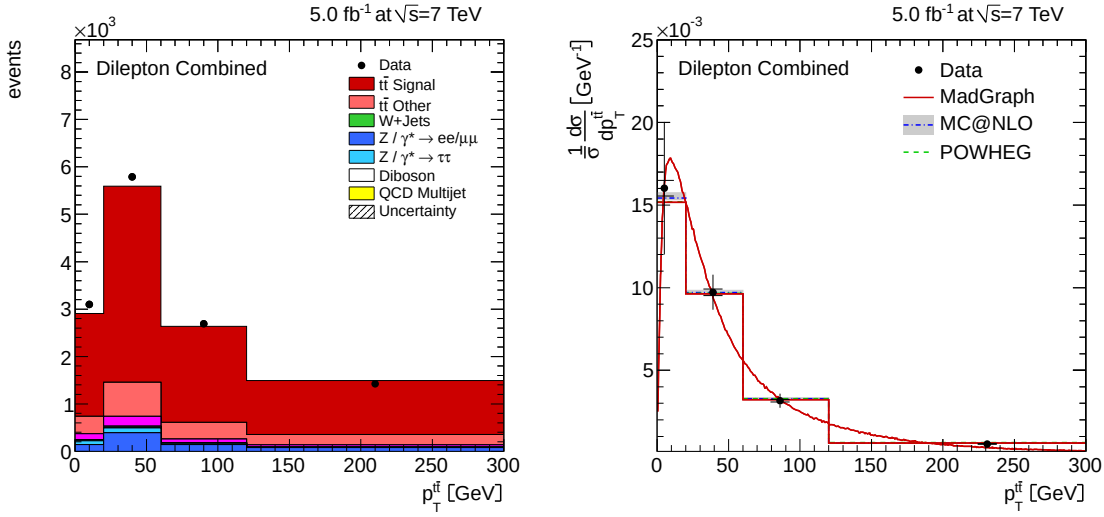


Figure 10.11: Event yield and normalised top pair $p_T^{t\bar{t}}$ cross section for 5 fb^{-1} [140, 150].

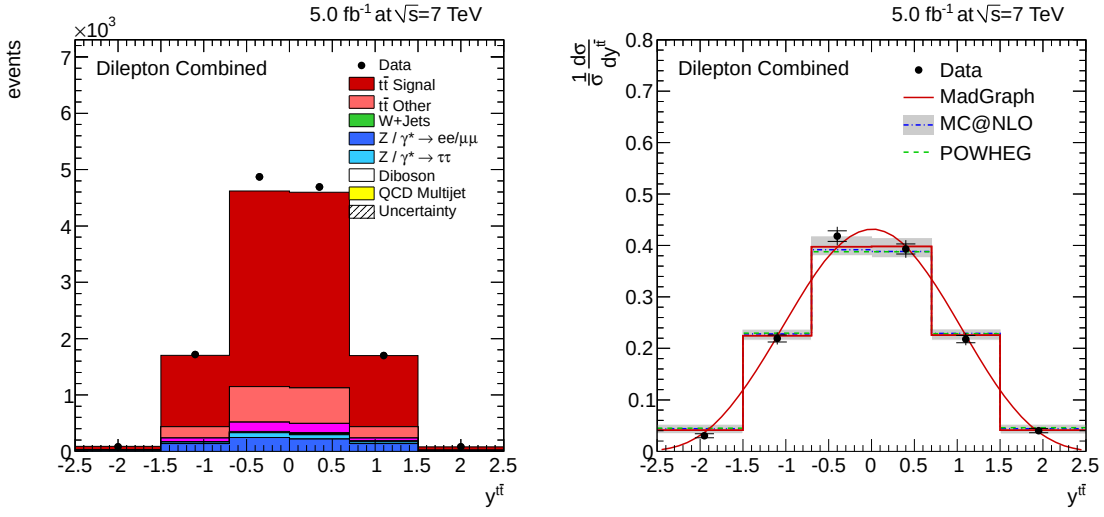


Figure 10.12: Event yield and normalised top pair rapidity $y^{t\bar{t}}$ cross section for 5 fb^{-1} [140, 150].

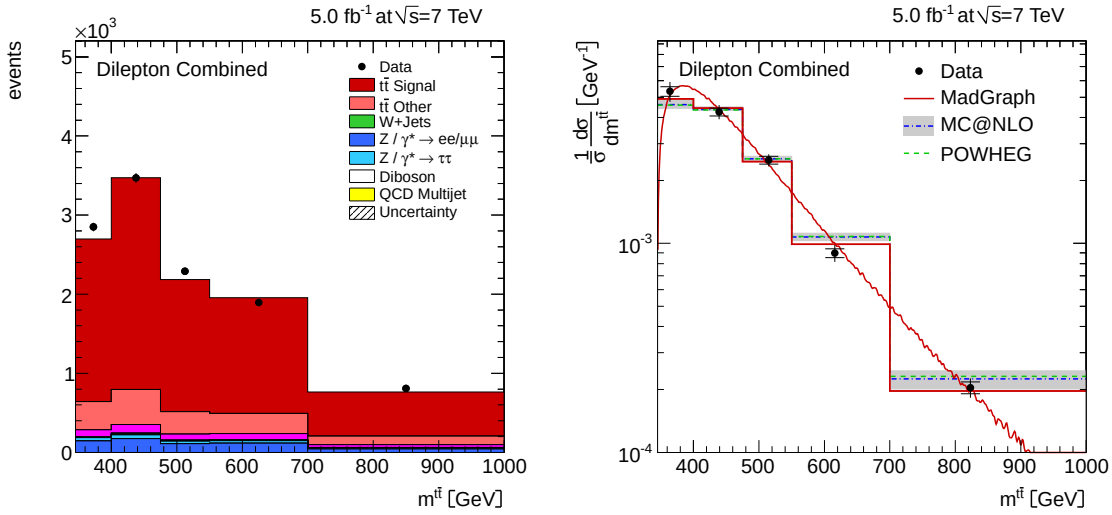


Figure 10.13: Event yield and normalised top pair $m^{t\bar{t}}$ cross section for 5 fb^{-1} [140,150].

10.5 Summary

The update to 5 fb^{-1} has reduced fluctuations such as the asymmetry in the lepton η distribution and made it possible to see that the MC@NLO and POWHEG generators agree better with data than MADGRAPH in the lepton pair quantities and the b jet rapidity, while MADGRAPH gives slightly better results for the $t\bar{t}$ rapidity and mass. Finally approximate NNLO calculations are even better as seen in the top quark transverse momentum. However, NNLO does not improve all predictions significantly: in the top quark rapidity the NNLO result is not much different from MADGRAPH, MC@NLO and POWHEG predictions.

Chapter 11

Summary and Outlook

11.1 Summary

Two different topics have been presented in this work, the results of the technical contribution to the CMS collaboration and the physics analysis.

In the technical part the deployment process of the CMS software on the Worldwide LHC Computing Grid has been presented. Software to automate the installation almost completely was developed and is in use since the beginning of 2011. It circumvents several possible problems which can occur during the installation and has lead not only to a faster but also to a more reliable deployment process. This work has also been presented at the CHEP 2010 conference and published in [76].

In the physics part of this work, normalised differential top quark pair production cross sections have been measured in all the three dilepton decay channels for the first time in proton-proton collisions. The dataset used has an integrated luminosity of 1.14 fb^{-1} and was recorded using the CMS detector in the first half of 2011 at a centre-of-mass energy of 7 TeV.

The measurement was performed differentially in the lepton p_T and η, the lepton pair p_T and mass, the top quark p_T and rapidity, and the top quark pair p_T, rapidity and mass. The results are compared to three QCD predictions by MadGraph, MC@NLO and Powheg. In general, a good agreement of the data with all three predictions is observed. All results have been included into the first preliminary result of the CMS collaboration [141] and made public in February 2012 (except for the invariant $\text{t}\bar{\text{t}}$ mass cross section, for which a dedicated analysis was available [129]).

A first application of this work is the $\text{t}\bar{\text{t}}$ background determination using the normalised differential invariant dilepton mass cross section in the $e\mu$ channel [149]. The result of this work was used in the publication of the Z/γ^*+b-jet cross section measurement of the CMS collaboration [148].

Finally the extension of this work to the full 2011 dataset has been presented, in which slightly different cross sections are measured due to the exclusion of intermediate taus and a different b jet definition. The results are similar and also show generally a good agreement of the predictions with the data. In the normalised top quark transverse momentum differential cross section the data is softer than the three predictions. Here,

also a NNLO prediction is compared to the data and found to be in excellent agreement with it. The extended analysis is currently in CMS collaboration wide review and will hopefully be submitted to EPJC soon [155].

The final conclusion of this work is that the shapes of the differential $t\bar{t}$ cross sections are all described amazingly well by the QCD theory predictions. Therefore no hints for new physics are found.

11.2 Outlook

This work covers the first measurement of top quark differential cross sections at the LHC. Several improvements have already been implemented in the update to the full statistics of 2011 (chapter 10). There are several ideas for further studies with the same dataset. Possible room for improvement is in the kinematic event reconstruction which currently allows top quark masses from 100 to 300 GeV to cover mismeasurements. This region could be limited to a smaller window around the top quark mass to have better reconstructed events.

Another field is the treatment of systematic uncertainties. Currently the total systematic uncertainty is calculated from the quadratic sum of the individual uncertainties which are thus treated as uncorrelated. Detailed studies could check the correlations and produce more accurate uncertainties. Before doing this, the statistics of the dedicated model variation samples should be increased.

The current analysis shows the lepton/antilepton and top quark/antitop quark quantities combined, i.e. with one entry for each of the particles in the plots. One could also add distributions for the leading (i.e. highest p_T) and the second leading lepton or top quark separately to check if also that is predicted correctly.

Finally, this analysis can also be applied to proton-proton collisions with higher centre-of-mass energies. As of August 19th, 2012 CMS has already recorded almost 11 fb^{-1} of 8 TeV data [156], twice the amount of the 2011 dataset, allowing for a more precise measurement. With so much data, one can trade statistics for the reduction of systematic uncertainties, for example by requiring two b tagged jets to select an even purer $t\bar{t}$ sample. This would also have an advantage for the kinematic fit which would only take the two b jets into account. Before work is invested in the precision measurement, one could repeat the current analysis and compare the results at different centre-of-mass energies.

To close the outlook, I hope that the presented results will be used many in other analyses, for example for the $t\bar{t}$ background determination in events with other final states.

Appendix A

Cross section plots for the individual channels

On the following pages figures showing the cross sections in the individual channels can be found. Each figure consists of 12 plots arranges in 4 rows and 3 columns. On the left side the efficiency, purity and stability plots are shown, in the middle the yields and on the right side the cross sections. From top to bottom, the ee, $e\mu$ and $\mu\mu$ channel s are shown. In addition, the combination plots are shown on the bottom, they are a repetition of what has already been shown in chapter 9.

In all cross section plots, the inner error bar shows the statistical error, the outer error bars statistical and systematic errors added in quadrature. The measurements are compared to the predictions from MadGraph, Powheg and MC@NLO generators. For the MadGraph comparison, a smooth curve and a bin-averaged histogram are shown.

In the event yields, the error band corresponds the the uncertainty on the luminosity and the $t\bar{t}$ cross section.

- the lepton transverse momentum is shown in figure A.1 on the following page
- the lepton pseudorapidity is shown in figure A.2 on page 149
- the lepton pair transverse momentum is shown in figure A.3 on page 150
- the lepton pair invariant mass is shown in figure A.4 on page 151
- the top quark transverse momentum is shown in figure A.5 on page 152
- the top quark rapidity is shown in figure A.6 on page 153
- the top quark pair transverse momentum is shown in figure A.7 on page 154
- the top quark pair rapidity is shown in figure A.8 on page 155
- the top quark pair invariant mass is shown in figure A.9 on page 156

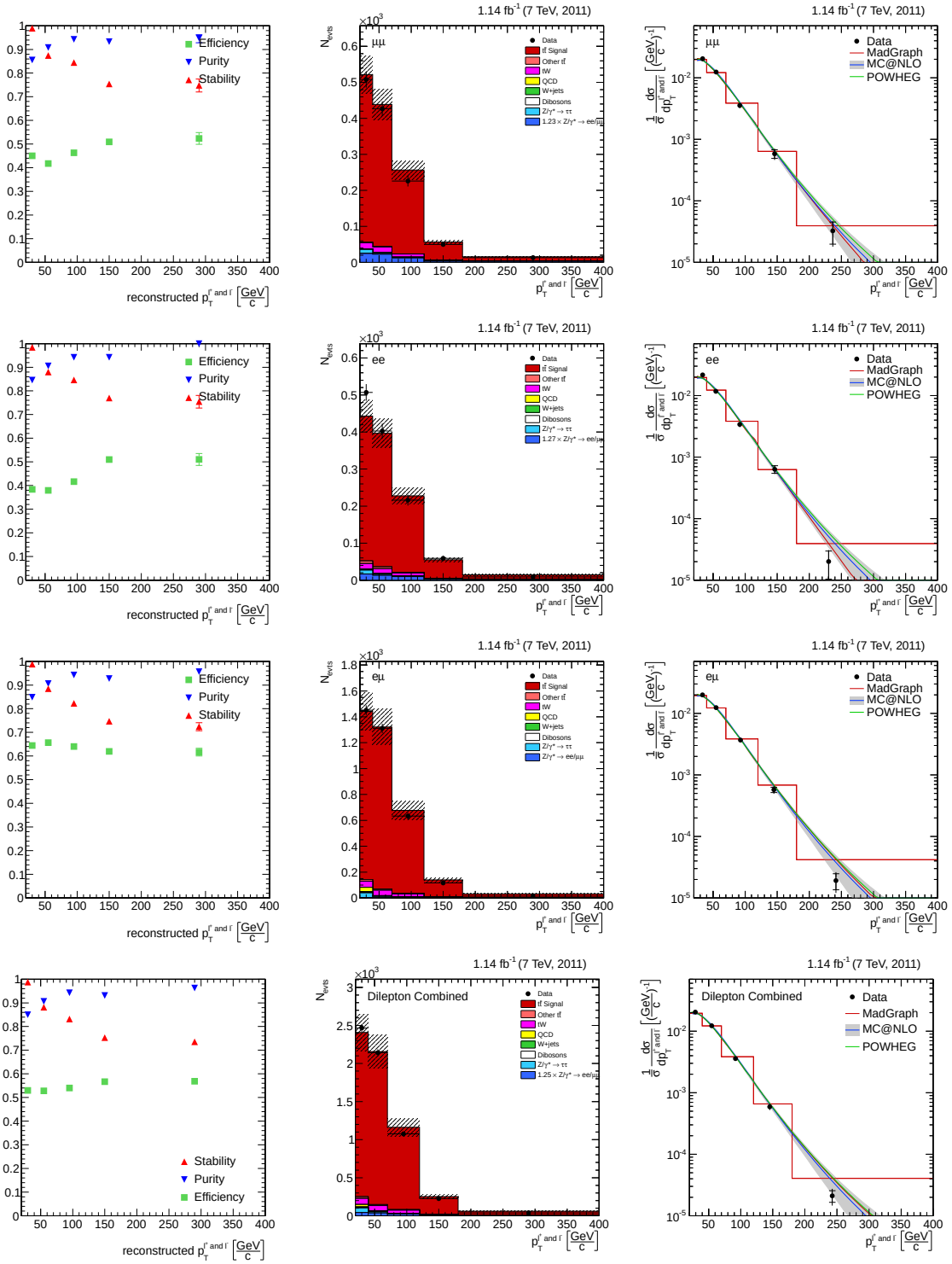


Figure A.1: Differential $t\bar{t}$ production cross section (right column) as a function of the transverse momentum of the leptons for the three decay channels $\mu^+\mu^-$ (top), e^+e^- (2nd row) and $\mu^\pm e^\mp$ (3rd row) and combined (bottom).

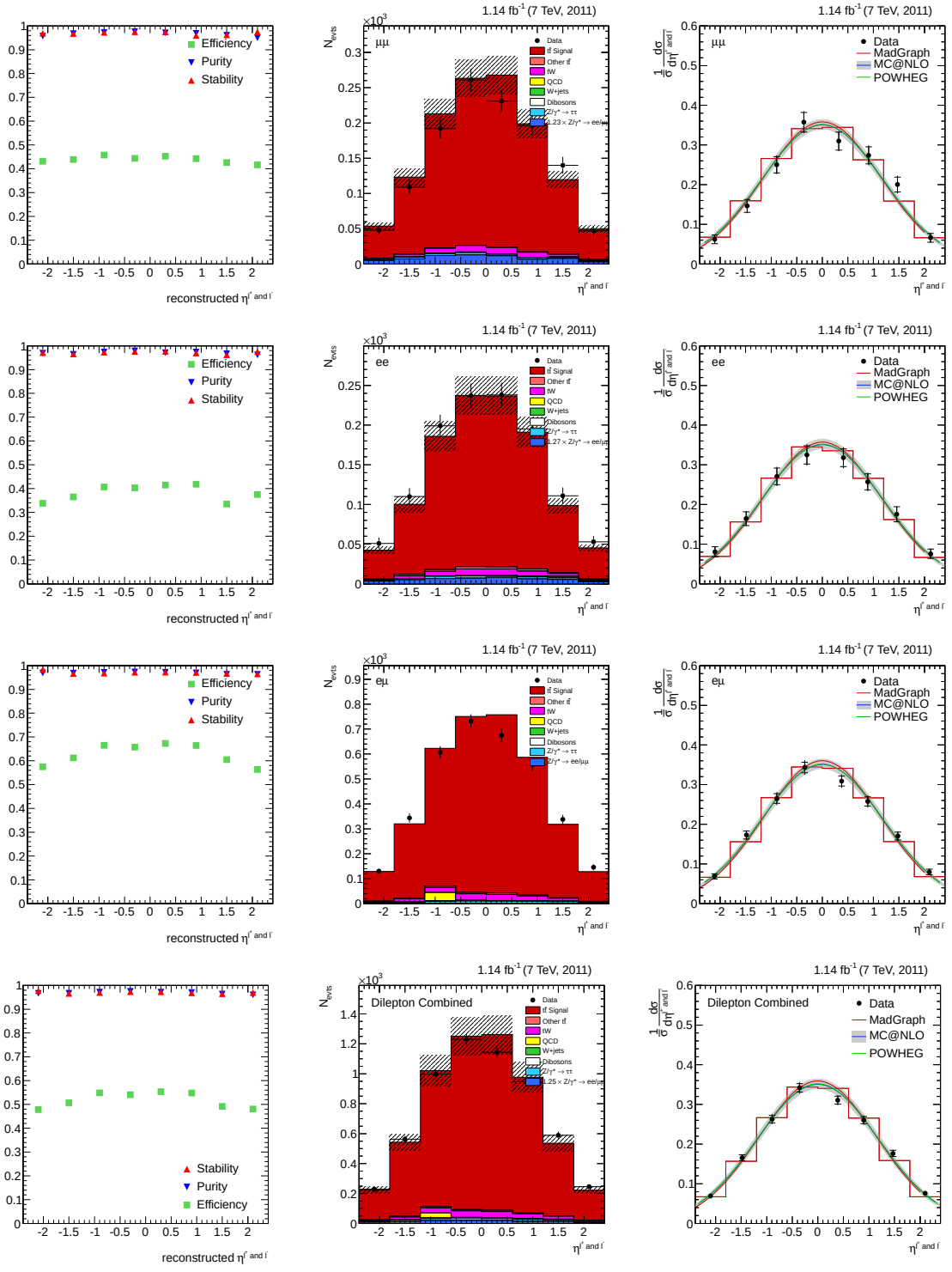


Figure A.2: Differential $t\bar{t}$ production cross section (right column) as a function of the pseudo-rapidity of the leptons for the three decay channels $\mu^+\mu^-$ (top), e^+e^- (2nd row) and $\mu^\pm e^\mp$ (3rd row) and combined (bottom).

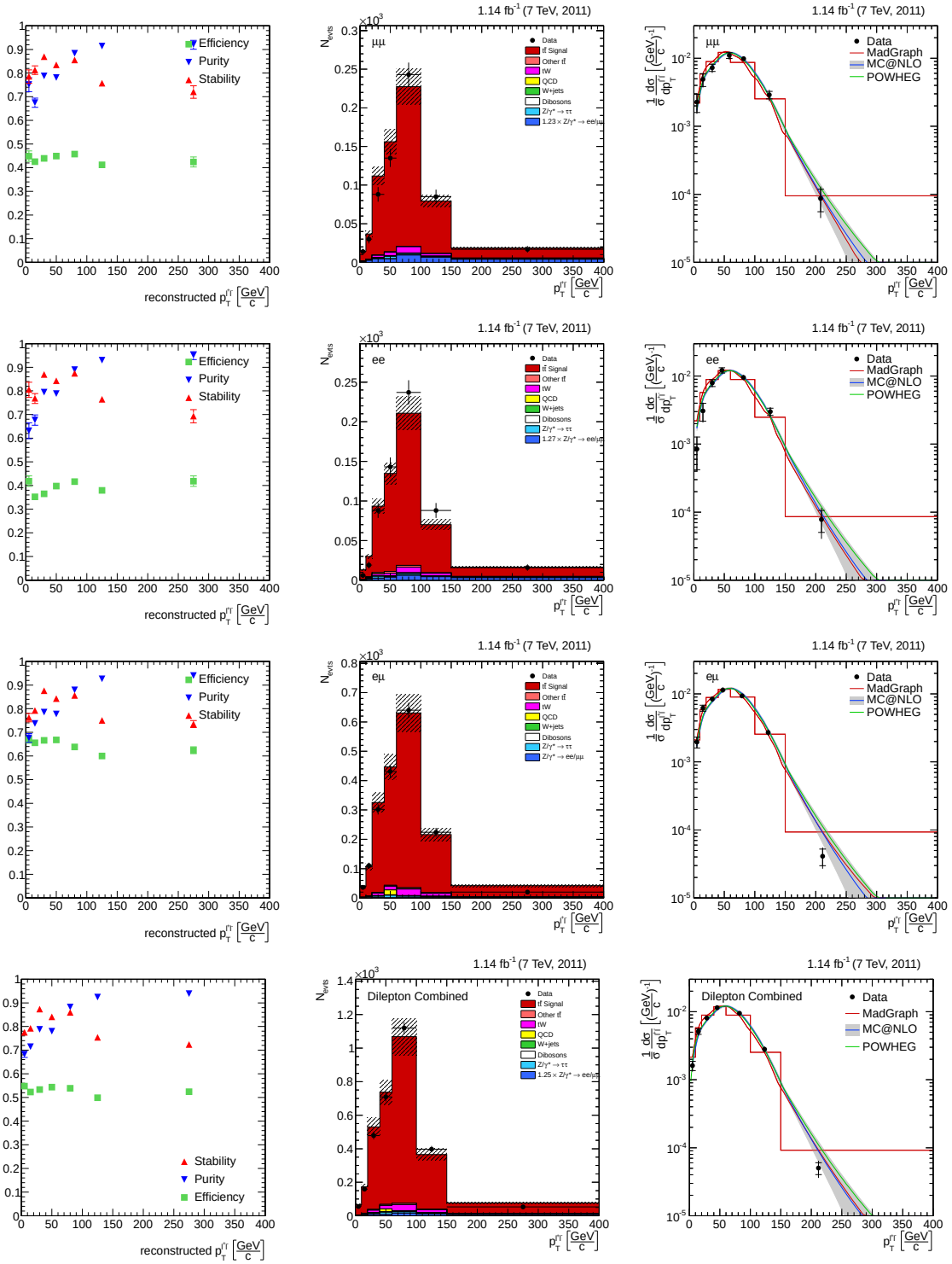


Figure A.3: Differential $t\bar{t}$ production cross section (right column) as a function of the transverse momentum of the lepton pair for the three decay channels $\mu^+\mu^-$ (top), e^+e^- (2nd row) and $\mu^\pm e^\mp$ (3rd row) and combined (bottom).

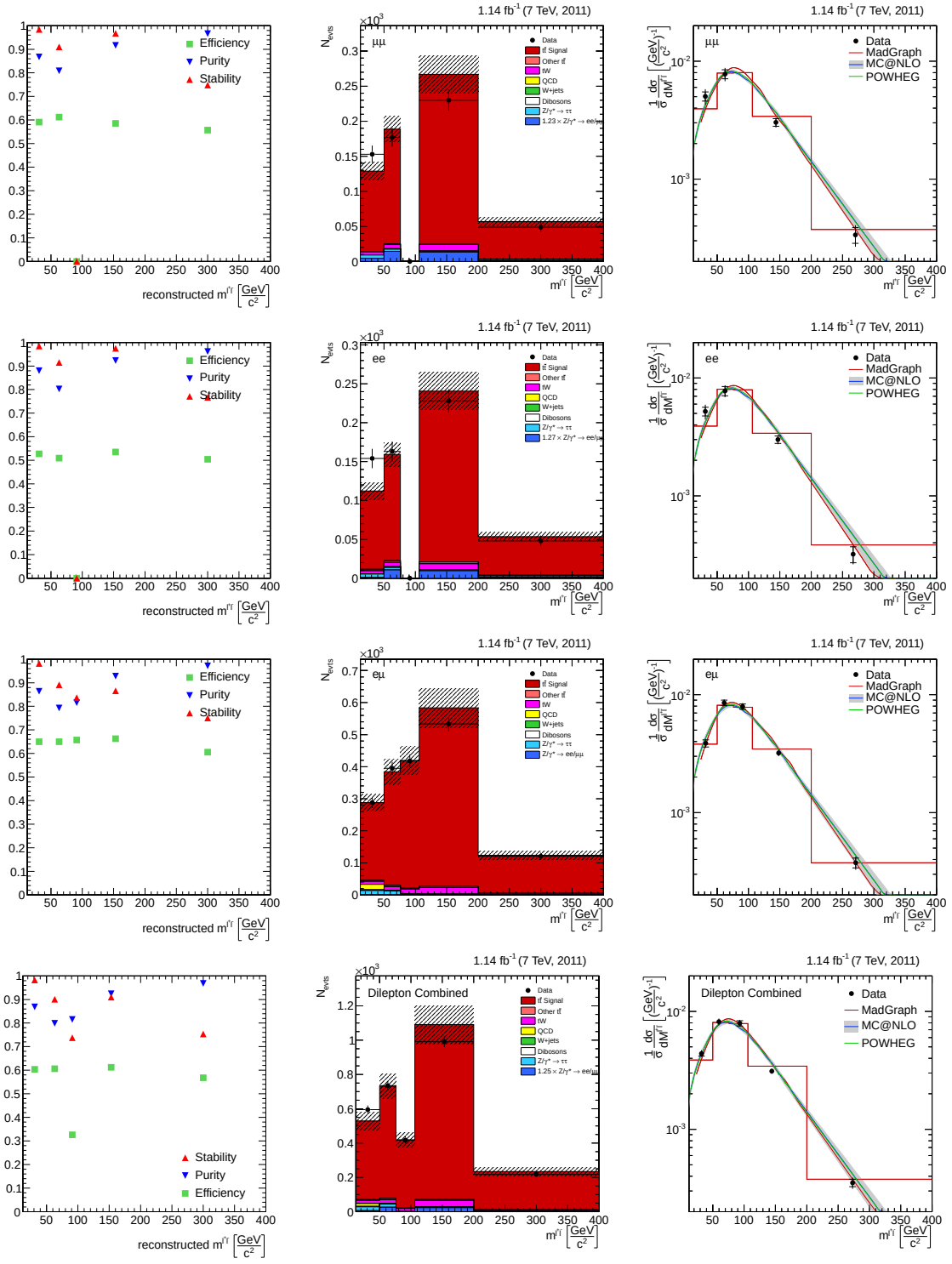


Figure A.4: Differential $t\bar{t}$ production cross section (right column) as a function of the invariant mass of the dilepton system for the three decay channels $\mu^+\mu^-$ (top), e^+e^- (2nd row) and $\mu^\pm e^\mp$ (3rd row) and combined (bottom). For the $\mu^+\mu^-$ and e^+e^- channels no measurement is available in the Z^0 veto bin $76 < m^{\ell\ell} < 106$ GeV.

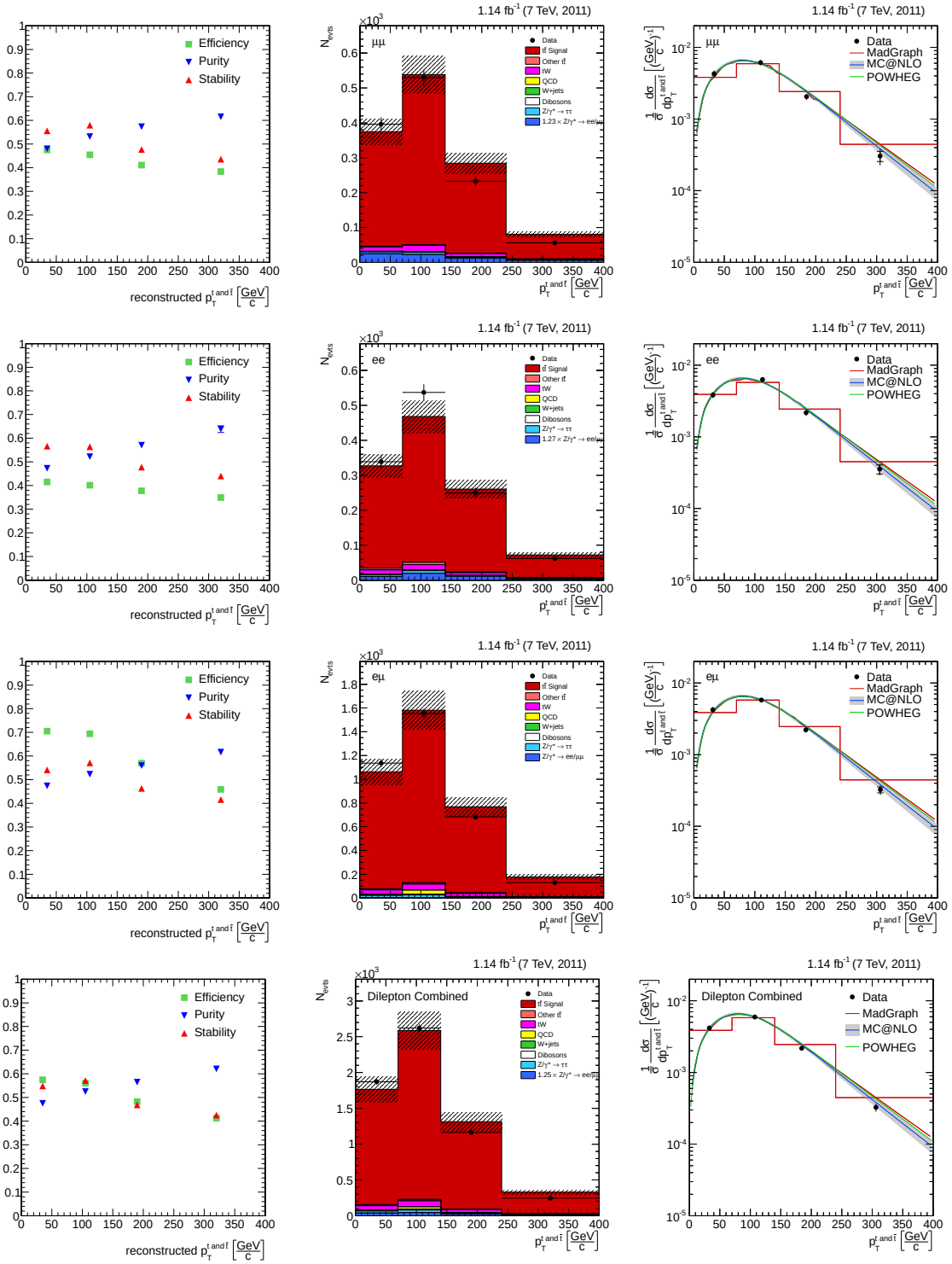


Figure A.5: Differential $t\bar{t}$ production cross section (right column) as a function of the transverse momentum of the top quarks for the three decay channels $\mu^+\mu^-$ (top), e^+e^- (2nd row) and $\mu^\pm e^\mp$ (3rd row) and combined (bottom).

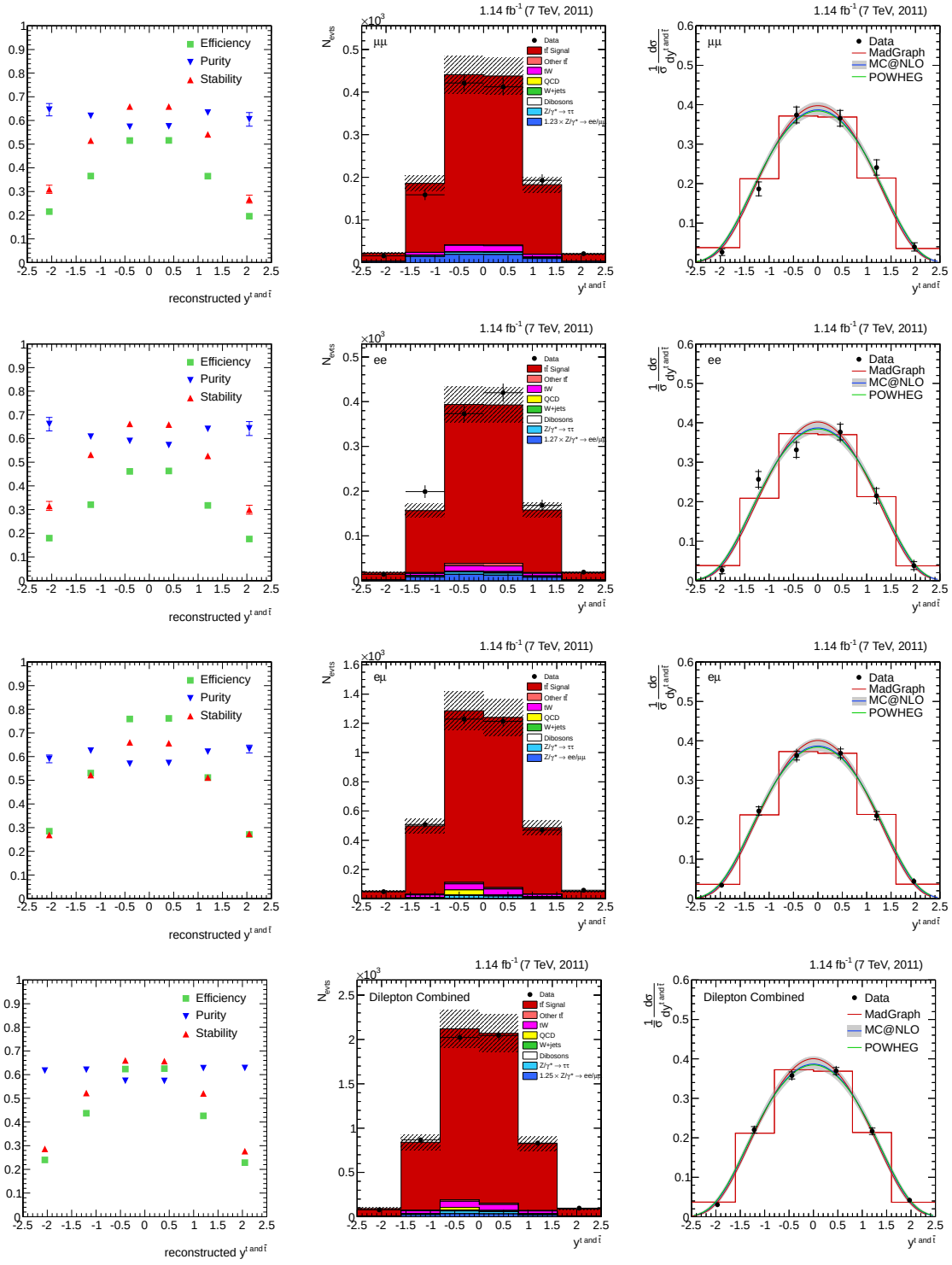


Figure A.6: Differential $t\bar{t}$ production cross section (right column) as a function of the rapidity of the top quarks for the three decay channels $\mu^+\mu^-$ (top), e^+e^- (2nd row) and $\mu^\pm e^\mp$ (3rd row) and combined (bottom).

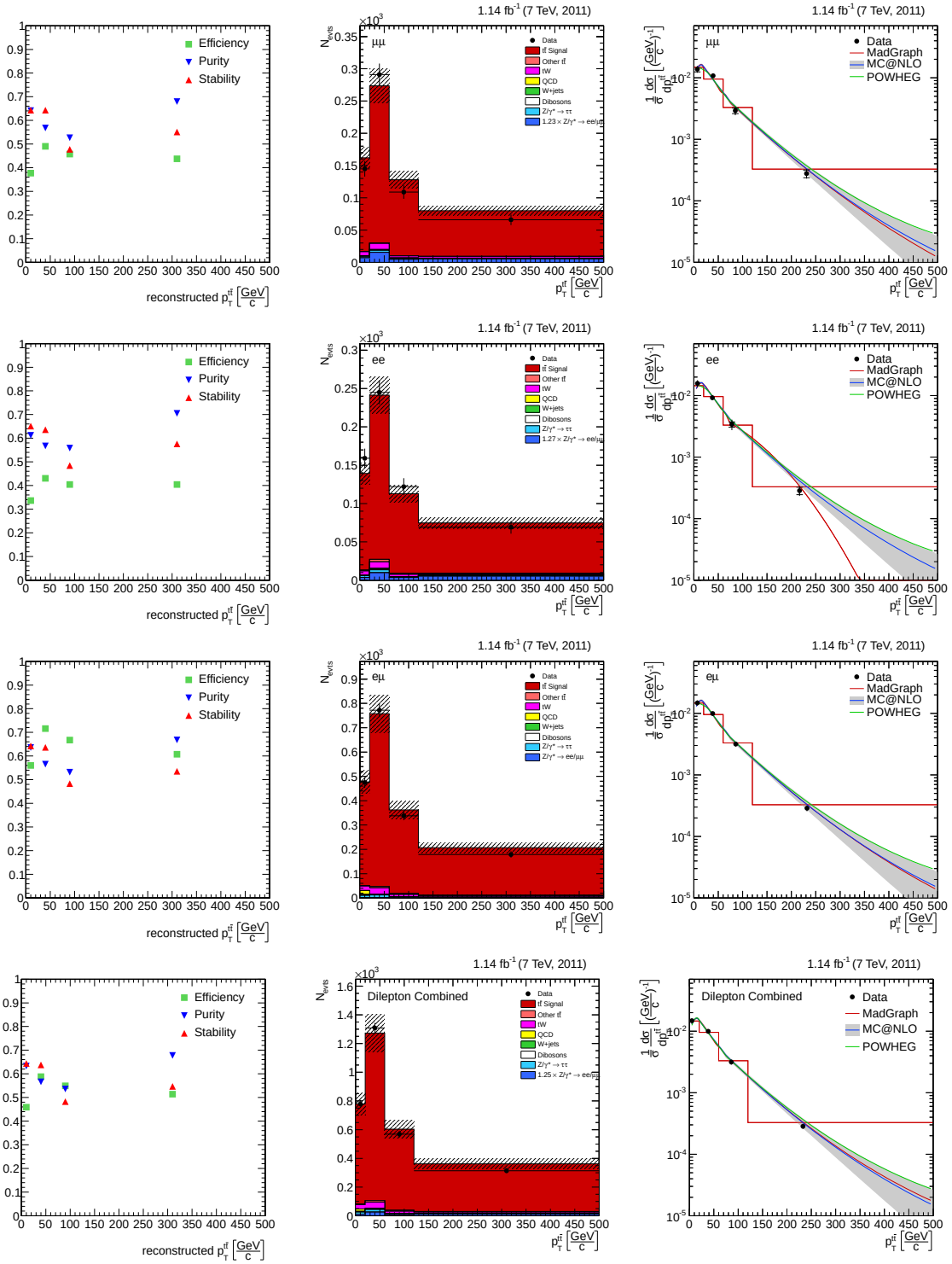


Figure A.7: Differential $t\bar{t}$ production cross section (right column) as a function of the transverse momentum of the $t\bar{t}$ quark pair for the three decay channels $\mu^+\mu^-$ (top), e^+e^- (2nd row) and $\mu^\pm e^\mp$ (3rd row) and combined (bottom).

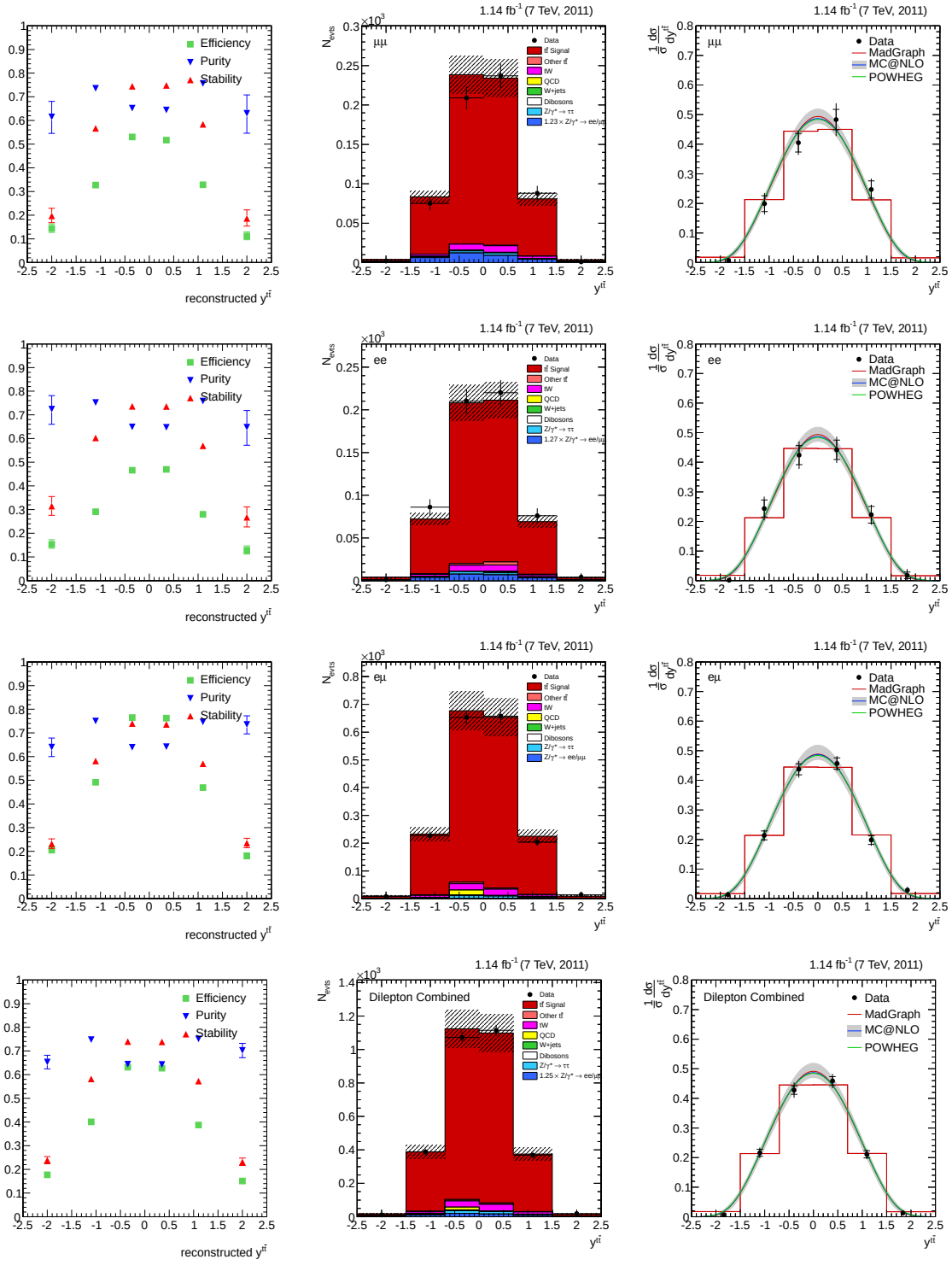


Figure A.8: Differential $t\bar{t}$ production cross section (right column) as a function of the rapidity y of the $t\bar{t}$ quark pair for the three decay channels $\mu^+\mu^-$ (top), e^+e^- (2nd row) and $\mu^\pm e^\mp$ (3rd row) and combined (bottom).

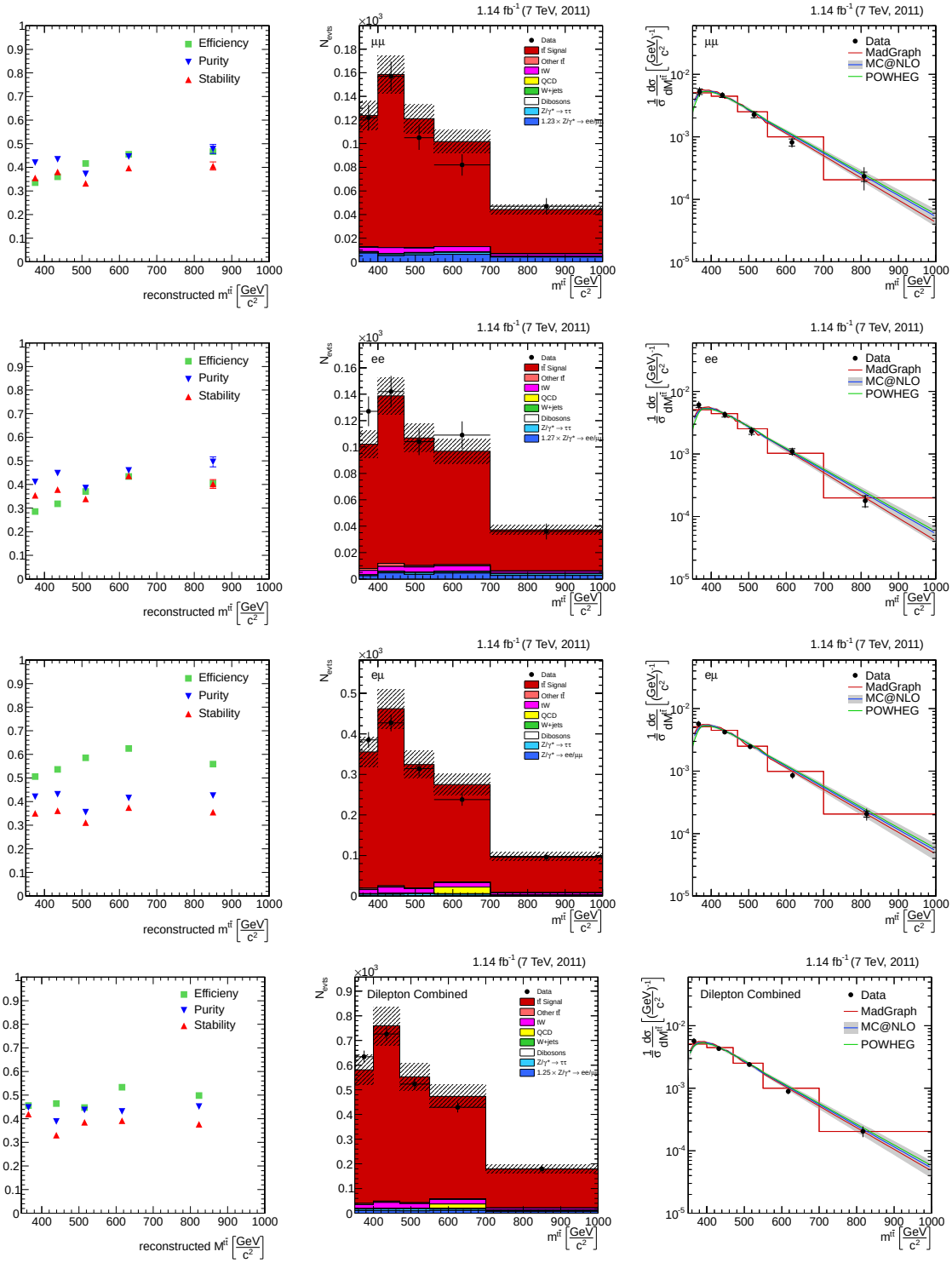


Figure A.9: Differential $t\bar{t}$ production cross section (right column) as a function of the invariant mass of the $t\bar{t}$ quark pair for the three decay channels $\mu^+\mu^-$ (top), e^+e^- (2nd row) and $\mu^\pm e^\mp$ (3rd row) and combined (bottom).

Appendix B

Systematics in individual channels

This section contains the same plots as shown in figure 8.7 on page 102, but for the three dilepton channels separately.

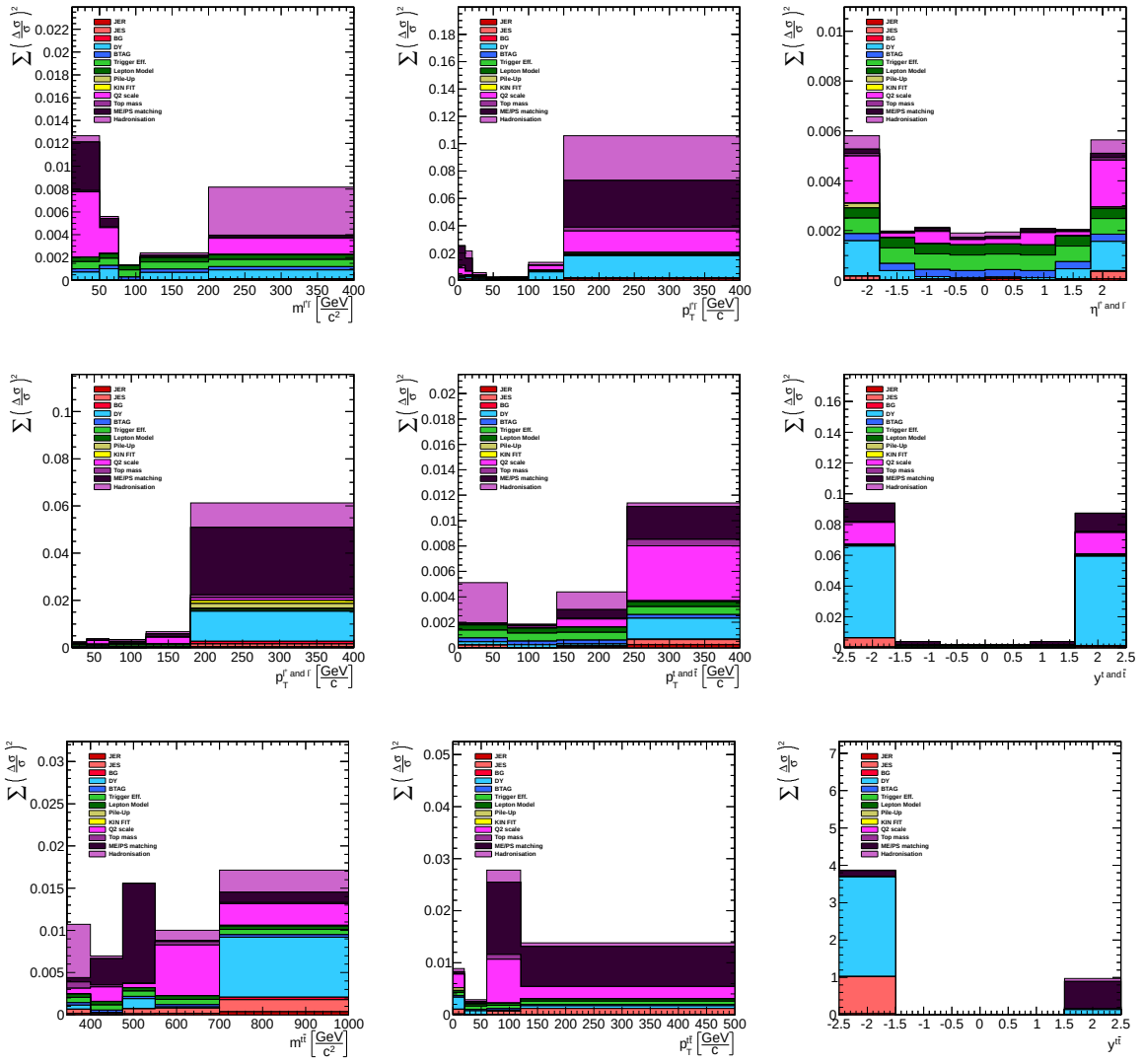


Figure B.1: The plots show the squares of the systematic uncertainties in all bins of all measured cross section quantities for the *ee* channel.

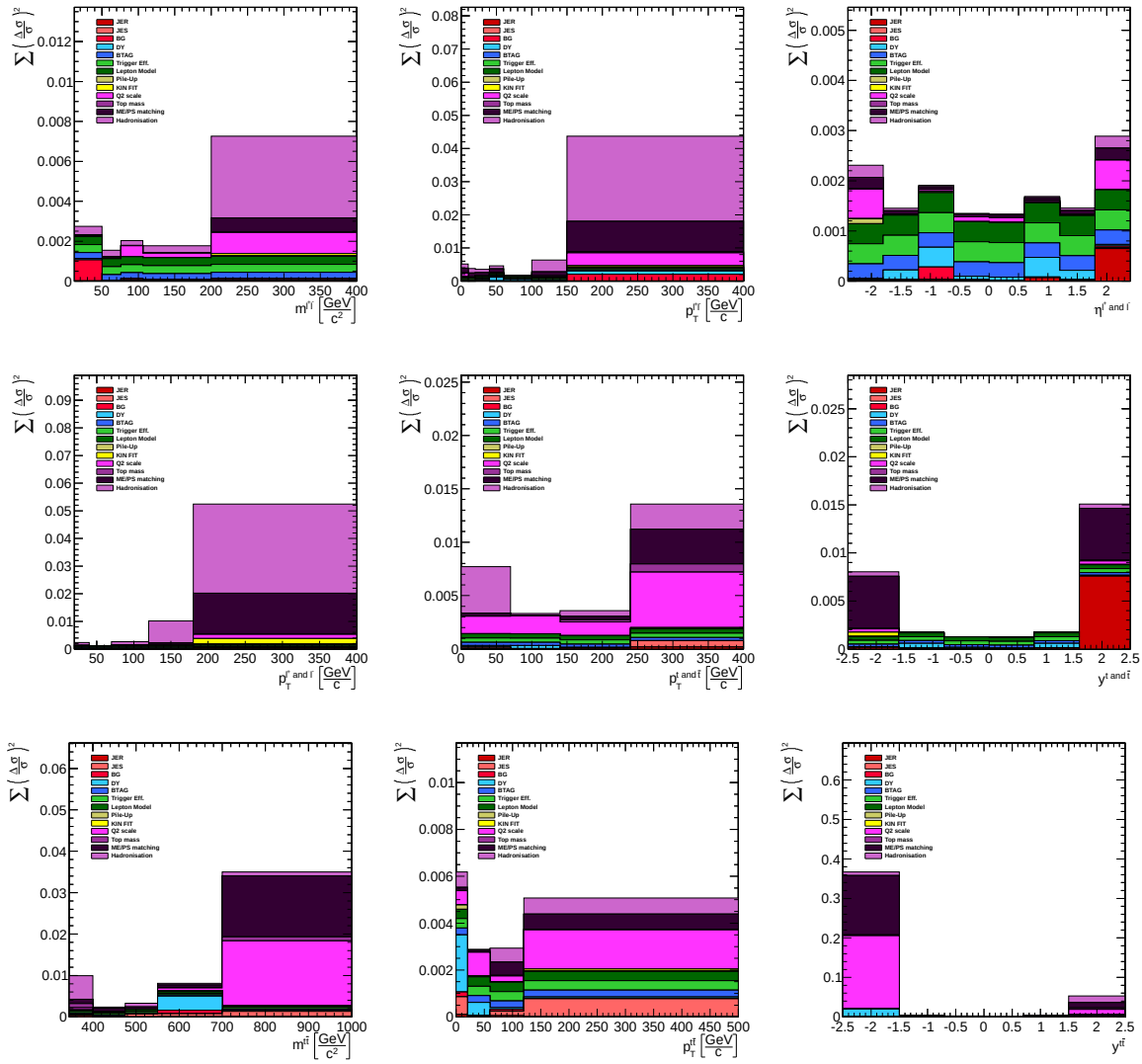


Figure B.2: The plots show the squares of the systematic uncertainties in all bins of all measured cross section quantities for the $e\mu$ channel.

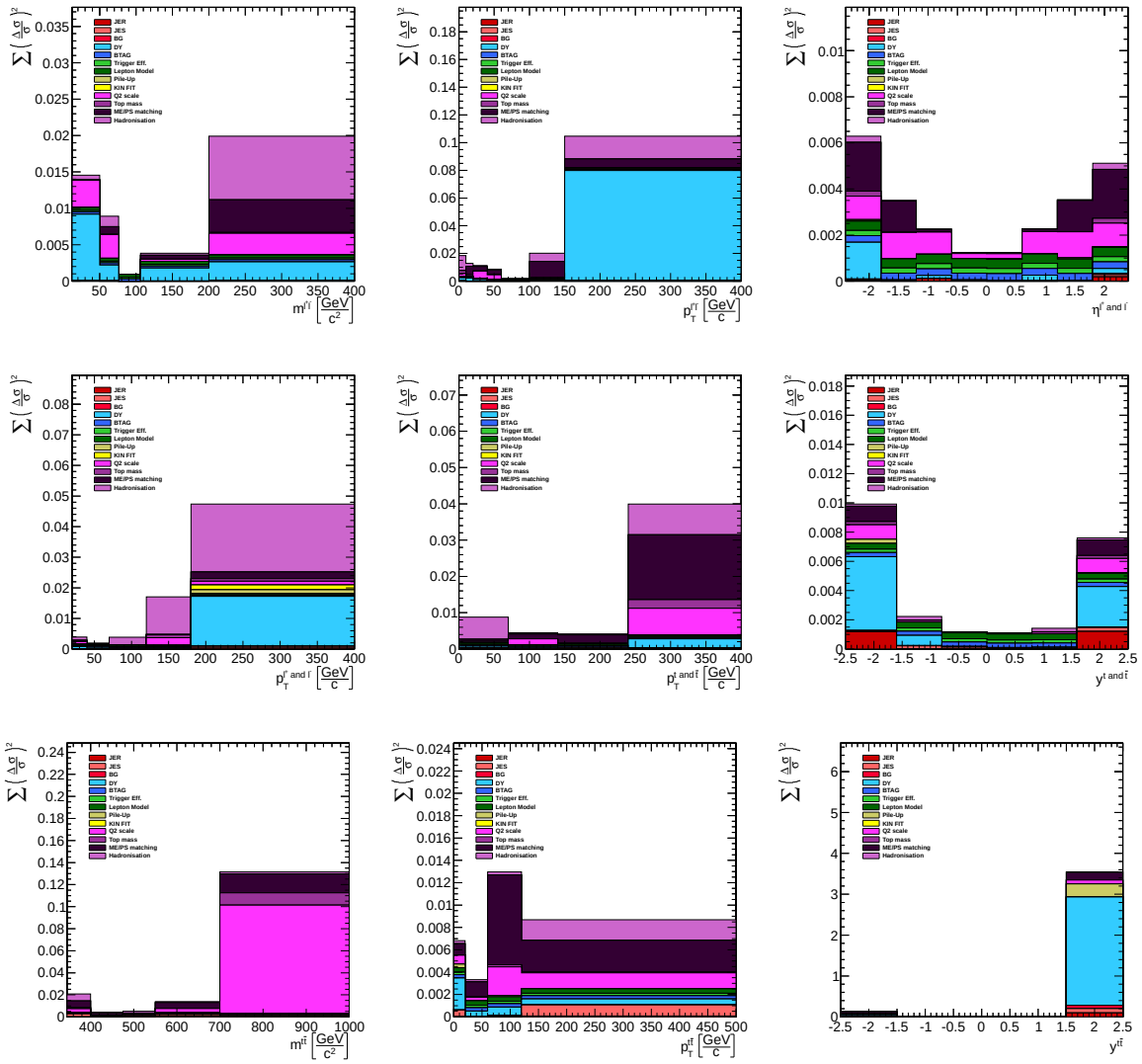


Figure B.3: The plots show the squares of the systematic uncertainties in all bins of all measured cross section quantities for the $\mu\mu$ channel.

Appendix C

Smoothing the differential cross sections

All normalised differential cross sections were calculated in histograms, even the smooth curves for the generator. The difference is that a very fine binning was chosen. This histogram was then drawn with ROOT's *C* option to retrieve a smooth curve. However, this didn't lead to good results in regions with low statistics. Therefore parts of the theory curves were fitted using a fit function as shown in table C.1. Furthermore, the fits were used to symmetrise η and y distributions. It is important that the fit functions have the only purpose to describe the simulation in a certain range, they do NOT describe a real functional dependence between the cross section and the corresponding quantity—in particular, they can produce completely wrong results outside the range in which they were applied. Checks were made by eye to make sure that the fit results really describe the binned histograms (done by putting both the histogram and the fit result in the same plot).

Quantity	Fit function	Fit range	Applied range
$p_T^{\ell^+ \text{ and } \ell^-}$	TMath::Exp(x*[0]+x*x*[1])*[2]	100 – 380 GeV	120 – 400 GeV
$p_T^{\ell\ell}$	TMath::Exp(x*[0]+x*x*[1])*[2]	150 – 470 GeV	150 – 400 GeV
$p_T^{t \text{ and } \bar{t}}$	TMath::Exp([0]*x)*[1]	200 – 360 GeV	200 – 400 GeV
$p_T^{\mathrm{t\bar{t}}}$	TMath::Exp(x*[0]+x*x*[1])*[2]	70 – 500 GeV	70 – 500 GeV
$\eta^{\ell^+ \text{ and } \ell^-}$	[0]*TMath::Gaus(x, [1], [2])	−2.4 – +2.4	−2.4 – +2.4
$y^{t \text{ and } \bar{t}}$, $y^{\mathrm{t\bar{t}}}$	TMath::Exp(x*x*[0]+x*x*x*x*[1]+ x*x*x*x*x*x*[2])*[3]	−2.5 – +2.5	−2.5 – +2.5
$m^{\ell\ell}$	TMath::Exp([0]*x)*[1]	170 – 350 GeV	175 – 400 GeV
$m^{\mathrm{t\bar{t}}}$	TMath::Exp([0]*x)*[1]	550 – 900 GeV	550 – 1000 GeV

Table C.1: Smoothing functions used to fit the final results. The functions are given in ROOT's syntax where [1], [2] and [3] stand for constants which need to be fitted. The *fit range* is the range in which the function was fitted to the histogram, the *applied range* is the range in which the fit was used.

Appendix D

Bias due to bin by bin corrections

In section 9.6.4 it is observed that the correlation between the generated and reconstructed top quark transverse momentum was only 0.67 and that thus the purity and stability were only in the order of 50 to 60 %. Therefore closure checks were performed for the top quark transverse momentum and for the top quark rapidity to show that the measurement is self-consistent and sensitive to changes in the data, i.e. that the bias introduced by the binwise calculation is acceptable.

To perform these checks, an equivalent of 1 fb^{-1} (i.e. same luminosity as in data) of simulation from the $t\bar{t}$ MadGraph $\text{t}\bar{\text{t}}$ sample is used to replace the data. These events are called *fake data* in the following.

In the next step, the new weights are assigned to the fake data events. The weighting is based on the true level values in the fake data, all other analysis steps use the reconstruction level values only because the fake data is then treated as real data. With these weighted fake data the full analysis is performed again, including the efficiency corrections from the unweighted standard $\text{t}\bar{\text{t}}$ MadGraph sample. As a last step, the results of these toy analyses are checked.

As only differences in the shape contribute to the final result, the test distributions are modified by a linear change of the weight. Both, the t and $\bar{t}$ quantities are modified and the separate weights are multiplied. For the top p_T distribution the following function is used to calculate the event weight w:

$$w = \min\left(\left(1 + \left(\frac{p_T^t}{\text{GeV}} - 100\right)s\right)\left(1 + \left(\frac{p_T^{\bar{t}}}{\text{GeV}} - 100\right)s\right), 0.1\right) \tag{D.1}$$

Here, p_T^t and $p_T^{\bar{t}}$ are the transverse momentum of the t resp. $\bar{t}$ at generator truth level, and s is the slope that is varied. The min function makes sure that each event keeps at least a weight of 0.1. The results for different values for positive slopes can be found in figure D.1, the results for the same slopes with negative sign can be found in figure D.2.

The calculation for the weights of the top rapidity distribution works similar. The following function is used:

$$w = \min\Big(\left(1 + (|y^t| - 1)s\right)\left(1 + (|y^{\bar{t}}| - 1)s\right), 0.1\Big) \tag{D.2}$$

Here, y is the rapidity at generator truth level, and s is again the slope that is varied. The result of the study for different values for positive slopes can be found in figure D.3, the results for the same slopes with negative sign can be found in figure D.4.

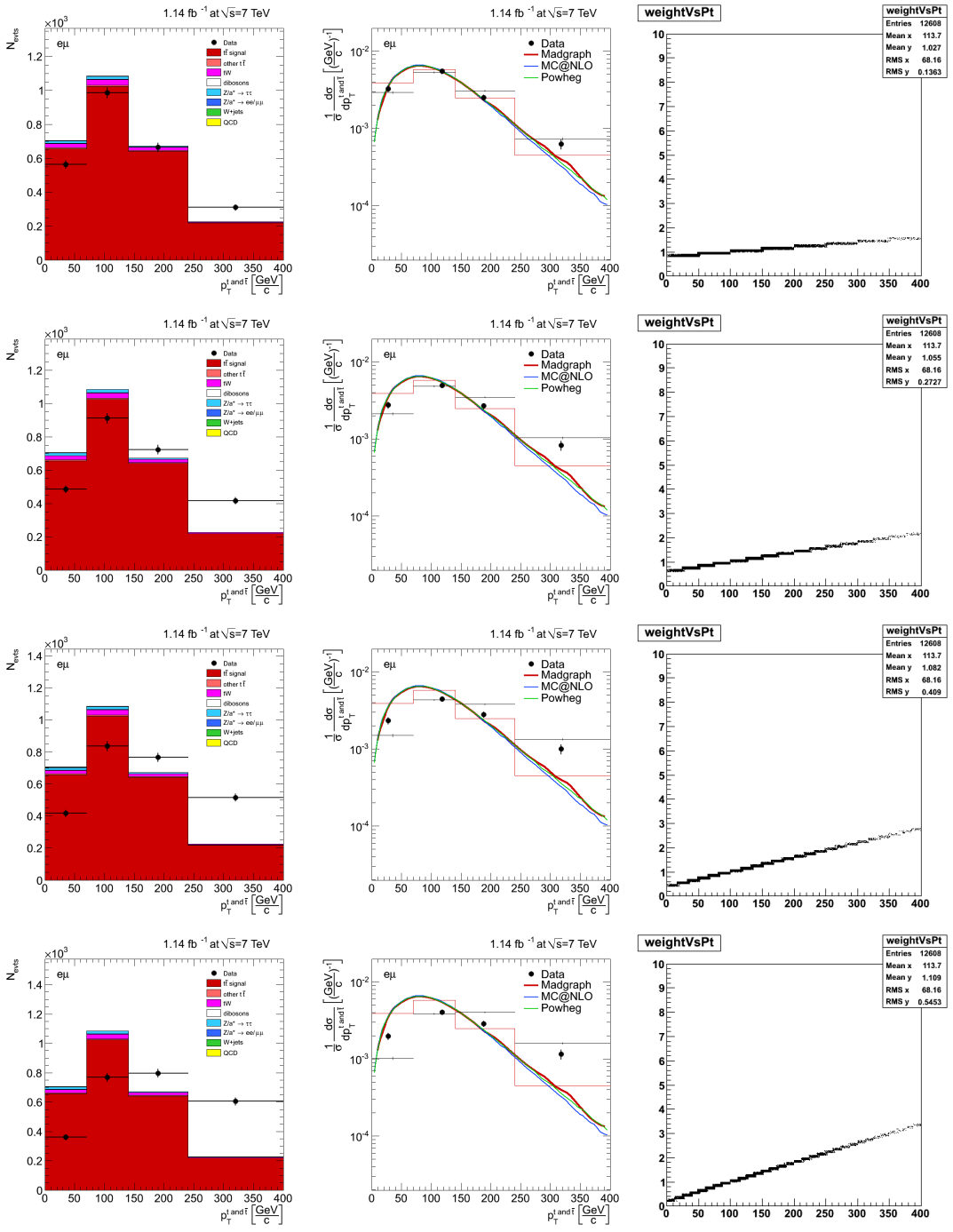


Figure D.1: Quantity: $p_T^{t\,\mathrm{and}\,\bar{t}}$
Event weight: $w = \min\left(\left(1 + \left(\frac{p_{T,t}}{\mathrm{GeV}} - 100\right)s\right)\left(1 + \left(\frac{p_{T,\bar{t}}}{\mathrm{GeV}} - 100\right)s\right), 0.1\right)$
Slopes (from top to bottom): $s = 0.002, 0.004, 0.006, 0.008$

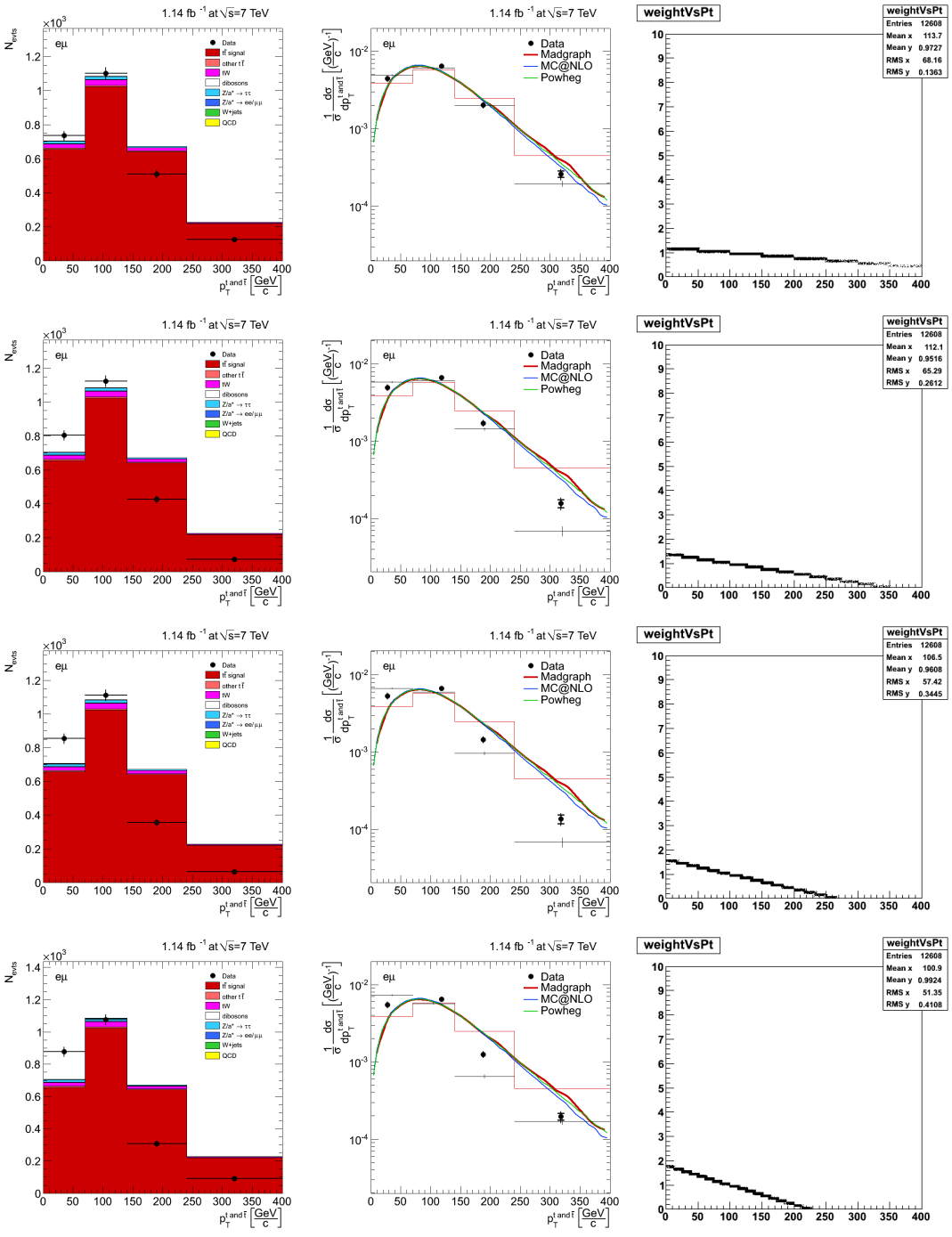


Figure D.2: Quantity: $p_T^{t\,\mathrm{and}\,\bar{t}}$
Event weight: $w = \min\left(\left(1 + \left(\frac{p_{T,t}}{\mathrm{GeV}} - 100\right)s\right)\left(1 + \left(\frac{p_{T,\bar{t}}}{\mathrm{GeV}} - 100\right)s\right), 0.1\right)$
Slopes (from top to bottom): $s = -0.002, -0.004, -0.006, -0.008$

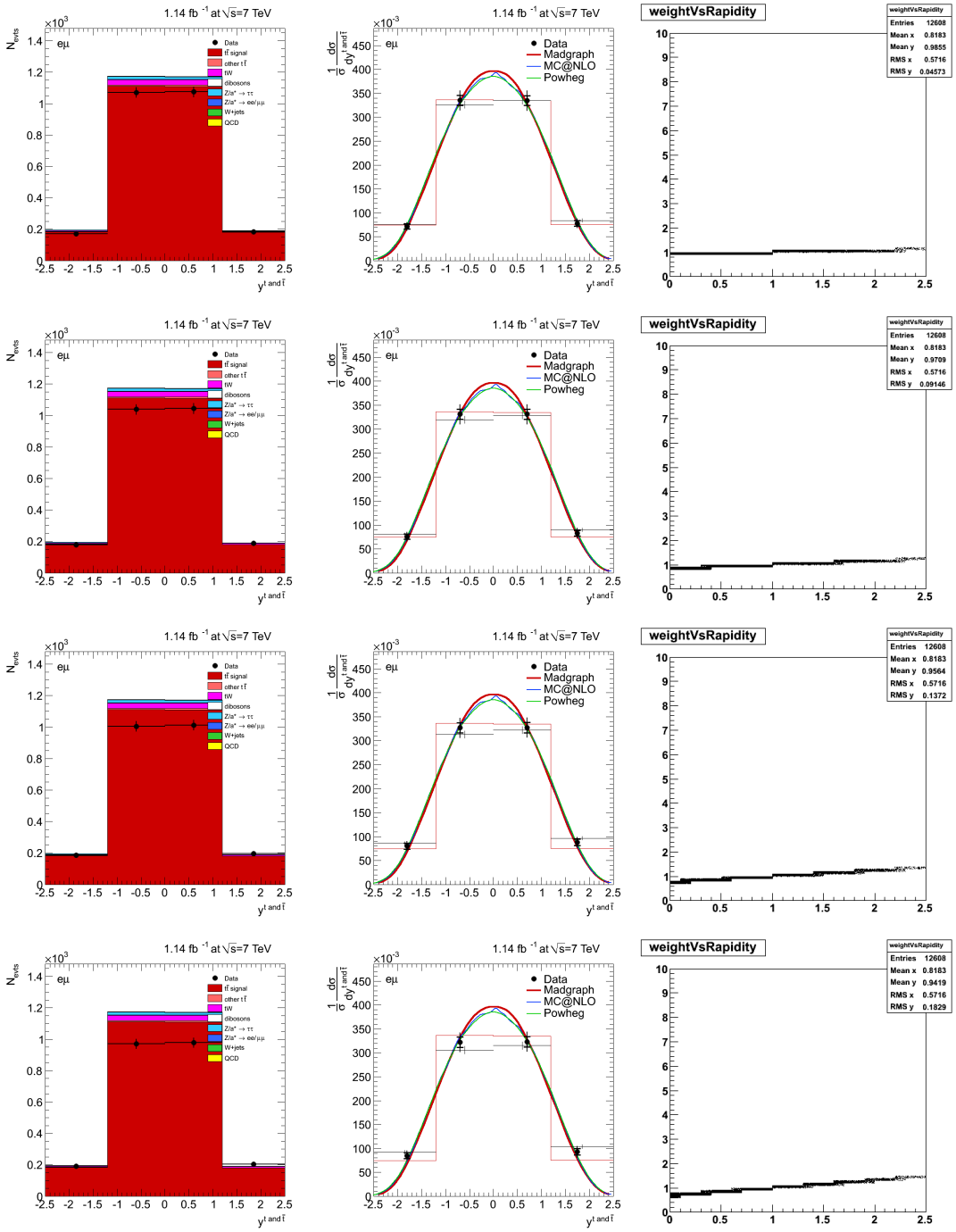


Figure D.3: Quantity: $y^{t\,\mathrm{and}\,\bar{t}}$
Event weight: $w = \min\Big(\left(1 + (|y_t| - 1)s\right)\left(1 + (|y_{\bar{t}}| - 1)s\right), 0.1\Big)$
Slopes (from top to bottom): $s = 0.08, 0.16, 0.24, 0.32$

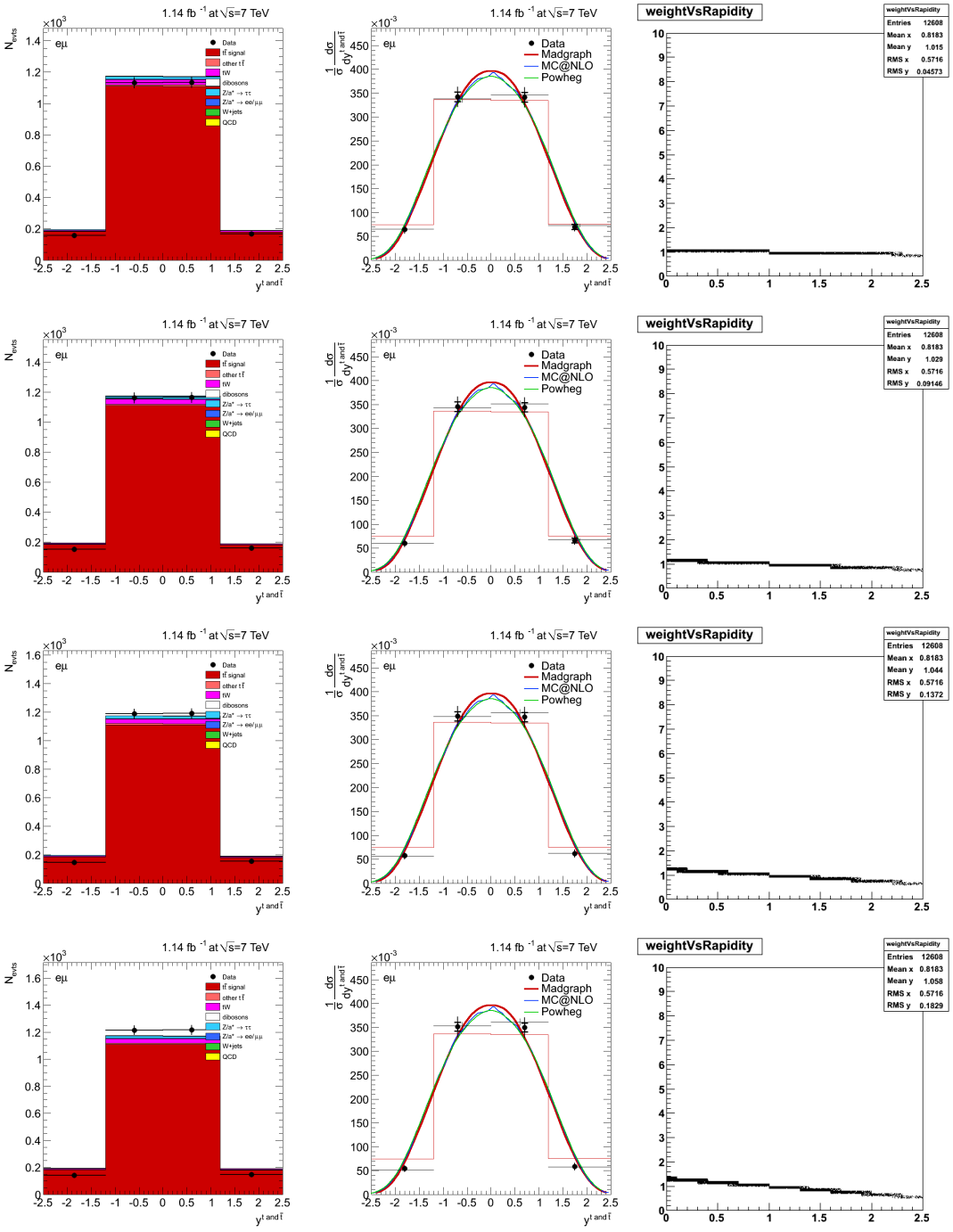


Figure D.4: Quantity: $y^{t \text{ and } \bar{t}}$

Event weight: $w = \min\Big((1 + (|y_t| - 1)s)\,(1 + (|y_{\bar{t}}| - 1)s)\,, 0.1 \Big)$

Slopes (from top to bottom): $s = -0.08, -0.16, -0.24, -0.32$

D.1 Summary of the shape variations

The resulting cross sections shown in figures D.1 to D.4 and the calculated fake data cross sections on true level are summarised in figure D.5.

The plots show the default (true level) cross section from MadGraph using black crosses. This is the sample used for the binwise cross section calculation. The red crosses show the true level cross section of the weighted fake data. There are eight different points in every bin corresponding to the eight different slopes. One can see that the resulting cross sections for the different slopes deviate from the default result. The red points finally correspond to the fake data measurement. The important thing to see is that the reconstructed distributions (red points) follow the modified generated values (red crosses) for small deviations from the standard MC simulation. This is valid for both observables.

A bias towards the default MadGraph result is however visible for large deviations of the fake data from the standard simulation. However, the event yield control plots on the left of figures D.1 – D.4 show that larger deviations from the standard simulation are well visible before the cross section calculation.

As the control plots for all nine different observables agree well with data, it can be concluded that that true cross sections in real data do not deviate much from the simulation. Therefore it is measured in a region where only a very small bias is introduced.

In case there would be a larger disagreement in data, one could either reweight the signal sample so that the measured cross section matches the reweighted signal sample cross section and redo the analysis; or one could take bin migrations from other bins into account. For a first differential cross section measurement and because of the good agreement, this is not necessary in this work.

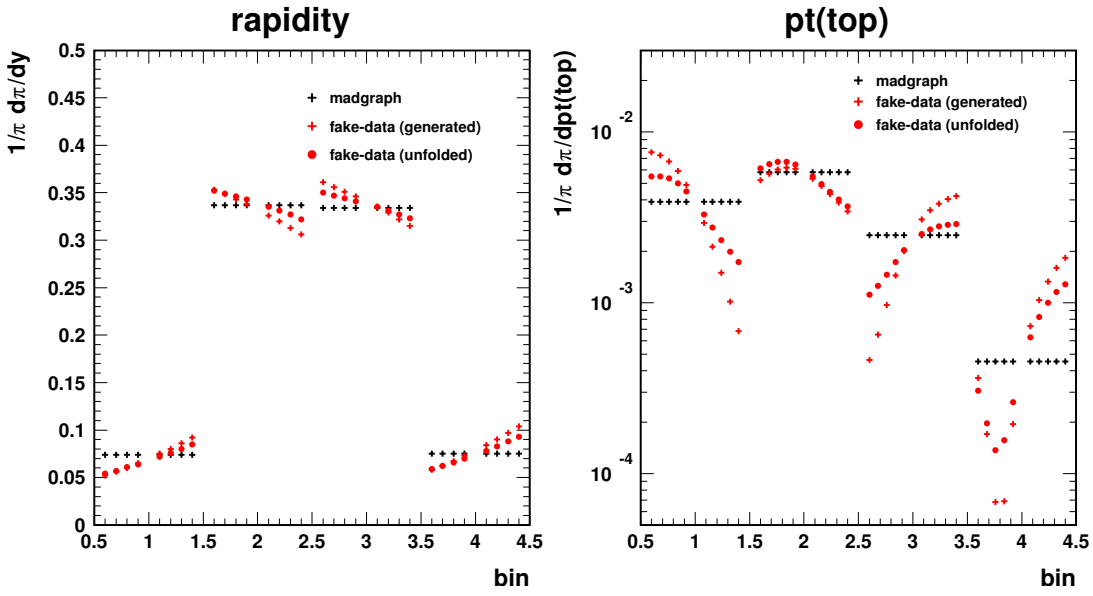
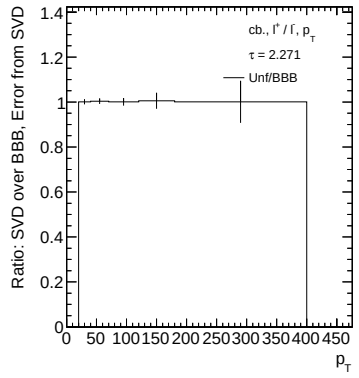


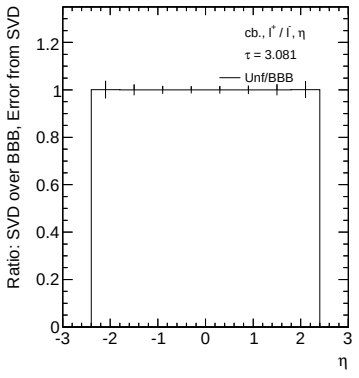
Figure D.5: Summary of the shape variations: the black crosses show the original MadGraph values, the red crosses show the true values of the varied fake data sample and the red points show the reconstructed values for the fake data determined with the non-reweighted MadGraph $t\bar{t}$ sample.
A thank goes to Andreas Meyer for putting all calculated variations into the single plots shown here.

D.2 Ratio SVD/BBB

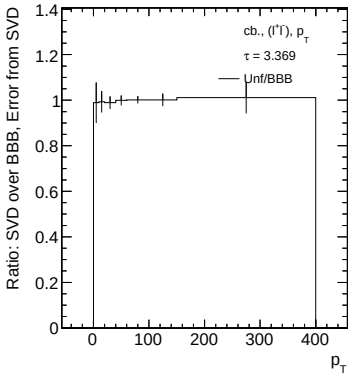
As a last proof that binwise cross section calculation works, the ratio between SVD unfolding and binwise calculation has been calculated in the updated analysis (chapter 10). The result is a good agreement even in top quark quantities, and even if the binning has changed and is now optimised for SVD. The ratios are shown in figure D.6 on the following page.



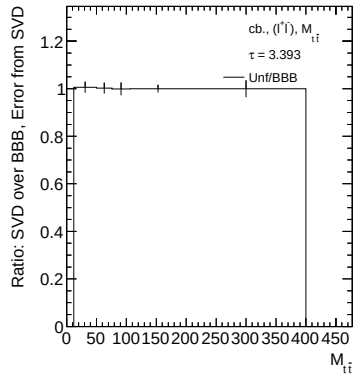
(a) Lepton p_T



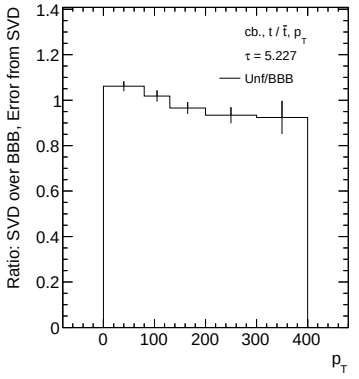
(b) Lepton η



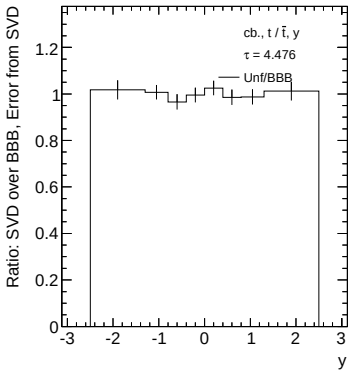
(c) Lepton pair p_T



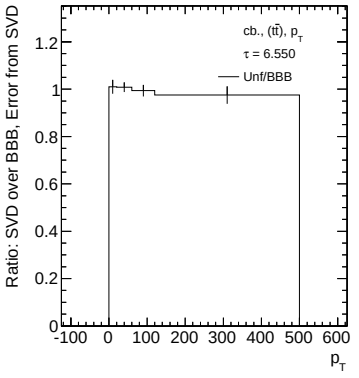
(d) Lepton pair mass



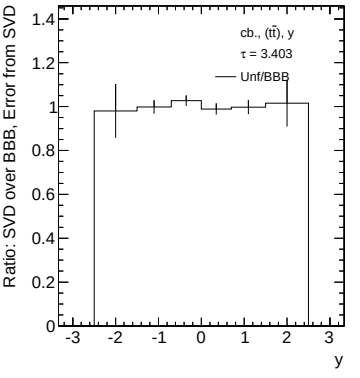
(e) Top quark p_T



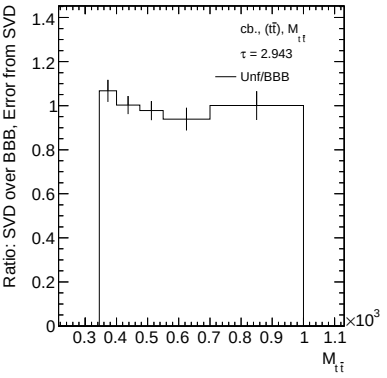
(f) Top quark rapidity



(g) Top quark pair p_T



(h) Top quark pair rapidity



(i) Top quark pair mass

Figure D.6: Comparison between full unfolding and binwise cross section calculations [153].

Appendix E

The nafJobSplitter

During this work, a small program called `nafJobSplitter.pl` has been developed to split any ntuple or histogram producing `cmsRun` job. Originally intended as a 5-minute proof of principle that job splitting should be easy, it is now not only in use by the DESY top group, but also by several people from the SUSY group and the University of Hamburg.

The program is simple to use: instead of running `cmsRun config.py`, one executes `nafJobSplitter.pl N config.py` where N is the number of jobs. There is no need to write configuration files or set up a special environment. The program will then automatically split the job into N sub-jobs, based on input files, i.e. one cannot have more jobs than input files. The sub-jobs are submitted to the batch farm of the National Analysis Facility (NAF). Once all jobs have run successfully, the output of the sub-jobs is merged automatically by the last job.

In case of errors, the sub-jobs that have failed can be determined and resubmitted automatically.

Should there be a large dataset to process and the jobs need more time to process the data than allowed by the batch system, then a terminate signal (actually, `SIGUSR1` or `SIGXCPU`) is sent shortly before the time limit is reached. This signal is caught and the corresponding job is terminated gracefully, i.e. it writes out the data it has already processed. The `nafJobSplitter.pl` now allows to resubmit this job such that it skips the events which have already been processed before the timeout. Therefore no processing time is lost even when the queue with a too short allowed runtime was selected.

There are also three main disadvantages. The first one is that the software is specialised for the NAF and doesn't run elsewhere. So if a dataset is not stored at DESY, a different software has to be used.

The second problem is that the program does not create independent batch jobs, it uses the local CMS software area in which `cmsenv` was executed, and this directory must be accessible from all batch worker nodes. That also means that one must not recompile or modify python files while jobs are running. In practice, this is usually not a big problem because one can simply setup a new software area in a different directory.

The third problem is that the software is only maintained by me, i.e. it might simply stop working, for example when changes are made on the batch system. On the other hand, it has been working fine for already more than a year and there was no need for substantial changes so far.

E.1 sumTriggerReports2.pl

Another tool in use also by other people is `sumTriggerReports2.pl`. It automatically joins the `TrigReport` sections of the output of `cmsRun` jobs to retrieve a cutflow. This script is used together with the `nafJobSplitter.pl` but can also be used to join output files from CRAB.

Both the `nafJobSplitter.pl` and `sumTriggerReports2.pl` can be found in the CMS CVS repository under `UserCode/Bromo/TopAnalysis/TopUtils/scripts`.

Bibliography

[1] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* **49** (1973) 652–657. doi:10.1143/PTP.49.652.

[2] S. W. Herb, D. C. Hom, L. M. Lederman et al., "Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions", *Phys. Rev. Lett.* **39** (1977) 252–255. doi:10.1103/PhysRevLett.39.252.

[3] The CDF Collaboration, F. Abe et al., "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab", *Phys. Rev. Lett.* **74** (1995) 2626–2631. doi:10.1103/PhysRevLett.74.2626.

[4] The DØ Collaboration, S. Abachi et al., "Observation of the Top Quark", *Phys. Rev. Lett.* **74** (1995) 2632–2637. doi:10.1103/PhysRevLett.74.2632.

[5] Brockhaus-Verlag, ed., "Brockhaus Enzyklopädie". F. A. Brockhaus GmbH, 21 edition, 2005.

[6] M. E. Peskin and D. V. Schroeder, "An Introduction to Quantum Field Theory". Addison-Wesley, 1995.

[7] MissMJ (pseudonym), "File:Standard Model of Elementary Particles.svg". http://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg, accessed on 2012-04-30.

[8] D. H. Perkins, "Hochenergiephysik". Addison-Wesley, 1991.

[9] P. Schmüser, "Feynman-Graphen und Eichtheorien für Experimentalphysiker". Springer, 1995.

[10] F. Jegerlehner, "The running fine structure constant alpha(E) via the Adler function", *Nucl. Phys. Proc. Suppl.* **181–182** (2008) 135–140, arXiv:0807.4206v1. doi:10.1016/j.nuclphysbps.2008.09.010.

[11] The L3 Collaboration, M. Acciarri et al., "Measurement of the Running of the Fine-Structure Constant", *Phys. Lett. B* **476** (2000) 40–48, arXiv:hep-ex/0002035v1. doi:10.1016/S0370-2693(00)00122-2.

[12] K. Nakamura et al. (Particle Data Group), "2010 Review of Particle Physics", *J. Phys. G* **37** (2010) 075021.

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

[14] S. L. Glashow, "Partial Symmetry of Weak Interactions", *Nucl. Phys.* **22** (1961) 579.

[15] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266.

[16] Nobelprize.org, "The Nobel Prize in Physics 1979". `http://www.nobelprize.org/nobel_prizes/physics/laureates/1979/`. [URL accessed 2012-08-19].

[17] H. Cavendish, "Experiments to Determine the Density of the Earth", *Phil. Trans. R. Soc.* **88** (1798) 469–526.

[18] The DØ Collaboration, V. M. Abazov et al., "Study of the decay $B_s^0 \to J/\psi f_2'(1525)$ in $\mu^+\mu^- K^+K^-$ final states", *submitted to Phys. Rev. D* (2012) `arXiv:1204.5723`.

[19] The DØ Collaboration, V. M. Abazov et al., "Dependence of the $t\bar{t}$ production cross section on the transverse momentum of the top quark", *Phys. Lett. B* **693** (2010) 515–521, `arXiv:1001.1900`. `doi:10.1016/j.physletb.2010.09.011`.

[20] The CDF Collaboration, T. Aaltonen et al., "First Measurement of the $t\bar{t}$ Differential Cross Section $d\sigma/dM_{t\bar{t}}$ in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. Lett.* **102** (2009) 222003, `arXiv:0903.2850`. `doi:10.1103/PhysRevLett.102.222003`.

[21] W. J. Stirling, "Standard Model cross sections as a function of collider energy". private communication, `http://www.hep.phy.cam.ac.uk/~wjs/plots/plots.html`.

[22] N. Kidonakis, "Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution", *Phys. Rev. D* **82** (2010) 114030. `doi:10.1103/PhysRevD.82.114030`.

[23] T. M. Liss and A. Quadt, "2012 Review of Particle Physics: The Top Quark", *Phys. Rev. D* **86** (2012) 010001.

[24] W. Beenakker, H. Kuijf, W. L. van Neerven et al., "QCD corrections to heavy-quark production in $p\bar{p}$ collisions", *Phys. Rev. D* **40** (1989) 54–82. `doi:10.1103/PhysRevD.40.54`.

[25] N. Kidonakis, "Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production", *Phys. Rev. D* **83** (2011) 091503. `doi:10.1103/PhysRevD.83.091503`.

[26] N. Kidonakis, "NNLL resummation for s-channel single top quark production", *Phys. Rev. D* **81** (2010) 054028. `doi:10.1103/PhysRevD.81.054028`.

[27] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a $W-$ or $H-$", *Phys. Rev. D* **82** (2010) 054018. doi:10.1103/PhysRevD.82.054018.

[28] The CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the fully hadronic decay channel in *pp* collisions at 7 TeV". CMS Physics Analysis Summary CMS-PAS-TOP-11-007, 2011.

[29] H. Enderle, M. Görner, J. Lange et al., "Measurement of the differential cross sections for top quark pair production in the lepton+jets channel". CMS AN-11-254 [CMS internal, http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2011_254_v11.pdf].

[30] M. Jeżabek and J. Kühn, "QCD corrections to semileptonic decays of heavy quarks", *Nucl. Phys. B* **314** (1989), no. 1, 1–6. doi:10.1016/0550-3213(89)90108-9.

[31] A. Quadt, "Top quark physics at hadron colliders", *EPJ C* **48** (2006) 835–1000. 10.1140/epjc/s2006-02631-6.

[32] A. Martin, W. Stirling, R. Thorne et al., "Parton distributions for the LHC", *EPJ C* **63** (2009) 189–285, arXiv:0901.0002. doi:10.1140/epjc/s10052-009-1072-5.

[33] J. Pumplin, D. R. Stump, J. Huston et al., "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis", *JHEP* **07** (2002) 012, arXiv:hep-ph/0201195.

[34] The H1 and ZEUS Collaboration, F. D. A. et al, "Combined Measurement and QCD Analysis of the Inclusive ep Scattering Cross Sections at HERA", *JHEP* **01** (2010) 109. doi:10.1007/JHEP01(2010)109.

[35] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964), no. 2, 132–133. doi:10.1016/0031-9163(64)91136-9.

[36] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509. doi:10.1103/PhysRevLett.13.508.

[37] K. Nakamura et al. (Particle Data Group), "2010 Review of Particle Physics: Hypothetical particles and Concepts, Higgs bosons: theory and searches", *J. Phys. G* **37** (2010) 075021.

[38] A. Stange, W. Marciano, and S. Willenbrock, "Higgs bosons at the Fermilab Tevatron", *Phys. Rev. D* **49** (1994) 1354–1362. doi:10.1103/PhysRevD.49.1354.

[39] W. Beenakker, S. Dittmaier, M. Krämer et al., "Higgs Radiation Off Top Quarks at the Tevatron and the LHC", *Phys. Rev. Lett.* **87** (2001) 201805. doi:10.1103/PhysRevLett.87.201805.

[40] CERN Press Office, "CERN experiments observe particle consistent with long-sought Higgs boson". CERN Press Release PR17.12 http://press.-web.cern.ch/press/PressReleases/Releases2012/PR17.12E.html. [online, accessed 2012-05-16].

[41] J. Incandela, "Status of the CMS SM Higgs Search". Slides from public CERN Higgs Seminar on July 4, 2012, https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=6125&filename=CMS_4July2012_Incandela.pdf. [online, accessed 2012-07-04].

[42] F. Gianotti, "Status of Standard Model Higgs searches in ATLAS". Slides from public CERN Higgs Seminar on July 4, 2012, https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=6126&version=1&filename=Higgs-CERN-seminar-2012.pdf. [online, accessed 2012-07-04].

[43] The ATLAS Collaboration, T. A. Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *submitted to Phys. Lett. B* (2012) arXiv:1207.7214.

[44] The CMS Collaboration, T. C. Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *submitted to Phys. Lett. B* (2012) arXiv:1207.7235.

[45] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008), no. 08, S08001.

[46] CERN Press Office, "Incident in LHC sector 3-4". CERN Press Release PR09.08 http://press.web.cern.ch/press/PressReleases/Releases2008/PR09.08E.html. [online, accessed 2012-05-16].

[47] The CMS Collaboration, "LHC Commissioning". http://lhc-commissioning.web.cern.ch/lhc-commissioning/. [online, accessed 2012-05-16].

[48] G. Papotti, "LHC operations: status and prospects". https://indico.cern.ch/getFile.py/access?contribId=0&resId=0&materialId=slides&confId=166286. [slides in Latest results from the LHC on 2012-07-12, URL accessed 2012-07-12].

[49] CERN Communication Group, "CERN FAQ – LHC the guide". http://doc.cern.ch//archive/electronic/cern/others/multimedia/brochure/brochure-2006-003-eng.pdf, January 2008.

[50] "PS – the Proton Synchrotron". http://public.web.cern.ch/public/en/Research/PS-en.html. [online, accessed 2012-05-17].

[51] "SPS – the Super Proton Synchrotron". http://public.web.cern.ch/public/en/Research/SPS-en.html. [online, accessed 2012-05-17].

[52] The ALICE Collaboration, K. Aamodt et al., "The ALICE experiment at the CERN LHC", *JINST* **3** (2008), no. 08, S08002.

[53] The ATLAS Collaboration, G. Aad et al., "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003. doi:10.1088/1748-0221/3/08/S08003.

[54] The LHCb Collaboration, A. Augusto et al., "The LHCb Detector at the LHC", *JINST* **3** (2008), no. 08, S08005.

[55] The LHCb Collaboration, R. Aaij et al., "Measurement of J/psi production in pp collisions at sqrt(s)=7 TeV", *Eur. Phys. J.* **C71** (2011) 1645, arXiv:1103.0423. doi:10.1140/epjc/s10052-011-1645-y.

[56] The LHCf Collaboration, O. Adriani et al., "The LHCf detector at the CERN Large Hadron Collider", *JINST* **3** (2008), no. 08, S08006.

[57] The TOTEM Collaboration, G. Anelli et al., "The TOTEM Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008), no. 08, S08007.

[58] The CMS Collaboration, S. Chatrchyan et al., "The CMS experiment at the CERN LHC", *JINST* **3** (2008), no. 08, S08004. doi:10.1088/1748-0221/3/08/S08004.

[59] The CMS Collaboration, S. Chatrchyan et al., "CMS TDR Volume 1 figure gallery". http://cmsdoc.cern.ch/cms/cpt/tdr/ptdr1-figs/Figures/Figure_001-002.pdf.

[60] The CMS Collaboration, "Description and Performance of the CMS Track and Primary Vertex Reconstruction". CMS Physics Analysis Summary CMS-TRK-11-001, 2012.

[61] The CMS Collaboration, "Electromagnetic calorimeter calibration with 7 TeV data". CMS Physics Analysis Summary CMS-EGM-10-003.

[62] The CMS Collaboration, S. Chatrchyan et al., "Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data", *JINST* **5** (2010), no. 03, T03012.

[63] The CMS Collaboration, S. Chatrchyan et al., "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011), no. 11, P11002.

[64] D. E. Groom, N. V. Mokhov, and S. I. Striganov, "Muon stopping power and range tables 10-MeV to 100-TeV", *Atom. Data Nucl. Data Tabl.* **78** (2001) 183–356. doi:10.1006/adnd.2001.0861.

[65] The CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV". CMS Physics Analysis Summary CMS-MUO-10-004, 2012.

[66] The CMS Collaboration, S. Dasu et al., "The TriDAS Project, Technical Design Report, Volume 1: The Trigger Systems". CERN-LHCC-2000-038 / CMS-TDR-006-1, http://cdsweb.cern.ch/record/706847.

[67] The CMS Collaboration, Sphicas, P., (ed.), "The TriDAS Project, Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger". CERN-LHCC-2002-026 / CMS-TDR-006-add-2, http://cdsweb.cern.ch/record/578006.

[68] The CMS Collaboration, W. Adam et al., "The CMS high level trigger", *EPJC* **46** (2006), no. 3, 605–667. doi:10.1140/epjc/s2006-02495-8.

[69] L. Tuura, A. B. Meyer, I. Segoni et al., "CMS data quality monitoring: systems and experiences", *JPCS* **219** (2010) 072020. doi:10.1088/1742-6596/219/7/072020.

[70] L. Tuura, G. Eulisse, and A. B. Meyer, "CMS data quality monitoring web service", *JPCS* **219** (2010) 072055. doi:10.1088/1742-6596/219/7/072055.

[71] C. Grandi, D. Stickland, and L. Taylor, "The CMS Computing Model", technical report, CERN, 2004. CERN-LHCC-2004-035/G-083.

[72] Worldwide LHC Computing Grid Collaboration, "Memorandum of Understanding for Collaboration in the Deployment and Exploitation of the Worldwide LHC Computing Grid". CERN-C-RRB-2005-01/Rev. 1, Version 29 March 2011, http://wlcg.web.cern.ch/collaboration/mou.

[73] E. C. Bailey, "Maximum RPM". Sams, 1997.

[74] E. Foster-Johnson, "Red Hat RPM Guide". Red Hat, 2003.

[75] S. Metson et al., "SiteDB: marshalling people and resources available to CMS", *JPCS* **219** (2010) 072044.

[76] W. Behrenhoff, C. Wissing, B. Kim et al., "Deployment of the CMS software on the WLCG Grid", *JPCS* **331** (2011), no. 7, 072041.

[77] J. Blomer, C. Aguado-Sanchez, P. Buncic et al., "Distributing LHC application software and conditions databases using the CernVM file system", *JPCS* **331** (2011), no. 4, 042003.

[78] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, arXiv:0709.2092. doi:10.1088/1126-6708/2007/11/070.

[79] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **06** (2002) 29, arXiv:hep-ph/0204244.

[80] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys. B* **126** (1977) 298. doi:10.1016/0550-3213(77)90384-4.

[81] Floßdorf, Alexander, "Higher Order QCD Radiation in Top Pair Production with the CMS Detector". DESY-Thesis-2009-038, ISSN 1435-8085.

[82] A. Buckley, J. Butterworth, S. Gieseke et al., "General-purpose event generators for LHC physics", *Phys. Rep.* **504** (2011), no. 5, 145–233. doi:10.1016/j.physrep.2011.03.005.

[83] F. James, "Monte Carlo theory and practice", *Reports on Progress in Physics* **43** (1980), no. 9, 1145.

[84] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175. doi:10.1088/1126-6708/2006/05/026.

[85] G. Corcella, I. Knowles, G. Marchesini et al., "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)", *JHEP* **01** (2001) 010, arXiv:hep-ph/0011363.

[86] S. Mrenna and P. Richardson, "Matching matrix elements and parton showers with HERWIG and PYTHIA", *JHEP* **05** (2004) 040.

[87] G. Dissertori, I. Knowles, and M. Schmelling, "Quantum Chromodynamics, High Energy Experiments and Theory". Oxford University Press, 2003.

[88] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph", *JHEP* **02** (2003) 027, arXiv:hep-ph/0208156.

[89] T. Chwalek, "Messung der W-Boson-Helizitätsanteile in Top-Quark-Zerfallen mit dem CDF II Experiment und Studien zu einer frühen Messung des $t\bar{t}$-Wirkungsquerschnitts mit dem CMS Experiment". CERN-THESIS-2010-255.

[90] J. Campbell, K. Ellis, and C. Williams, "MCFM - Monte Carlo for FeMtobarn processes". `http://mcfm.fnal.gov/`.

[91] D. Dammann, "Production Cross Section Measurement of Top-Antitop Pairs in the Dimuon Decay Channel at $\sqrt{s} = 7\,\text{TeV}$ with the CMS Experiment". DESY-Thesis-2011-016, ISSN 1435-8085.

[92] S. Alioli, P. Nason, C. Oleari et al., "NLO vector-boson production matched with shower in POWHEG", *JHEP* **07** (2008) 060, `arXiv:0805.4802`. `doi:10.1088/1126-6708/2008/07/060`.

[93] S. Jadach, J. H. Kuhn, and Z. Was, "TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons", *Comput. Phys. Commun.* **64** (1990) 275–299. `doi:10.1016/0010-4655(91)90038-M`.

[94] Z. Was, "TAUOLA for simulation of tau decay and production: perspectives for precision low energy and LHC applications", *Nucl. Phys. Proc. Suppl.* **218** (2011) 249–255, `arXiv:1101.1652`. `doi:10.1016/j.nuclphysbps.2011.06.040`.

[95] S. Agostinelli et al., "Geant4–a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250–303. `doi:10.1016/S0168-9002(03)01368-8`.

[96] The CMS Collaboration, S. Chatrchyan et al., "Performance of CMS muon reconstruction in pp collision events at sqrt(s) = 7 TeV", *Submitted to JINST* (2012) `arXiv:1206.4071v1`.

[97] The CMS Collaboration, "Slice of the CMS detector". `http://cms-project-cmsinfo.web.cern.ch/cms-project-cmsinfo/Resources/Website/Media/Videos/Animations/files/CMS_Slice.gif`.

[98] The CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s} = 7$ TeV". CMS Physics Analysis Summary CMS-EGM-10-004.

[99] The CMS Collaboration, ""Cuts in Categories" (CiC) Electron Identification". `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideCategoryBasedElectronID`. CMS Public TWiki, URL accessed 2012-07-06].

[100] Sani, Matteo and Branson, Jim, "Cuts in Categories Electron Identification". `https://indico.cern.ch/getFile.py/access?contribId=2&resId=1&materialId=slides&confId=92412`. slides from Egamma Meeting, 2010-04-28, CMS internal, URL accessed 2012-07-06].

[101] D. Evans et al., "Realistic Analysis Level ElectronID Comparisons". `https://indico.cern.ch/getFile.py/access?contribId=6&resId=0&materialId=slides&confId=93821`. slides for Egamma Meeting, 2010-05-06, CMS internal, URL accessed 2012-07-06].

[102] P. Schieferdecker, "Jet Algorithms". Slides from Vivian's Meeting, accessed on 2012-07-01: `https://twiki.cern.ch/twiki/bin/viewfile/Sandbox/Lecture?rev=1;filename=Philipp_Schieferdeckers_Lecture.pdf`, 2009-04-17.

[103] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm", *JHEP* **04** (2008) 063.

[104] S. Catani, Y. L. Dokshitzer, M. H. Seymour et al., "Longitudinally invariant $K(t)$ clustering algorithms for hadron hadron collisions", *Nucl. Phys. B* **406** (1993) 187.

[105] Y. L. Dokshitzer, G. D. Leder, S. Moretti et al., "Better jet clustering algorithms", *JHEP* **08** (1997) 001.

[106] The CMS Collaboration, "Jet Performance in pp Collisions at $\sqrt{s} = 7\,\text{TeV}$". CMS Physics Analysis Summary JME-10-003, `http://cdsweb.cern.ch/record/1279362?ln=en`, 2010.

[107] M. Schröder, "Quality of Jet Measurements and Impact on a Search for New Physics at CMS". PhD Thesis, to be published 2012.

[108] K. Kousouris, "Update on dijet pt balance". `https://indico.cern.ch/getFile.py/access?contribId=2&resId=0&materialId=slides&confId=105277`. [slides in Jet Energy Corrections Meeting on 2010-08-27, CMS internal, URL accessed 2012-07-04].

[109] The CMS Collaboration, "Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET". CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.

[110] F. Beaudette, "Performance of the particle flow algorithm in CMS", Technical Report CMS-CR-2010-276, CERN, Geneva, 2010.

[111] The CMS Collaboration, "Particle flow reconstruction of 0.9 TeV and 2.36 TeV collision events in CMS". CMS Physics Analysis Summary CMS-PAS-PFT-10-001, 2010.

[112] The CMS Collaboration, "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV". CMS Physics Analysis Summary CMS-PAS-PFT-10-003, 2010.

[113] E. Schlieckau, "Private communication".

[114] C. Bernet, M. Bluj, M. Edelhoff et al., "An Introduction to PF2PAT". Slides from the PAT Tutorial on April 7, 2011 at CERN at `https://indico.cern.ch/getFile.py/access?contribId=18&resId=0&materialId=slides&confId=128291`. [online, CMS internal, accessed 2012-04-02].

[115] The CMS Collaboration, M. C. Llatas, "New world record – first pp collisions at 8 TeV". Event visualisation from the press release, inverted colours `http://cms.web.cern.ch/news/new-world-record-first-pp-collisions-8-tev`. [online, accessed 2012-07-04].

[116] A. Rizzi, F. Palla, and G. Segneri, "Track impact parameter based b-tagging with CMS". CMS Note 2006/019, `http://cms.cern.ch/iCMS/jsp/openfile.jsp?type=NOTE&year=2006&files=NOTE2006_019.pdf`, 2006.

[117] M. Marienfeld, "Measurement of the Top Quark Pair Production Cross Section in the Muon-Electron Decay Channel at $\sqrt{s} = 7\,\text{TeV}$ with the CMS Experiment". DESY-Thesis-2011-023, ISSN 1435-8085.

[118] JSON file July 6, `/afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions11/7TeV/Reprocessing/Cert_160404-163869_7TeV_May10ReReco_Collisions11_JSON_v2.txt`.

[119] JSON file July 6, `/afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions11/7TeV/Prompt/Cert_160404-167913_7TeV_PromptReco_Collisions11_JSON.txt`.

[120] The CMS Collaboration, "Measurement of the top dilepton cross section at $\sqrt{s} = 7\,\text{TeV}$ with 1.14 fb^{-1} of pp collisions". CMS Physics Analysis Summary TOP-11-005, `http://cdsweb.cern.ch/record/1377323/`, 2011.

[121] R. Chierici, F. Cossutti, G. G. Ceballos et al., "Standard Model Cross Sections for CMS at 7 TeV". `https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSections?rev=24`. [revision 24, CMS internal, URL accessed 2012-07-10].

[122] O. Gonzalez, "issue with madgraph sample?". `https://hypernews.cern.ch/HyperNews/CMS/get/generators/1228.html`, 2011-09-14. [send to *Generator Tools Discussions* hypernews, CMS internal].

[123] E. Migliore et al., "LHE-Pythia bug with 5 GeV cut". https://indico.cern.ch/getFile.py/access?contribId=4&resId=0&materialId=slides&confId=153867. [slides in VHF group meeting on 2011-11-01, CMS internal, URL accessed 2012-07-10].

[124] The DØ Collaboration, B. Abbott et al., "Measurement of the top quark mass using dilepton events", *Phys. Rev. Lett.* **80** (1998) 2063–2068, arXiv:hep-ex/9706014. doi:10.1103/PhysRevLett.80.2063.

[125] The CMS Collaboration, "Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update". CMS Physics Analysis Summary CMS-PAS-SMP-12-008, 2012.

[126] M. Hildreth, "Pileup Reweighting for 2011 Data and Monte Carlo". https://twiki.cern.ch/twiki/bin/view/CMS/Pileup_2011_Reweighting. [online, revision 2, CMS internal, accessed 2012-04-01].

[127] M. Hildreth, "Estimating Systematic Errors Due to Pileup Modeling". https://twiki.cern.ch/twiki/bin/view/CMS/PileupSystematicErrors?rev=3. [online, revision 3, CMS internal, accessed 2012-04-01].

[128] J. Andrea, A. Calderon, M. Cardaci et al., "Measurement of the top dilepton cross section using b-tagging at $\sqrt{s} = 7$ TeV with 882 pb^{-1} in pp collisions". CMS AN-11-334 [CMS internal, http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2011_334_v4.pdf].

[129] S. Choi, J. Goh, T. J. Kim et al., "Measurement of the $t\bar{t}$ differential cross section $d\sigma/dM_{t\bar{t}}$". CMS AN-11-225 [CMS internal, http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2011_225_v11.pdf].

[130] The CMS Collaboration, S. Chatrchyan et al., "Measurement of the t t-bar production cross section and the top quark mass in the dilepton channel in pp collisions at sqrt(s) =7 TeV", *JHEP* **07** (2011) 049, arXiv:1105.5661. doi:10.1007/JHEP07(2011)049.

[131] The CMS Collaboration, S. Chatrchyan et al., "Measurement of the $t\bar{t}$ production cross section in pp collisions at 7 TeV in lepton + jets events using b-quark jet identification", *Phys. Rev. D* **84** (2011) 092004. doi:10.1103/PhysRevD.84.092004.

[132] M. Aldaya, W. Behrenhoff, D. Dammann et al., "Measurement of Top Quark Pair Differential Cross Sections in the Dilepton Final State at $\sqrt{s} = 7$ TeV". CMS AN-11-186 [CMS internal, http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2011_186_v14.pdf].

[133] The CMS Collaboration, "Jet Energy Corrections determination at $\sqrt{s} = 7$ TeV". CMS Physics Analysis Summary CMS-PAS-JME-10-010, 2010.

[134] The CMS Collaboration, "Jet Energy Resolution in CMS at $\sqrt{s} = 7$ TeV". CMS Physics Analysis Summary CMS-PAS-JME-10-014, 2010.

[135] The CMS Collaboration, "Performance of b-Jet Identification in CMS". CMS Physics Analysis Summary CMS-PAS-BTV-11-001.

[136] D. Dammann, "Private communication".

[137] P. Silva, M. Gallinaro, and J. Varela, "Probing the heavy flavor content of top quark events in the dilepton channel with $\sqrt{s} = 7$ TeV data". CMS AN-2011-152 [CMS internal, `http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2011_152_v1.pdf`].

[138] The CMS Collaboration, S. Chatrchyan et al., "Measurement of the $\mathrm{t\bar{t}}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 7\,\mathrm{TeV}$", *submitted to JHEP* (2012).

[139] The CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 7$ TeV with a luminosity of 1.14 fb^{-1}". CMS Physics Analysis Summary CMS-PAS-TOP-11-005, 2011.

[140] T. Dorland, "Private communication".

[141] The CMS Collaboration, "Measurement of Top Quark Pair Differential Cross Sections at $\sqrt{s} = 7$ TeV". CMS Physics Analysis Summary CMS-PAS-TOP-11-013, 2012.

[142] J. Hartung, B. Elpelt, and K.-H. Klösener, "Statistik: Lehr- und Handbuch der angewandten Statistik". Oldenbourg Wissenschaftsverlag, 15 edition, 1999.

[143] M. Botje et al., "The PDF4LHC Working Group Interim Recommendations", `arXiv:1101.0538`.

[144] G. Lafferty and T. Wyatt, "Where to stick your data points: The treatment of measurements within wide bins", *Nucl. Instrum. Meth.* **A355** (1995) 541–547. `doi:10.1016/0168-9002(94)01112-5`.

[145] S. Naumann-Emme et al., "Overview of Top studies at the LHC (CMS and ATLAS)". Talk at the LC Forum 2012, `https://indico.desy.de/getFile.py/access?contribId=21&sessionId=11&resId=0&materialId=slides&confId=4980`.

[146] W. Behrenhoff et al., "Measurement of top quark differential cross sections at sqrt(s) = 7 TeV". Poster presented at the PLHC2012 conference, https://indico.cern.ch/getFile.py/access?contribId=18&sessionId=30&resId=0&materialId=poster&confId=164272.

[147] M. Aldaya et al., "Differential cross sections in top pair events at CMS". Talk at the ICHEP2012 conference, http://indico.cern.ch/getFile.py/access?contribId=586&sessionId=67&resId=0&materialId=slides&confId=181298.

[148] The CMS Collaboration, S. Chatrchyan et al., "Measurement of the Z/γ^*+b-jet cross section in pp collisions at 7 TeV", *JHEP* **06** (2012) 126, arXiv:1204.1643. doi:10.1007/JHEP06(2012)126.

[149] J. Hauk, "Measurement of Associated Z^0-Boson and b-Jet Production in Proton-Proton Collisions with the CMS Experiment". DESY-THESIS-2012-030, ISSN 1435-8085.

[150] M. Aldaya, W. Behrenhoff, I. A. Cruz et al., "Measurement of Top Quark Pair Differential Cross Sections in the Dilepton Final State". CMS AN-12-065 [CMS internal, http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2012_065_v5.pdf].

[151] 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.

[152] V. Blobel, "An Unfolding method for high-energy physics experiments", arXiv:hep-ex/0208022.

[153] D. J. Fischer, "Private communication".

[154] N. Kidonakis, "Top quark rapidity distribution and forward-backward asymmetry", *Phys. Rev. D* **84** (2011) 011504. doi:10.1103/PhysRevD.84.011504.

[155] The CMS Collaboration, S. Chatrchyan et al., "Measurement of Differential Top Quark Pair Production Cross Sections at $\sqrt{s} = 7$ TeV", *to be submitted to EPJC* (2012).

[156] The CMS Collaboration, "CMS Luminosity – Public Results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. [revision 80, URL accessed 2012-08-20].

Acknowledgements

All the people who have supported me during my time as a PhD student deserve a big *thank you*! Without you, the completion of this PhD thesis would not have been possible.

I would like to thank especially my supervisor Dr. Achim Geiser who made this work possible. He not only gave valuable help and made lots of suggestions for my work but he also taught me that writing down numbers on a piece of paper and doing a calculation manually is often very helpful, even when searching for a programming bug in a larger piece of code.

Very much appreciated is also that the referees of the dissertation and disputation, Prof. Peter Schleper and JProf. Christian Sander were available on short notice.

For the cross section measurement, Andreas Meyer gave lots of ideas to improve the analysis. He was always available for discussions about the physics, the technical implementation details, and organisational issues.

Furthermore, I received a lot of help from Maria Aldaya, for example when preparing the internal analysis approval talk under time pressure. She also coordinated the efforts of all groups participating in the CMS differential cross section measurement.

For all issues with computing and all questions about the deployment of the CMS software, Christoph Wissing was my contact person. With lots of explanations, he guided me through my technical work.

I also wish to thank all former and current colleagues who have contributed to an inspiring office life.

A big thank also goes to all my friends outside of the "physics world". Especially our yearly trips to Denmark gave me the opportunity to forget about my work and relax. We always had a lot of fun, whether we were walking on the beach, swimming in the pool or playing countless rounds of board games.

Finally I want to thank my parents for more than 30 years of support!