Measurement of the Cross Sections of $t\bar{t}$ Production in the Boosted Regime with the CMS Detector at $\sqrt{s} = 8$ TeV

Dissertation

zur Erlangung des Doktorgrades

an der Fakulät für Mathematik, Informatik und Naturwissenschaften

Fachbereich Physik

der Universität Hamburg

vorgelegt von

Ganna Dolinska

aus Kyiv (Ukraine)

Hamburg

2016

Folgende Gutachter empfehlen die Annahme der Dissertation:
PD Dr. Olaf Behnke
Prof. Dr. Peter Schleper

Datum der Disputation: 12. September 2016

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, den

Abstract

In this work the measurement of the $t\bar{t}$ production in "boosted" topologies, i.e. with high transverse momenta of the top quark, $p_T(t) > 400$ GeV, in pp collisions at the LHC is presented. It is the first measurement of this kind in the dileptonic final state.

By the end of 2012 the CMS detector recorded a data sample which contains of the order of 100 thousand $t\bar{t}$ pairs, which further decay dileptonically ($t\bar{t} \to W^+ b\, W^- \bar{b} \to l^+ \nu b\, l^- \bar{\nu} \bar{b}$). The dileptonic final state has a unique signature – two leptons (electrons or/and muons), which are well reconstructed in the detector. The small branching ratio (low statistics) of the dileptonic decay channel of the $t\bar{t}$ system is compensated by an almost background free measurement of the $t\bar{t}$ production. In the presented analysis the kinematics of the $t\bar{t}$ final state is fully reconstructed. For this purpose six kinematic constrains are used to reconstruct the two undetected neutrinos in the $t\bar{t}$ final state.

Two specific techniques are employed to account for the requirements of the boosted top quark analysis. Firstly, the spatial isolation of the leptons from hadronic jets in the events, which is important to reject fake leptons out of jets, was optimized by implementing an additional isolation criterion based on the relative transverse momentum of the electron to the jet. Secondly, to effectively suppress significant migrations from the true low $p_T(t)$ values to the boosted region, a kinematic restriction on the ratio of the invariant mass of the visible objects from the $t\bar{t}$ decay to the reconstructed $t\bar{t}$ mass is applied.

The obtained cross section results in the high $p_T(t)$ region have a good precision, which is comparable to the one from measurements in the semi-leptonic channel, where hadronically decaying top quarks are reconstructed in the boosted regime as a single "fat jet". The measured production cross sections in the dileptonic and in the semi-leptonic decay channels are consistent with each other.

The production dynamics of the boosted top quarks is elucidated in this work in more detail – for this several differential production cross sections as a function of kinematic observables, such as the rapidity of the boosted top quark, are determined. Similar studies are performed in this work in the region of high invariant masses of the $t\bar{t}$ system.

The comparison to Standard Model predictions based on four different QCD Monte Carlo simulation programs evidences no significant discrepancies. However, one can observe some trends in particular distributions, which indicate that the models are far from being perfect.

Kurzfassung

In dieser Arbeit wird eine Messung der $t\bar{t}$ Produktion in pp Kollisionen am LHC mit 'geboosteten' Topologien, d.h. mit hohen Transversalimpulsen der Top Quarks $p_T(t) > 400$ GeV vorgestellt. Es handelt sich dabei um die erste Messung dieser Art im dileptonischen Endzustand. Bis zum Ende von 2012 wurde mit dem CMS Detektor eine Datenmenge aufgezeichnet die ca. 100000 produzierte $t\bar{t}$ Paare enthält die dileptonisch zerfallen ($t\bar{t} \to W^+ b\, W^- \bar{b} \to l^+ \nu b\, l^- \bar{\nu} \bar{b}$). Dieser Endzustand hat eine einzigartige Signatur – zwei Leptonen (Elektronen und/oder Muonen), die im Detektor sehr gut identifiziert und rekonstruiert werden können. Das kleine Verzweigungsverhältnis für Zerfälle in diesen Endzustand wird durch eine grosse Signalreinheit der erhaltenen Datenmenge kompensiert. In der Analyse wird der $t\bar{t}$ Endzustand vollständig kinematisch rekonstruiert. Dabei werden sechs kinematische "Constraints" benutzt um die nicht direkt detektierbaren zwei Neutrinos im Endzustand zu rekonstruieren. Zwei spezielle Techniken wurden angewandt um den Bedürfnissen der Analyse "geboosteter" Top Quarks Rechnung zu tragen. Die räumliche Isolation der Leptonen von hadronischen Jets im Ereignis, die wichtig ist um keine falschen Leptonen aus Jets zu selektieren, wurde durch Hinzunahme eines weiteren Isolationskriteriums, dem relativen Transversalimpuls des Leptons zu den Jets optimiert. Als zweites wurden signifikante Ereignismigrationen von kleinen wahren Werten von $p_T(t)$ zu grossen rekonstruierten hin durch kinematische Schnitte auf das Verhältnis der Masse aller direkt sichtbaren Teilchen aus dem $t\bar{t}$ Zerfall zu der rekonstruierten $t\bar{t}$ Masse effektiv unterdrückt. Die erhaltenen Wirkungsquerschnitte für hohe $p_T(t)$ haben eine gute Präzision, vergleichbar mit Messungen im semileptonischen Zerfallskanal bei denen das hadronisch zerfallende Top Quark "geboostet" ist und als einzelner "Fat-Jet" rekonstruiert wurde. Die gemessenen Wirkungsquerschnitte in dem dileptonischen und dem semileptonischen Zerfallskanal sind miteinander konsistent.

Die Produktionsdynamik für "geboostete" Top Quarks wird in der Arbeit genauer beleuchtet – dafür werden mehrere differenzielle Produktionswirkungsquerschnitte als Funktion von kinematischen Observablen wie z.B. der Rapidität des geboosteten Top Quarks bestimmt. Ähnliche Studien werden in der Arbeit für den Bereich hoher invarianter Massen des $t\bar{t}$ Systems durchgeführt. Ein Vergleich mit Standard-Modell Vorhersagen basierend auf vier verschiedenen QCD Monte Carlo Simulationsprogrammen zeigt keine stark signifikanten Diskrepanzen auf, jedoch sieht man an einigen Trends in bestimmten Verteilungen dass die Beschreibung der Modelle bei weitem nicht perfekt ist.

Contents

1 | Introduction

The modern elementary particle physics states that all the visible matter around us consists of twelve different fermionic elementary particles: six quarks and six leptons. The heaviest out of them is the top quark.

The top quark has unique properties in comparison to the other quarks. In particular, it has an extremely short lifetime ($\sim 10^{-25}$ s), which is less than the typical hadronization time ($\sim 10^{-23}$ s). Thus it decays before hadronizing, providing a possibility to study the properties of a bare quark. Moreover, it is the only weakly decaying particle which decays to a non-virtual W-boson. The top quark has also the largest coupling to the Higgs boson. All the stated properties make the top quark thrilling to investigate.

However, due to its large mass, it took relatively long until one was able to discover the top quark. For instance, quarks were predicted in the middle of 1960s, while the first quarks (u, d and s quarks) were already observed in 1968 [1, 2]. The partner of the s quark, the c quark, was discovered in 1974 in a $c\bar{c}$ bound state, or a J/ψ resonance [3, 4]. The top-partner, the bottom quark, which has the second highest mass among the quarks, was discovered in 1977 [5]. The top quark was discovered only in 1995 at the TEVATRON collider at Fermilab [6, 7].

The LHC collider, which is the most energetic collider ever built, provides a unique possibility to study the top quark properties with unprecedented detail. With the energies and luminosities it operates with it acts like a real top factory. Various top physics precise measurements have already been performed with the data samples collected from the LHC runs in 2011 and 2012.

At the LHC, the top quarks are dominantly produced together with an antitop quark in the gluon-gluon fusion process ($gg \to t\bar{t}$). This work presents a measurement of the production cross section of the $t\bar{t}$ production in the boosted regime (when the top quark is produced with large transverse momentum). The boosted top quark production gives not only the possibility to test the Quantum Chromodynamics (QCD) predictions and probe the Standard Model, but it may also point to new physics, as the top-quark pair may be a product of a heavy resonance decay.

One of the main features of the boosted top production is that the decay products of the top quark with a high momentum travel close to each other and sometimes is is hard to distinguish between them. Usually, in the hadronic final state of the top decay, the decay products of the boosted top quark are not separated and are reconstructed as a single "fat jet". In this work the $t\bar{t}$ pair is reconstructed in the dileptonic final states (see Fig. 1.1). Using this final state all the products of the top quark decay are reconstructed separately providing the full information about the final state $t\bar{t}$. This provides a possibility to cross check the measurement strategies with "fat jet" techniques. To effectively select the $t\bar{t}$ events in boosted regime, the specific spatial isolation of leptons from jets is applied in

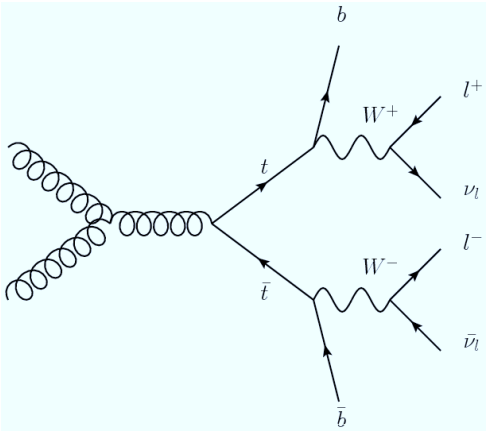

Figure 1.1: The diagram of the production of the $t\bar{t}$ pair in the gluon-gluon fusion and its further dileptonic decay. The plot is taken from [8].

this analysis.

From the studies on simulated MC samples it is observed that a significant number of events, where the transverse momenta of neutrinos are small, is reconstructed with too high neutrino momenta. This also results in artificially high reconstructed momenta of top quarks. To reject these events, additional cleaning conditions are applied.

To study the production dynamics of the $t\bar{t}$ production in boosted regime the cross sections in the boosted region (with transverse momentum $p_T(t) > 400$ GeV) are measured in this thesis as a function of the $p_T(t)$ minimum thresholds, absolute rapidity of the top quark, $|y(t)|$, absolute rapidity of the $t\bar{t}$ system, $|y(t\bar{t})|$, and pseudorapidity separation between top and antitop from the $t\bar{t}$ pair. The selection and the cleaning conditions applied in this analysis also improve the reconstruction of events with high $M(t\bar{t})$. That is why the measurements of the cross sections as a function of $|y(t)|$, $|y(t\bar{t})|$ and pseudorapidity separation between top and antitop are additionally performed in the high $M(t\bar{t})$ region, in which some discrepancies in the description of the measurements by QCD predictions were observed previously [9].

The thesis is structured as follows:

An overview of the modern understanding of elementary particle physics is given in Chapter 2. A brief description of the Standard Model of particle physics in general and top quark physics in particular is provided.

The experimental facilities, which provided the data for the measurements performed in this thesis, are described in Chapter 3. The main features of the LHC collider as well as the essential parts of the CMS detector are elucidated.

In any high energy physics experiment there is a need to correct measured data for detector effects. For this Monte Carlo simulations (MC) are used. Chapter 4 provides an overview of the MC programs which were used in this thesis for the simulation of the experiment.

In Chapter 5 the building of the physical objects from the detector signals is described which are used in the further analysis. The further selection of the events based on the characteristics of the reconstructed objects is elucidated in Chapter 6. The goal of the event selection is to pick the $t\bar{t}$ signal events (production of the $t\bar{t}$ pair with a decay chain which ends up in a final state with two leptons, two jets and two neutrinos).

The kinematic reconstruction, which is used to obtain the full information on the $t\bar{t}$ dileptonic final state is elucidated in Chapter 7.

The studies of the specific lepton isolation and of the cleaning conditions, which are applied to improve the selection and reconstruction procedures for the boosted topologies, are presented in Chapter 8.

Chapter 9 contains the description of the procedure of the $t\bar{t}$ cross sections determination, including the corrections of the experimental data for the detector resolution, migrations and acceptance effects. The different sources of systematic uncertainties, which are relevant for this analysis, and the assessment of the resulting uncertainties on the measured cross sections are discussed in Chapter 10.

The final results including consistency checks and comparisons to other measurements and theoretical models are presented in Chapter 11.

A summary of the whole analysis and of the main conclusions are given in Chapter 12.

2 | Modern Elementary Particle Physics

Mankind was always trying to find out, what are the building bricks of the world around us. This question is not driven by pure curiosity, but also by an urge to understand the nature of different phenomena and interactions. Moreover, a deep knowledge of the structure of matter may allow us to answer the question about the origin and long term evolution of our Universe.

Already ancient philosophers made great efforts to elucidate the question of the structure of matter. Along with simplistic and incorrect models, like the one proposed by Anaximenes of Miletus, who claimed that everything consists of water, earth, fire and air, there were quite sophisticated guesses, like the prediction of the basic elementary pieces of matter *atomos*, by Democritus.

In general, the knowledge about the structure pieces of matter was developing as shown in Fig. 2.1. The understanding of a minimum piece of everything went down from molecules and atoms to electrons, nuclei and quarks.

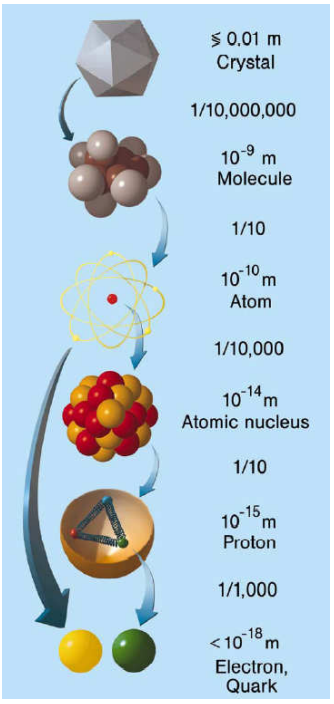

Figure 2.1: Building blocks of matter on different scales. The plot is taken from [10].

The **S**tandard **M**odel (SM) is the theory which successfully describes the particle physics in our modern understanding. In the Standard Model the smallest indivisible

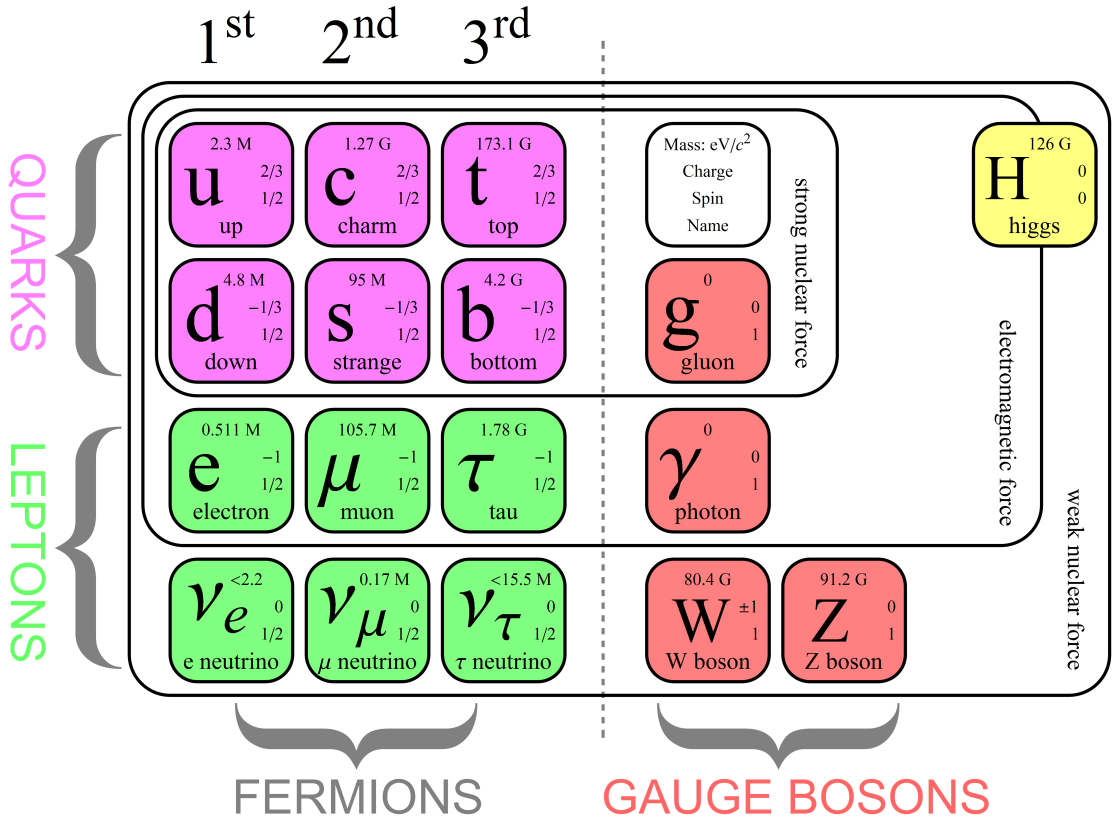

Figure 2.2: The Standard Model set of elementary particles and gauge bosons with their masses, electric charges and spins. The plot is taken from [11].

particles are leptons and quarks (see Fig. 2.2). The interaction between them is mediated by the field bosons.

After the discovery of a Higgs boson [12, 13], the predicted set of particles for the Standard Model was completed. However, one can not call the SM a complete theory, as there are some experimental facts which don't fit into this model. For example:

- the description of dark matter and dark energy is not included in the Standard Model;
- there is no answer on the question, why there are exactly three generations of particles (as shown in Fig. 2.2);
- the asymmetry between matter and antimatter, observed in the Universe, is not explained;
- the gravity is not included in the Standard Model.

There will be no detailed discussion of the open questions of the Standard Model in this chapter but rather a brief overview of the main principles of the SM. Additionally, the $t\bar{t}$ pair production in pp collisions and the subsequent decay will be elucidated.

2.1 Remark on the Units

As the interactions and properties of the elementary particles are related to extremely small scale (small sizes, relatively small energies and masses), the standard units of the S.I. system, which are defined on larger scales, are not convenient to describe them. For this reason another set of variables is usually used in particle physics. It is displayed in the table below:

Quantity	Particle physics units	S.I. units
Energy	1 eV (electron-Volts)	1.602×10^{-19} J
Area	1 barn	10^{-28} m^2

Taking to account the Einstein energy-momentum relation $E^2 = p^2c^2 + m^2c^4$, where E is the particle energy, p is the particle momentum, m is the mass of the particle and c is the speed of light, one can conclude that the masses in the elementary particle physics are expressed in eV/c^2 and the momenta – in eV/c.

For a more comfortable description of the equations and units in particle physics, one usually assumes the *natural units*, which means that the main constants are assumed to equal unity:

$$\hbar = c = 1. \tag{2.1}$$

From here on in this thesis the natural units are assumed[1].

2.2 Constituents of Matter

As mentioned before, according to the Standard Model, the matter consists of elementary fermions[2] – *leptons* and *quarks*. They are grouped into 3 generations (see Fig. 2.2). The particles from different generations apparently differ only by their masses: the higher the generation – the heavier the masses are. Other then that, there are no difference in the quantum numbers or properties (which are not related to the mass).

All the fermions have their "antipartners" with the same mass, but with inverted charges. These are the *antiparticles*. Their existence can be explained by the Dirac equation [14], which describes all the fermions. This equation has different solutions with positive and negative energy states – corresponding to particles and antiparticles.

There are three charged and three neutral leptons (not counting their antiparticles): electron (e^-), muon (μ^-), tauon (τ^-) and corresponding neutrinos – ν_e, ν_μ and ν_τ. Leptons carry a special charge called *lepton number*, which is +1 for the leptons and -1 for the antileptons. The total lepton number is conserved in the interactions. The *lepton flavour* (e, μ or τ) is also conserved in all the processes except in the neutrino oscillations [15].

There are six types of quarks, or six different *quark flavours*: up (u) and down (d) quarks are in the first generation, charm (c) and strange (s) quarks are in the second generation and top (t) and bottom (b) quarks are in the third generation of the elementary particles. Quarks have non-integer charge (in units of the electron charge e), either

[1] This also means that the masses, energies and momenta will all be expressed in eV.

[2] A *fermion* is any particle characterized by Fermi-Dirac statistics and which has half-integer spin.

$+2/3e$ or $-1/3e$. Another special quark property is the presence of the *color charge* – the quantum number which is responsible for the ability to interact strongly (will be discussed in Sec.2.3.5). There are three colors (red (r), green (g) and blue (b)), which have corresponding anticolors. The sum of all three colors (anticolors), or a sum of the color and corresponding anticolor result in a final state with no color.

Quarks can form colorless bound states. The most simple and the only states observed experimentally and unambiguously so far are the *mesons* and the *baryons*. Mesons are composed of pairs of quark and antiquark and have integer spins, while the baryons are formed of three quarks and have half-integer spins. By now, the only known stable baryon is the proton and all the mesons are unstable.

2.3 Interactions

The interactions between any objects in the Universe are governed by four basic forces: gravity, electromagnetism, weak and strong interaction. For comparison, the relative strength of all four forces acting between two elementary particles for the typical energies of ~ 1 GeV is presented in the following table:

Force	Strength
Strong	1
Electromagnetic	10^{-2}
Weak	10^{-15}
Gravity	10^{-41}

Talking about the interaction between elementary particles, the masses of which are very low, the gravity can be neglected. As gravity is also not included to the Standard Model, it will not be discussed in this chapter.

The strength of the different interactions is described by *coupling constants*, which are dimensionless constants depending on the actual interaction potential strength.

2.3.1 Interaction Carriers

In the Standard Model all the interactions are mediated by the gauge bosons[3]. There are massles gauge bosons – the *gluon* (g) and the *photon* (γ). The gluon is responsible for the strong interaction, it has two colors (a color and an anticolor) and zero electrical charge. The photon is the mediator of the electromagnetic interaction, it has neutral electrical charge.

The massive gauge bosons ($W^{\pm}$ and Z^0) are responsible for the weak interaction.

2.3.2 Electromagnetic Interaction

All the particles and objects which have a non-zero (total) electric charge undergo the electromagnetic interaction. This interaction is transmitted through the photon exchange and the theory, which describes electromagnetic interaction, is **Q**uantum **E**lectro**d**ynamics

[3] A *boson* is any particle, which has an integer spin and which is characterized by the Bose-Einstein statistics

(QED). The coupling of the QED gauge boson (γ) to the charged fermions can be described by a simple interaction vertex (see Fig. 2.3). The more complex processes are constructed of such vertices.

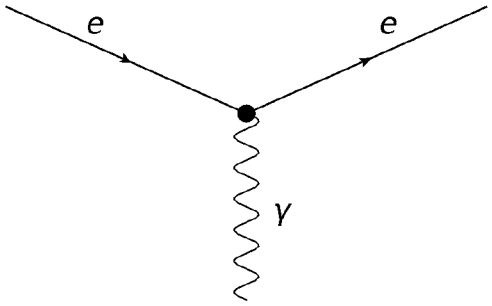

Figure 2.3: QED interaction vertex.

The coupling constant for the electromagnetic interaction is expressed as follows (in the natural units):

$$\alpha = \frac{e^2}{4\pi}, \tag{2.2}$$

where e is the electric charge of the electron. The electromagnetic coupling constant $\alpha \sim \frac{1}{137}$, which is much smaller than unity. That means that for the estimation of interaction probabilities one can use the α constant for perturbative expansions. One can calculate the expansion up to a certain order of α. Thus, the processes may be given with **l**eading **o**rder (LO), **n**ext-to-**l**eading **o**rder (NLO), **n**ext-to-**n**ext-to-**l**eading **o**reder (NNLO), etc., accuracy.

The general properties of the electromagnetic interaction are that it results in attraction (repulsion) between the particles with the electric charge of the different (same) sign. The range of the electromagnetic interaction is unlimited and it (the force) gets weaker proportionally to the distance squared between the interacting objects. The additional effective decreasing of the electromagnetic force depending on the scale can be attributed to the *screening effect*. The main idea is that the vacuum around a charged particle polarizes and causes the attraction of the virtual dipoles around the real charge. The virtual charges screen the real charge or reduce it. The effect of screening by virtual particles is very weak on the short distances to the real charge, but the further the spectator moves away from the real charge, the stronger the screening becomes.

2.3.3 Weak Interaction

The weak interaction is called that way as the strength of it (see Sec. 2.3) is much smaller than for the electromagnetic and strong interactions. The charge responsible for the weak interaction is called *weak isospin*. The weak interaction is mediated by three massive gauge bosons: electrically neutral Z^0 boson and electrically charged $W^{\pm}$ bosons. Related to that, there are two kinds of weak interactions – via neutral current, or via charged current exchange. The basic vertices of the weak interaction are presented in Fig. 2.4.

The neutral current, as in the case of strong and electromagnetic interactions, doesn't produce lepton or quark flavour exchange, while the charge current does both. The

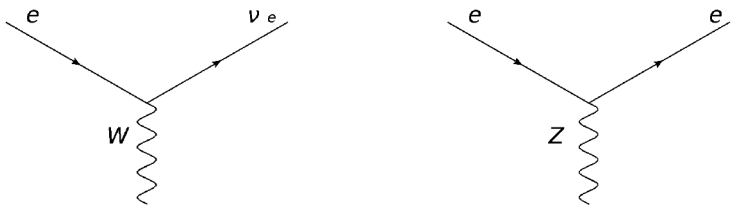

Figure 2.4: Weak interaction vertices: charged current (left) and neutral current (right).

quark flavour exchange may happen only within one generation, according to the quark weak eigenstates. However, the weak eigenstates differ from the mass eigenstates. The transformation between them is given the following way:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.3}$$

Here d, s and b are the weak eigenstates and d', s' and b' are the mass eigenstates. The matrix V is the **C**abibbo-**K**obayashi-**M**askawa (CKM) matrix [16]. It describes the transition probabilities between different quark states. All the values of the matrix were experimentally determined [17]:

$$V = \begin{pmatrix} 0.974 & 0.225 & 0.003 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix}. \tag{2.4}$$

The diagonal elements of the CKM matrix are close to unity, which means that the transition of the quark falvour inside one generation is most likely to happen. The off-diagonal elements are much lower. The lowest is the probability of flavour exchange between the first and the third generations.

2.3.4 Electroweak Unification

The main goal of any physical studies is to provide the most universal laws, which would describe as many aspects of the Universe as possible. For example, electricity and magnetism are already unified to a common theory of electomagnetism. The unification of electromagnetic and weak interactions was achieved by Glashow, Salam and Weinberg providing a theory of electroweak interaction [18].

The weak interaction through charged current is associated with the SU(2) gauge symmetry. The SU(2) local gauge invariance is satisfied by introducing three fields, which correspond to the three gauge bosons: W^1, W^2 and W^3. All these fields couple to fermions with a coupling constant g. The fields couple to left-handed chiral[4] particles and right-handed chiral antiparticles. The W^1, W^2 couple to charged fermions, thus the $W^\pm$ can be expressed as their linear combination:

$$W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2). \tag{2.5}$$

[4]The *chirality* defines the right-handed or left-handed state of the particles. The chirality is changed in a parity transformation.

The W^3 couples to neutrinos, thus it is tempting to express the neutral Z boson in its terms. However, in reality the Z boson doesn't couple only to the left-handed chiral particles, while the W^3 couples to the left-handed chiral states only. The Z boson is expressed through a linear combination of W^3 and B, the gauge boson of the electromagnetic U(1) symmetry group (the corresponding coupling constant is g'). The photon is also expressed in a similar way:

$$\gamma = B\cos\theta_W + W^3\sin\theta_W, \tag{2.6}$$

$$Z = -B\sin\theta_W + W^3\cos\theta_W. \tag{2.7}$$

Here θ_W is the *weak mixing angle*. It is measured experimentally and results around 28.7^o [17]. This angle can be expressed in terms of the coupling constants g and g':

$$\tan\theta_W = \frac{g'}{g}. \tag{2.8}$$

The charge, responsible for the electroweak interaction, is called *weak hypercharge* Y and it is expressed the following way:

$$Y = 2Q - I_3, \tag{2.9}$$

where Q is the electric charge in units of the electron charge, and I_3 denotes the weak isospin.

The electroweak symmetry is possible under the assumption that all the gauge bosons are massless. However, the $W^{\pm}$ and Z bosons have a non-zero mass [17]. This phenomenon is explained by the *electroweak symmetry breaking*, in frames of which the gauge bosons obtain their mass via *Higgs mechanism* [19], which introduces an additional field - the Higgs field. The particles couple to this field and obtain their masses. The Higgs boson is a quantum of the Higgs field.

2.3.5 Strong Interaction

All the particles which have non-zero color are able to interact strongly via gluon exchange. The theory, which describes the strong interaction, is called **Q**uantum **C**hromo**d**ynamics (QCD). Unlike the electromagnetic interaction, the mediator of the strong interaction (gluon) has a color charge, which means that not only quarks, but also other gluons can couple to gluons. The other consequence of it is that there is color exchange present in the strong interaction. The basic vertices of the strong interaction are presented in Fig. 2.3.

There are many experimental evidences of the existence of quarks and gluons [20], although they were never observed in the free state. The reason for that is one of the unique properties of the strong interactions: *color confinement*. The main idea of this property is that the colored objects are always confined to states with zero color. The states with non-zero color can't propagate as free particles. There is still no accurate analytic description of the color confinement, however some models successfully describe it. One approach is to assume that the field lines of the strong interaction are squeezed into a narrow tube. When one attempts to separate two quarks the field potential is growing proportionally to the volume of the tube, $V(\vec{r}) \sim \kappa\vec{r}$. Here $\vec{r}$ is a distance between the

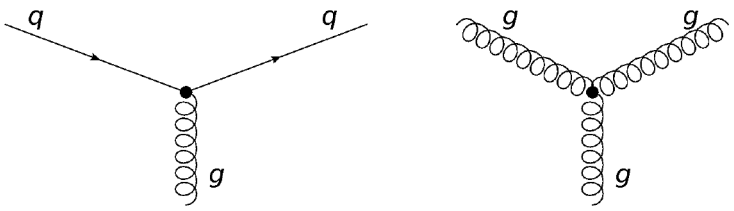

Figure 2.5: QCD interaction vertices: coupling of quarks to gluons (left) and gluon self-coupling (right).

two quarks and κ is an experimentally defined constant ($\kappa \sim 1$ GeV/fm). As soon as the potential gets large enough, another pair of quarks is produced. This process is illustrated in Fig. 2.6.

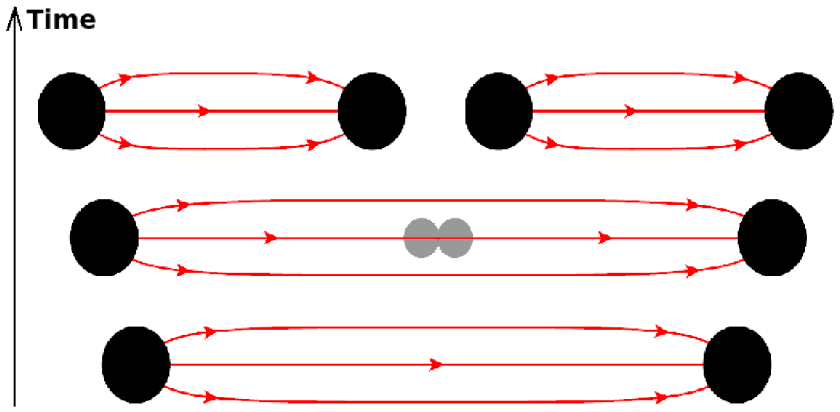

Figure 2.6: Illustration of the color confinement.

In the big particle physics experiments the strong interactions is artificially initiated in the high energetic collisions. In these interactions due to small interaction distances and high collision energies free quarks may be produced or gluons may be irradiated. As these quarks can't propagate as separate particles, the confinement process starts: quarks pick the partners for the bound states out of the vacuum producing mesons and baryons, which flow in the same direction. These groups of particles are called *hadronic jets*.

The coupling constant of the strong interaction, α_s, depends on the energy scale of the interaction. This property is called *running* α_s. The constant α_s has been measured experimentally at different energy scales confirming the running (see Fig. 2.7). The larger the energy scale of the process is, the lower the α_s gets. This running can be explained by the *color anti-screening effect*. This is opposite to the screening effect, observed in QED. The screening of the color charge around a quark can be described similarly to the screening of the electric charge. However, the carriers of the strong interactions (gluons)

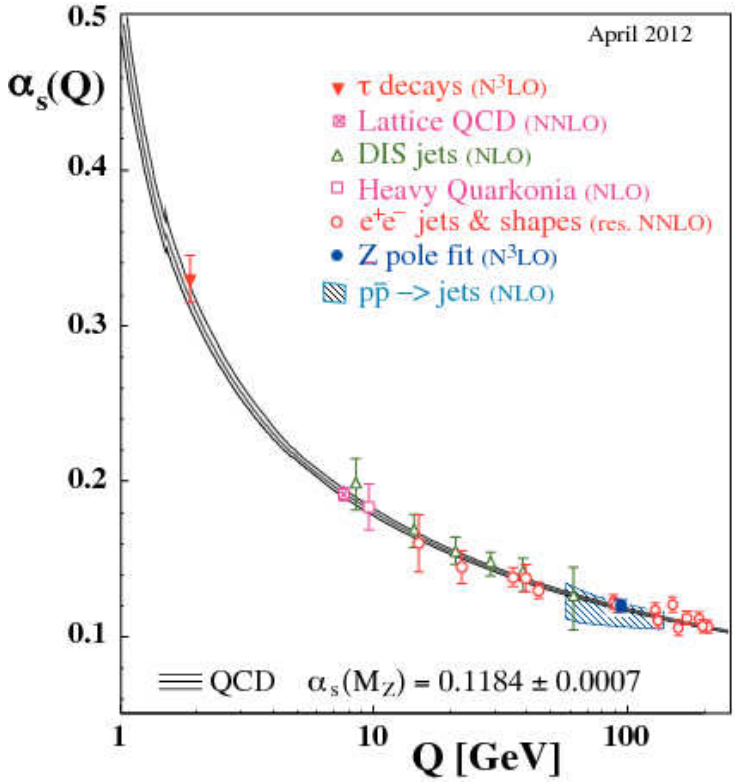

Figure 2.7: Illustration of the α_s running. Different points represent the result of different measurements in different processes. The band represents theoretical calculations. The plot is taken from [21].

also have a color and an anticolor and for them the effect is the opposite, they anti-screen the gluon color. The gluon color anti-screening is strongly dominating over the quark color screening.

Processes with low energy scale have large α_s, thus they can't be calculated in perturbation theory. The processes with high energy scale, on the opposite, have lower values of α_s and can be theoretically calculated using perturbation theory, which corresponds to the cross section calculations via the expansion in a series of α_s: $\sigma_{pert} = \sum_{k=min}^{\infty} \alpha_s^k \cdot C_k$, where *min* is the minimum possible power of α_s for a corresponding process and C_k is the coefficient which represents the hard scattering cross section. The higher the power of α_s is, the more loops or additional radiations are described in the corresponding term of the expansion. The processes contributing to the terms can be graphically represented by Feynman diagrams such as those shown in Fig. 2.11-2.13.

The property, that on the very high energy scales the α_s approaches zero is called *asymptotic freedom*. In this state quarks are quasi-free.

The strong force is a force of attraction. It's range is limited due to the confinement.

2.4 Proton Structure and Factorization

Tests of the main principles of the Standard Model and searches beyond the SM are performed in modern particle physics experiments. The collider experiments play a very

important role in this field. The largest collider experiments running nowadays are situated at the **L**arge **H**adron **C**ollider (LHC) (which will be discussed in Chapter 3). The LHC provides collisions of protons in which different interaction and production processes are initiated.

The production cross section of any hard process $pp \to A$ at the LHC can be described by the factorization theorem [22]. This theorem represents the cross section of any strong hard process in pp collision as a convolution of the partonic cross sections, which can be calculated in perturbative QCD (pQCD), and a non-perturbative phenomenological part. The artificial limit between two parts is chosen arbitrary and is called the *factorization scale* μ_F. Finally, the cross section of the $pp \to t\bar{t}$ can be expressed the following way [23]:

$$\sigma_{pp\to A} = \sum_{i,j=all\ partons} \int dx_i dx_j f_i(x_i, \mu_F^2) f_j(x_j, \mu_F^2) \cdot \sigma_{ij\to A}(s, x_i, x_j, m_t, \mu_F, \mu_R, \alpha_s). \quad (2.10)$$

Here $x_{i,j}$ are the longitudinal fractions of the proton momentum carried by a parton i or j, $\sigma_{ij\to AB}$ is a partonic cross section of the process $ij \to A$ (this part is calculated in pQCD), s is the center-of-mass energy of the pp system, μ_R denotes the renormalization scale and $f_{i,j}$ are the **P**arton **D**istribution **F**unctions (PDFs) of the partons i or j (the phenomenological part of the convolution).

The PDF is the probability to find a parton i with a given longitudinal proton momentum fraction x in the proton. The PDFs can't be fully determined in perturbative QCD calculations. They are first parametrized at some starting scale Q_0 and further extrapolated to the other Q^2 scales with the DGLAP evolution [24–26]. Following this Ansatz, the PDFs are fitted to different experimental data at different Q^2 scales [27, 28].

Different groups, like CTEQ [29] or MMHT [30], provide sets of PDFs, which are based on the experimental data fits. Most of the information on the modern PDFs are obtained from the ep HERA **D**eep **I**nelastic **S**cattering (DIS) data, where the proton structure was probed by gauge bosons (γ, Z, W) emitted from the electron beam. An example of the PDFs based on HERA data is shown in Fig. 2.8.

2.5 Top Quarks

The top quark is the heaviest known elementary particle [32]. It has a very short lifetime (10^{-25} s), which is smaller than the typical time for hadronization. Thus, the top quark decays before hadronizing. It means that there is no bound state (meson or baryon), which has a top quark as a constituent. These properties are unique among the other quarks, which makes the top quark a thrilling object for studies.

All the results of the top quark production, decay and coupling, which were obtained by now are pointing to the fact that it is a Standard Model particle and no traces of physics beyond Standard Model have been observed in its properties. However, the searches of possible evidences of physics beyond Standard Model are ongoing. In many of those searches some new resonances are directly looked for through a decay process involving top quarks. In particular, the resonances decaying into $t\bar{t}$ are searched for. In case the new resonances are too heavy and can't be produced under the conditions of the LHC collisions, another kind of searches is performed. They are focused on finding deviations in

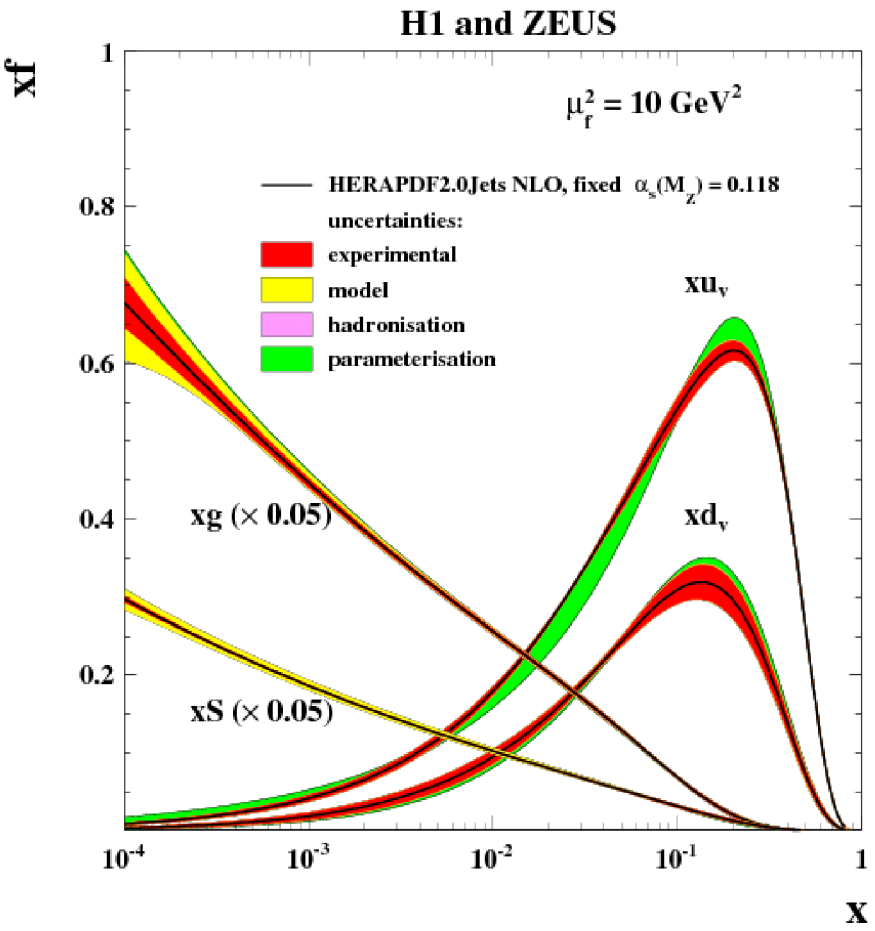

Figure 2.8: The PDFs from HERAPDF2.0, which were obtained using combined HERA data. Here, $\mathbf{u_v}$, $\mathbf{d_v}$, $\mathbf{S}$ and $\mathbf{g}$ correspond to densities of valence u-quark, valence d-quark, sea quarks and gluons respectively. The plot is taken from [31].

couplings of the top quark to the other particles from the Standard Model. The searches of the physics beyond Standard Model are also ongoing in a sector of top quark decay: the limits on the rare top quark decays (like top quark decaying to a light charged Higgs boson, which is not predicted in the Standard Model) are set.

So far the basic properties of the top quark, like mass, couplings, production cross sections, etc. have been measured at the TEVATRON and at the LHC. No other machine provided enough energy and data at the same time to observe the top quark.

In this work the production of the top quarks at the LHC is analyzed. The LHC itself with its high design collision energy (14 TeV for pp collisions, see Sec. 3.1) is a laboratory where one can study many Standard Model processes with a high production rate (see Fig. 2.9). Moreover, the production cross section of the top quark at the LHC design collision energy is two orders of magnitude higher than for $p\bar{p}$ collisions at the TEVATRON, where the top quark was discovered. This means that the LHC is a real top quark factory.

In this section a short overview of the top quark production and decay will be given.

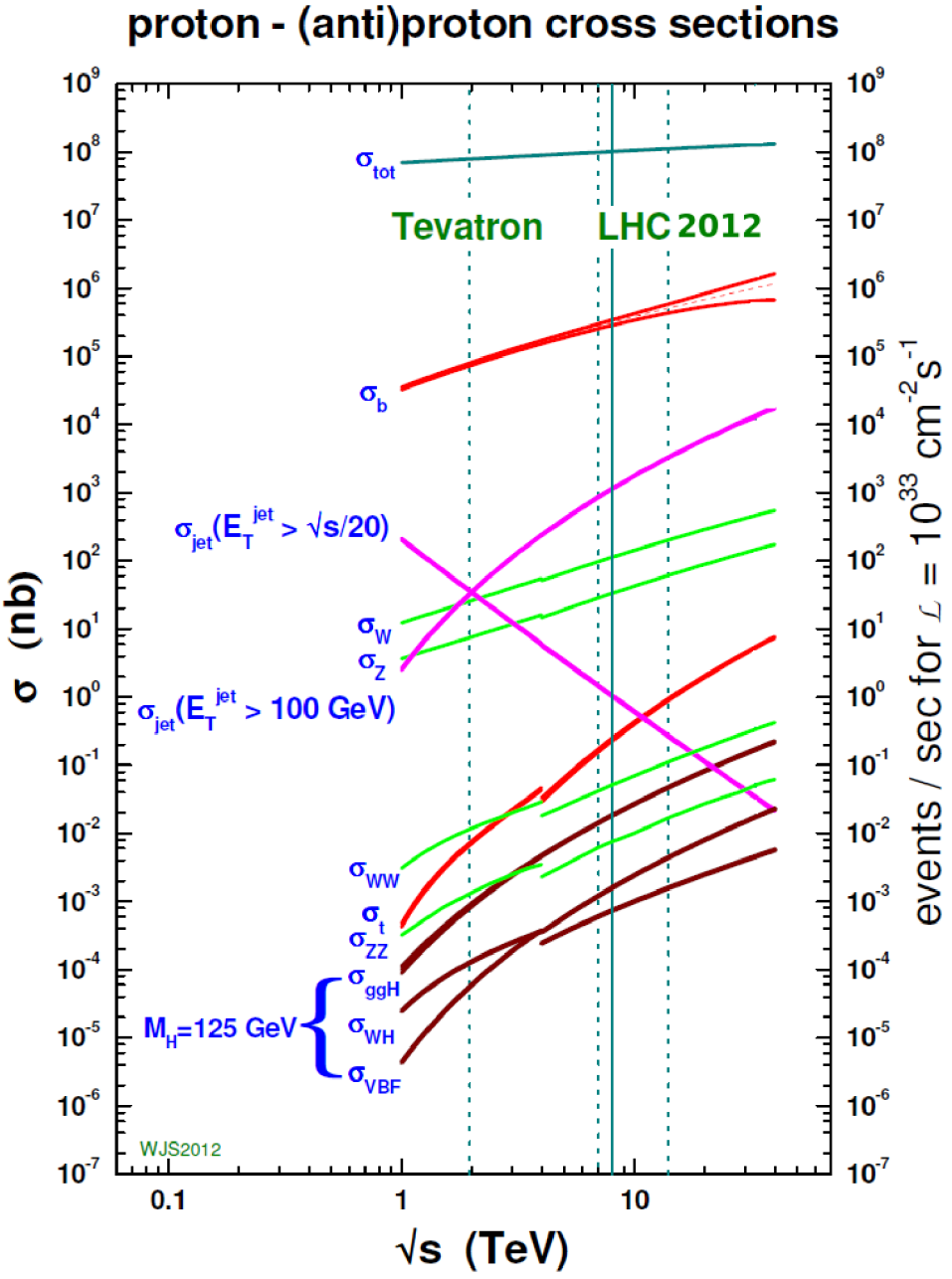

Figure 2.9: Standard Model cross sections as functions of collider energy. The plot is modified from [33].

2.5.1 Top Quark Production

The top quarks in the pp collisions can be produced either as single particles (see Fig. 2.10) or in pairs (see Fig. 2.11, Fig. 2.12 and Fig. 2.13). Their production cross sections can be calculated using the factorization theorem (see eq. 2.10).

Single top quarks are produced via the electroweak interaction, almost exclusively through the W_{tb} vertex (as the $|V_{tb}|$ element of the CKM matrix is much larger than $|V_{td}|$ and $|V_{ts}|$). Thus, the single top production is directly sensitive to V_{tb}. This process is very interesting to probe the Standard Model and might have sensitivity to physics beyond the SM.

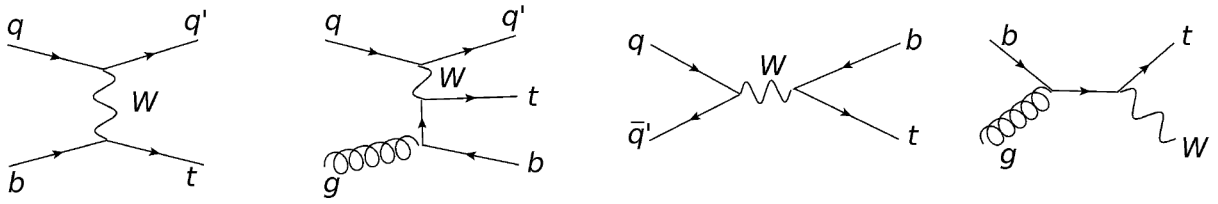

Figure 2.10: Feynman diagrams for single top production.

Top quark pairs from pp collisions are dominantly produced via the strong interaction.

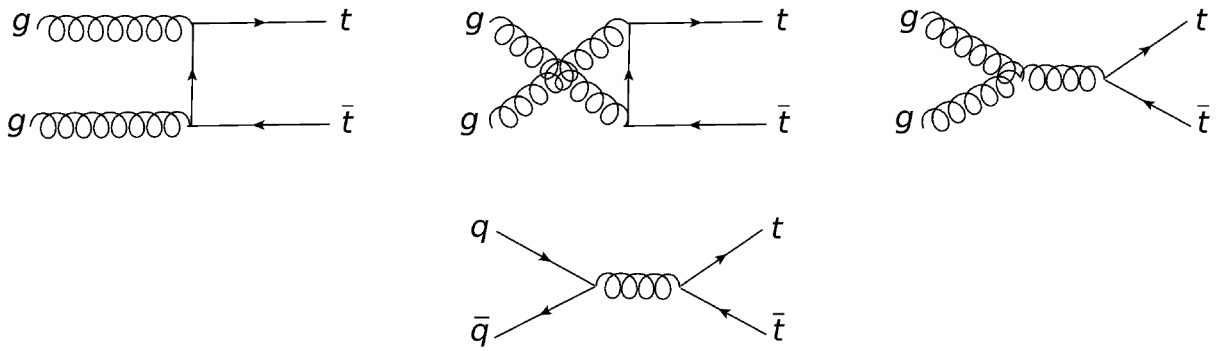

Figure 2.11: Leading order Feynman diagrams for the top pair production.

Figure 2.12: Examples of NLO Feynman diagrams for top pair production.

At the LHC, the $t\bar{t}$ pairs are dominantly produced in strong interactions: via gluon-gluon fusion (first three diagrams in Fig. 2.11), which is dominating ($\sim 80\%$ for $\sqrt{s} = 8$ TeV pp collisions), or via quark-antiquark annihilation ($\sim 20\%$, last diagram in Fig. 2.11).

The measurement of the $t\bar{t}$ production cross sections is a very important test of the Standard Model. The total $t\bar{t}$ production cross sections are currently calculated up to

NNLO accuracy in QCD (Fig. 2.13 shows examples of NNLO diagrams), thus a comparison with experimental data provides a precision test of perturbative QCD. The differential $t\bar{t}$ production cross sections have sensitivity to high-x gluons in the proton and the density of high-x gluons (with $x \gtrsim 0.5$) is not well known. Thus, the top cross sections might be used to constrain PDFs. The studies presented in this work are particularly focused on the production of the top quarks with high momenta. These top quarks may occur from the decay of resonances with very high masses, thus may be sensitive to new physics.

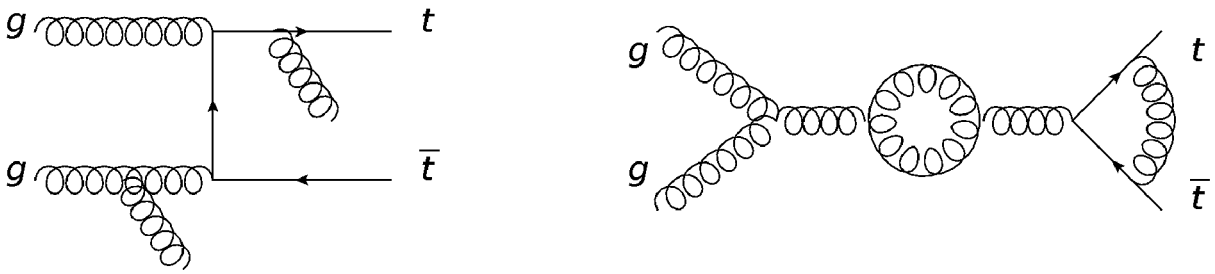

Figure 2.13: Examples of NNLO Feynman diagrams for top pair production.

Different measurements of the integrated $t\bar{t}$ production cross sections are presented in Fig. 2.14. The differential [34, 35] and double differential [9] $t\bar{t}$ production cross sections were measured with the LHC data.

As the $t\bar{t}$ pair production is the object of the studies described in this work, the further discussion will be focused on it.

2.5.2 Top Pair Decay

As was discussed in the previous section, the CKM element $|V_{tb}|$ is close to unity, thus the top quark decays almost exclusively to a $W^{\pm}$ boson and a b-quark. The $W^{\pm}$ boson is a resonance with a short lifetime[5], which can either decay to a pair of quark and antiquark or to a lepton and a corresponding anti-neutrino. The decay channel of the top-pair is thus determined by the decay mode of the $W^{\pm}$ bosons.

The $t\bar{t}$ pair has the following decay modes:

- *Fully hadronic* decay mode. Both W bosons decay to a pair of quark and antiquark.
- *Semi-leptonic* decay mode. One of the Ws decays hadronically (to a quark and antiquark) and the other one decays to a lepton and an anti-neutrino.
- *Dileptonic* decay mode. One W boson decays to a lepton and an anti-neutrino and the other to an antilepton and neutrino.

The $t\bar{t}$ decay modes with the probabilities of them to occur (*branching ratios*) are summarized in Fig. 2.15. As one can see from this figure, the fully hadronic decay channel has the highest branching ratio. However, it implies the presence of six jets in the final state. This leads to several challenges: first, it is hard co correctly assign the six jets correctly to the two top quarks and second, it is easy to mix up the jets from the $t\bar{t}$

[5] The lifetime of the $W^{\pm}$ boson is $\sim 3 \cdot 10^{-25}$ s.

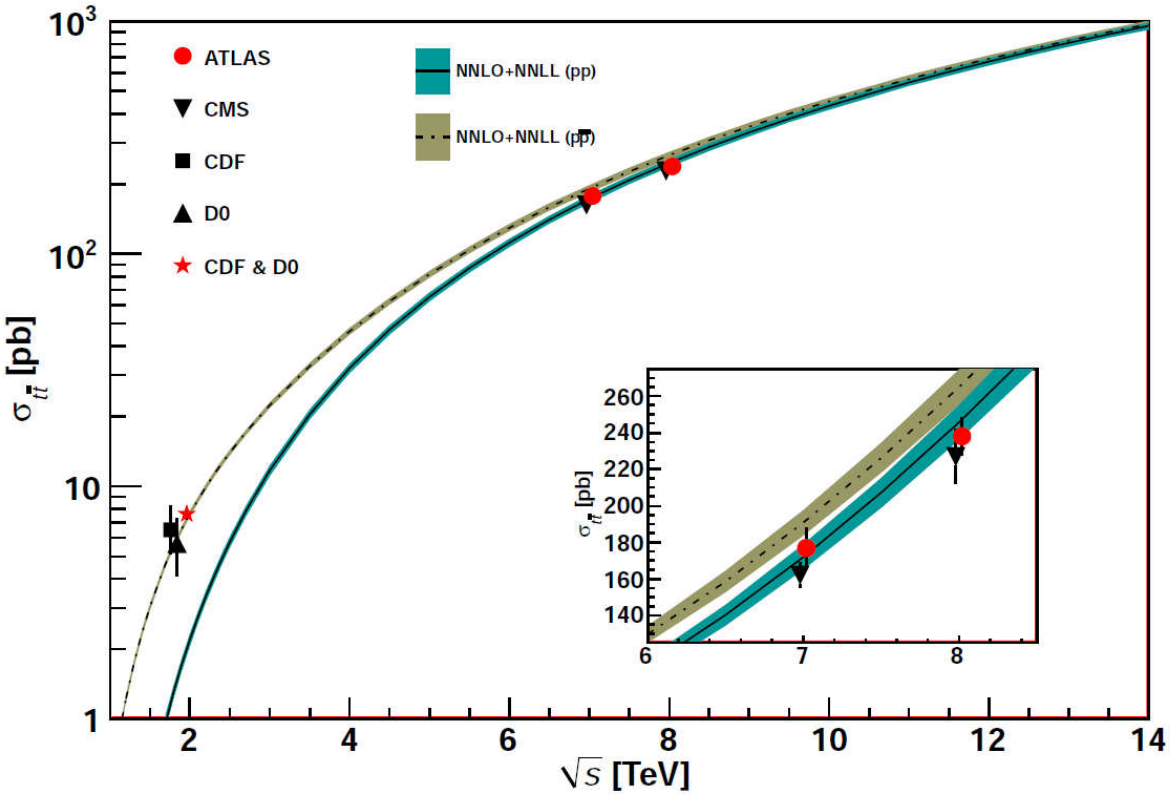

Figure 2.14: Measured $t\bar{t}$ production cross sections at the LHC and TEVATRON experiments compared to the NNLO predictions. The results from TEVATRON are taken from [36–40] and the results from the LHC are taken from [41–45]. The plot is taken from [32].

decay with the other jets in the event since in the environment of the LHC collisions, the number of additionally produced hadronic jets in the event can be large. In this sense the semileptonic and dileptonic decay channels are easier to identify. However, they have lower branching ratios.

In this work the dileptonic decay of the $t\bar{t}$ system is studied. It is the channel with the lowest branching ratio, but the presence of two leptons makes it relatively easy to identify. Depending on into which leptons the Ws from top quarks decay, there are six possible dileptonic final states: ee, $e\mu$, $\mu\mu$, $e\tau$, $\mu\tau$ and $\tau\tau$[6]. The τ lepton can further decay leptonically in the following decay channels: $\tau \to l\nu_l\nu_\tau$, or hadronically into ν_τ + hadrons. Here l is either an electron e or a muon μ. Only the ee, $e\mu$ and $\mu\mu$ final states are treated in this analysis as signal processes.

[6]Here and further in the text the signs of the electric charges of the leptons from the top decay are skipped when appearing in the name of the $t\bar{t}$ decay channel

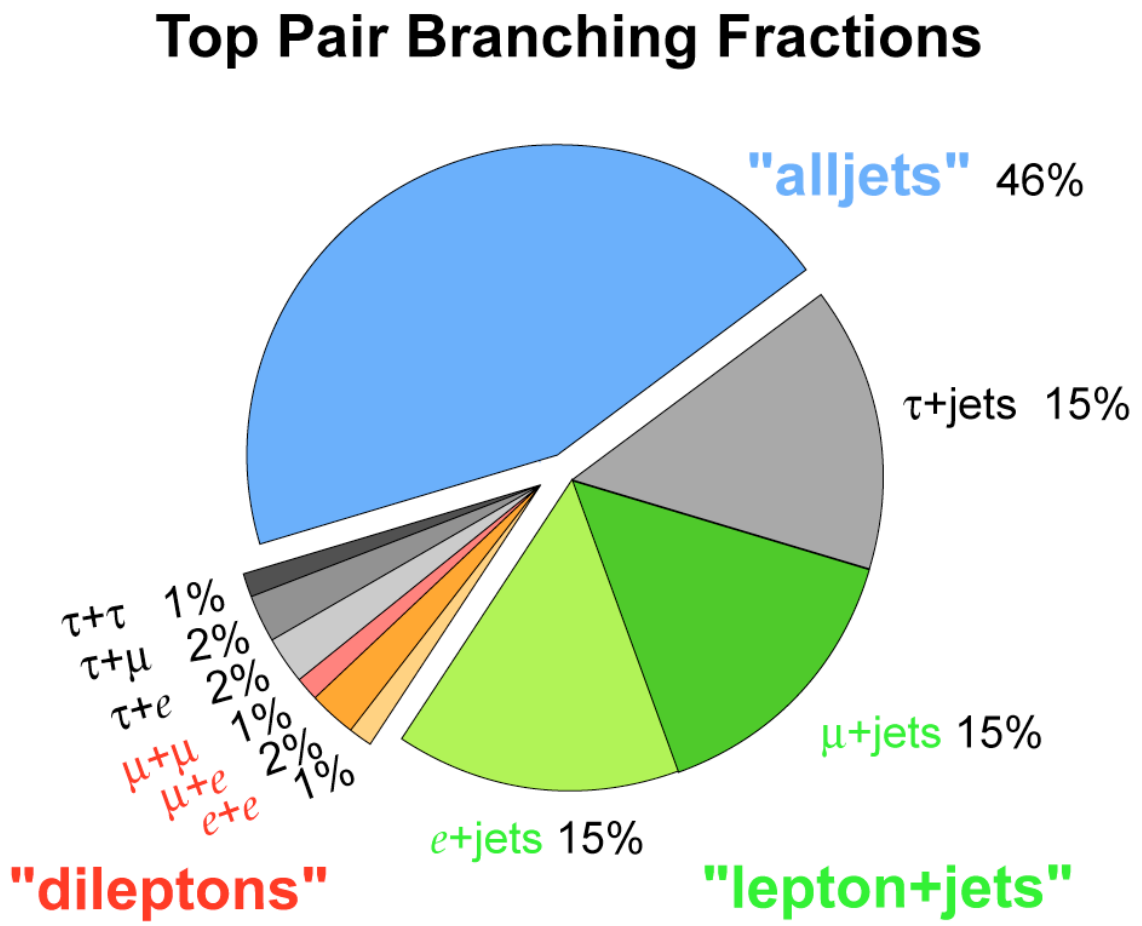

Figure 2.15: $t\bar{t}$ decay modes with the corresponding branching ratios. The plot is taken from [46].

3 | CMS Experiment at the LHC

To artificially produce elementary particles under controlled conditions, special experimental facilities have to be set up – the particle accelerators and colliders. The particles, like electrons, protons, or heavy nuclei are accelerated first and then brought to collisions (either one on another, or on a fixed target). After the collision, the reaction products are registered and memorized by the particle detectors.

The analysis presented in this thesis is based on the data collected by the **C**ompact **M**uon **S**olenoid (CMS) experiment at the **L**arge **H**adron **C**ollider (LHC). This chapter gives an overview of the LHC and provides a brief description of the CMS parts.

3.1 The LHC

The LHC [47] is the worlds biggest and most powerful synchrotron accelerator facility ever built. The LHC is a part of the CERN [48] (European Organization for Nuclear Research) accelerator complex. It is exploited to investigate physics on the TeV scale. The LHC is designed to collide protons (pp collisions) or heavy lead nuclei. The results presented in this work are obtained from the proton-proton collisions, thus this overview will focus on describing this type of collisions.

The design centre-of-mass energy of the pp collisions provided by the LHC is 14 TeV (in 2012 the energy of collisions was 8 TeV). However, the LHC itself doesn't accelerate the protons from the very beginning till the maximum energy. To optimize the acceleration process, a system of pre-accelerators [49] is exploited to inject beams of protons to the LHC for the final stage acceleration (see Fig. 3.1):

- The protons from the H_2 source enter the LINAC2[1] linear accelerator. The acceleration is performed by radiofrequency cavities. The protons exit LINAC2 with an energy of 50 MeV.
- The next acceleration step is performed in the **P**ROTON **S**YNCHRONTRON **B**OOSTER (PSB). It accelerates the protons from 50 MeV to 1.4 GeV . The PSB consists of four superimposed synchrotron rings, which allows to increase the number of protons being injected to the next accelerator by factor 100.
- The **P**ROTON **S**YNCHRONTRON (PS) follows the PSB in the LHC preaccelerator chain. The PS is a synchrotron ring with a circumference of 628 m, which accel-

[1] Starting from 2018 the LINAC4 [51] will be the first preaccelerator for the LHC.

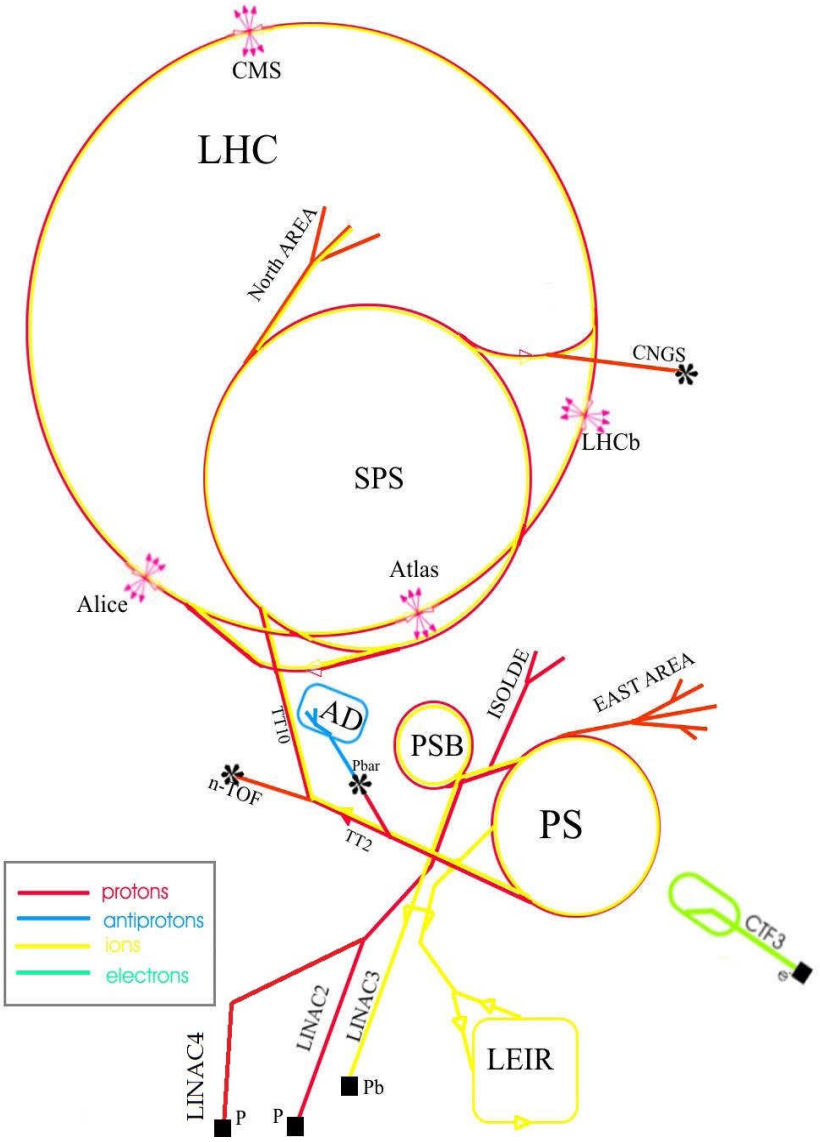

Figure 3.1: The CERN accelerator complex. The sketch also shows the particles which are accelerated and the main experiments at the LHC ring. The sketch is taken from [50].

erates protons[2] to an energy of 25 GeV and injects them to the **Super Proton Synchrontron** (SPS).

- The SPS is the second largest CERN accelerator with 7 kilometers in circumference. It injects the proton beams with an energy of 450 GeV directly to the LHC.

The LHC is a two-ring-superconducting-hadron accelerator and collider with a circumference of 26.7 km installed in the already existing tunnel which was formerly used for the e^+e^- collider LEP [52]. The LHC (LEP) tunnel is located 170 m below the ground on the territory of France and Switzerland. The bunches of protons in the LHC circulate in two different directions (clockwise ans anticlockwise) and after they reach their maximum

[2] Additionally to protons, the PS can accelerate heavy ions, alpha particles, sulphur and oxygen nuclei, electrons, positrons and antiprotons.

energy, they can collide on each other. The maximum energy of protons in one bunch in 2012 was 4 TeV, thus the pp collision centre-of-mass energy was 8 TeV. The design energy per proton at the LHC is 7 TeV.

The pp collisions, which are interesting for some physical analyses and searches, are called **events**. The number of events produced at the LHC can be expressed as follows:

$$N_{event} = L\sigma_{event}. \tag{3.1}$$

Here the σ_{event} is the cross section of the process which is studied and L is the integrated **luminosity** provided by the accelerator over some time. The integrated luminosity can be expressed through the the *instantaneous luminosity* $\mathcal{L}$, which is defined as the number of protons over the collision area per second as follows:

$$L = \int \mathcal{L} dt. \tag{3.2}$$

The design value of instantaneous luminosity for the LHC is 10^{34} cm^{-2} s^{-1}, while in 2012 the machine was providing $7.7 \cdot 10^{33}$ cm^{-2} s^{-1}. In 2012 the integrated luminosity of 23.3 fb^{-1} was provided by the collider.

3.1.1 Experiments at the LHC

There are four main experiments which analyze the data from the pp and heavy ion collisions at the LHC (see Fig. 3.1): ALICE, LHCb, ATLAS and CMS. Each of these four experiments is located around or in the vicinity of the point where the protons collide.

- The **ALICE** (**A** **L**arge **I**on **C**ollider **E**xperiment) [53] is a detecting facility which analyses the heavy ion collisions at the LHC. The main physical goals of this experiment are studies of the strongly interacting matter in the extreme density state – the quark-gluon plasma. The ALICE detector is 26 m long, 16 m high and 16 m wide and its weight is around 10000 tons.

- The **LHCb** (**L**arge **H**adron **C**ollider **b**eauty) [54] is focusing on the CP violation measurements by studying the b quarks produced in the pp collisions. The detector facility is located only on one side of the collision point. The dimensions of the LHCb detector are the following: 21 meters long, 10 meters high and 13 meters wide with a weight of 5600 tons.

- **ATLAS** (**A** **T**oroidal **L**HC **A**pparatu**S**) [55] is a general purpose detector focusing on various physics goals (Higgs boson searches, dark matter particles searches, studies of SM particles, etc.). It has a 4π geometry around the pp collision point. ATLAS is the largest LHC detector: its height and width are both 25 m and its length is 46 m. The weight of the ATLAS detector is 7000 tons,

- The **CMS** [56] is a general purpose detector as ATLAS, but it uses different technical solutions. The structure of the CMS detector will be discussed in more detail in this chapter.

There are also smaller experiments at the LHC. However, they are not located at the collision point. They are built close to the bigger experiments described above to investigate the forward particles, which are produced almost in line with the colliding protons and continue flying in the directions of the proton beams. These experiments are **TOTEM** [57] and **LHCf** [58].

3.2 The CMS Experiment

The CMS [56] is a general purpose detector. It is the heaviest experimental apparatus at the LHC – its weight is about 14000 tons with a length of 21 m and height and width of 15 m.

The design of the CMS detector has an onion-like structure formed by "layers", which are detector parts of different functionality. The positioning and structure of each detector part is in agreement with the physical goals this detector part aims to gain.

The design of the CMS detector is also restricted by the conditions of the LHC collider. The design collision frequency of the LHC is 25 ns, which leads to the need of exploiting a fast electronics to make the response of some parts of the detector, which are responsible for the event filtering, not longer than 25 ns. It is crucial not to mix up the particles from different collisions. The other restricting factor from the collider is the high luminosity which leads to 20 inelastic interactions (on average) per one bunch crossing. This means that every time the proton bunches collide around 1000 charged particles emerge from the interaction region. To distinguish the large number of reaction products, the detector parts have to be granulated enough.

Combining the need to reach the stated physical goals (detailed studies of the Standard Model particles, Higgs searches, searches beyond the Standard Model, etc.), the harsh restrictions on the LHC environment and the limited budget, the final CMS structure was formed including the following detector parts (see Fig. 3.2):

- The inner silicon tracking system insures a good charged particles momentum and spatial resolution, sufficient for the offline b identification.
- The electromagnetic calorimeter with a wide geometric coverage allow accurate measurements of the energy of the charged leptons and photons.
- The hadronic calorimeter with a hermetic coverage makes it possible to measure the whole energy deposit from the hadronic particles.
- The solenoid magnet with a strong magnetic field of 3.8 T, which is enough to bend the charged particles with a high energy and "sweep away" very low p_T particles.
- The muon detecting system, which provides an excellent muon identification with high momentum resolution and almost unambiguous muon charge identification.

More details on each detector part will be given in the following subsections.

To analyze the data derived from each detector part consistently, it is important to define a common CMS coordinate system. The coordinate system adopted by CMS has a center at the nominal collision point inside the experiment. The x-axis points towards the center of the LHC ring, the y-axis points straight upward and the z-axis points along

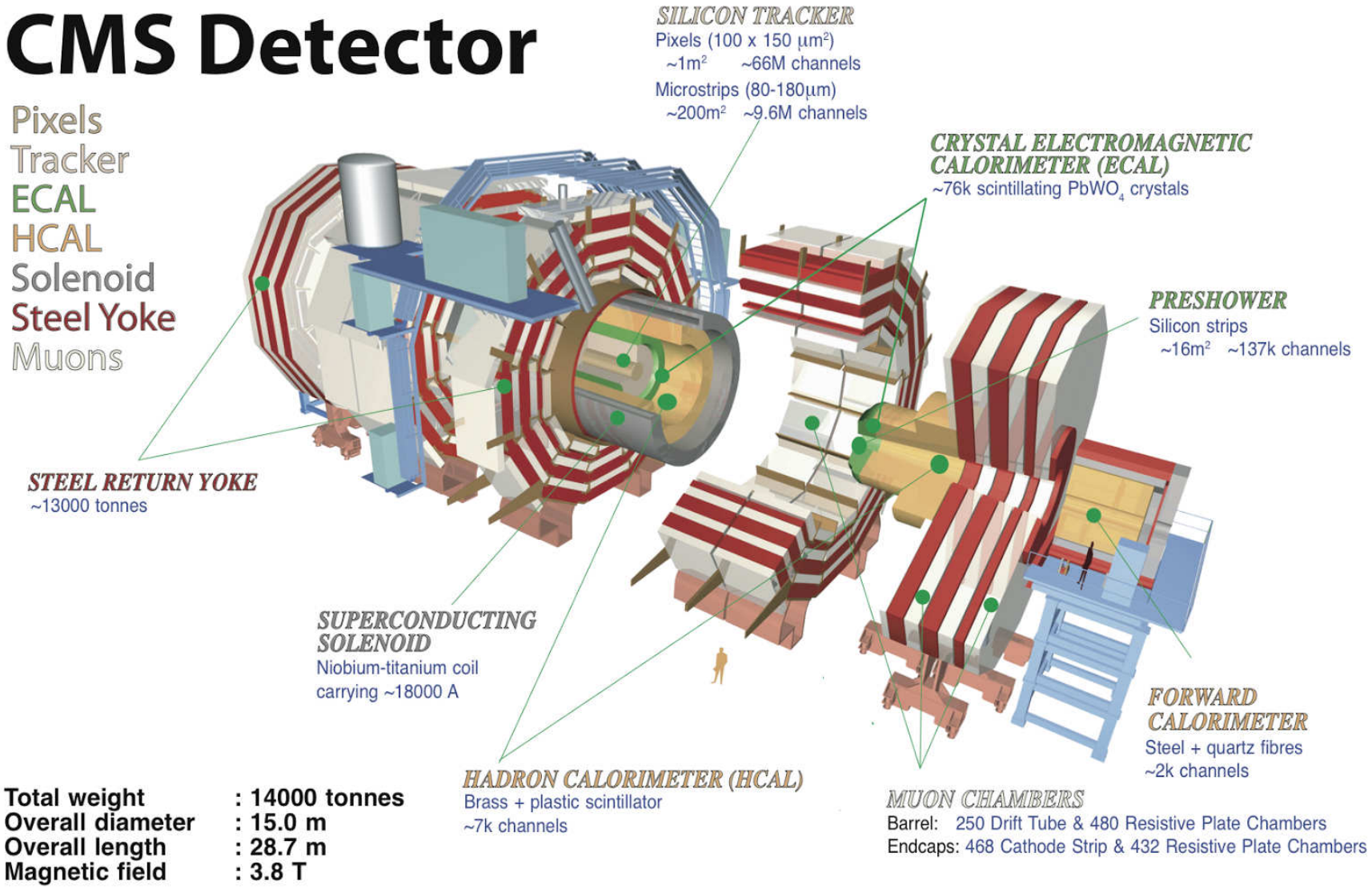

Figure 3.2: Schematic view of the CMS detector. The plot is taken from [59].

the beam direction towards the Jura mountains [56]. For the spherical coordinates, the azimuthal angle ϕ is measured in the $x-y$ plane from the x axis and the radial coordinate r is also defined in the same plane. The polar angle θ is measured from the z-axis. It is sometimes more convenient to use the *pseudorapidity* η instead of the polar angle:

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

The pseudorapidity can be also expressed through the momenta of the particles – particle momentum three-vector $\vec{p}$ and the projection of the particle momentum onto the z-axis p_z:

$$\eta = \frac{1}{2}\ln(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z}). \tag{3.4}$$

For the particles with large masses the variable *rapidity* y is often used instead of the pseudorapidity:

$$y = \frac{1}{2}\ln(\frac{E + p_z}{E - p_z}). \tag{3.5}$$

Here E is the energy of the particle. One can see from the equations 3.4 and 3.5 that if the mass of the particle is zero (or negligibly small compared to the momentum), than $\eta = y$.

The other convenient variables which are often used to analyze the data from the collisions are the transverse momenta (p_T) and transverse energies (E_T) of the objects identified using the CMS detector.

3.2.1 Silicon Tracking Detectors

The inner tracker is the detector part closest to the beam line. The task of the tracking detectors is the precise measurement of the trajectories of the charged particles emerging from the LHC collisions. As discussed before, the flux of the charged particles from the pp collisions is very high. Many of the particles, which are interesting for the further physical studies, as well as their decay products, have a short life time. That is why the CMS tracking detector, as the part of CMS closest to the collision point, has to distinguish and precisely characterize as many particles as possible, because some of them will not reach the other detector parts. The decision has been taken to make the CMS tracker fully silicon to high granularity and fast response [56]. The schematic view of the CMS tracking detector is shown in Fig. 3.3.

An average achieved transverse momentum resolution of the tracker is 2.8% for muons with $p_T = 100$ GeV and $|\eta| < 1.4$. The corresponding resolutions in transverse and longitudinal impact parameter[3] is 10 μm and 30 μm, respectively [61].

The tracking detector consists of the pixel and strip detector parts.

Pixel Tracking Detector

The pixel detector is the most inner part of the CMS tracker. It covers the pseudorapidity range $-2.5 < \eta < 2.5$, where the magnet field is uniform (see Sec. 3.2.4). The pixel

[3] An *impact parameter* is the shortest distance of a certain track to a certain vertex.

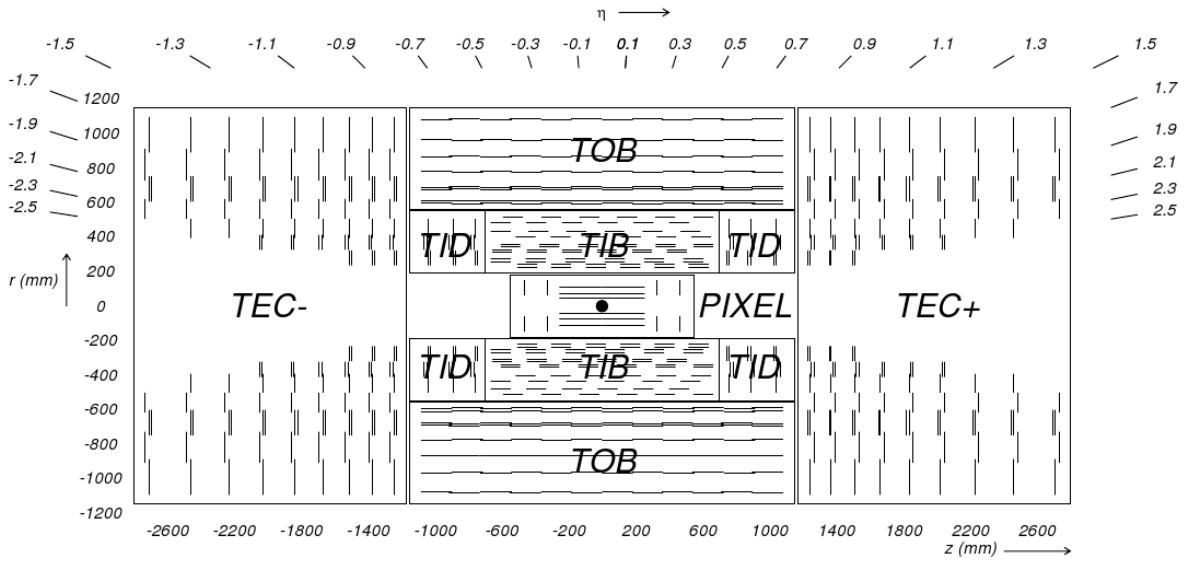

Figure 3.3: View of the CMS tracker in $r - z$ plane. It shows the pixel and the strip detectors. The plot is taken from [60].

detector consists of the barrel part (BPIX) and the forward endcap discs (FPIX) (see Fig. 3.3). The length of the BPIX part is 53 cm (in the z-direction). It consists of three concentric cylindric layers with radii 4.4, 7.3 and 10.2 cm. The BPIX consists of 48 million pixels in total. The size of each pixel is 100×150 μm^2. This results in a hit position resolution of 15-20 μm [62].

The FPIX discs are extending from a radius of 6 cm to 15 cm. They are placed on each side of the BPIX at $z = \pm 34.5$ cm and $z = \pm 46.5$ cm. The FPIX has in total 18 million pixels of the same size as for the BPIX.

The BPIX and the FPIX are aligned in such a manner that each particle, which crosses the tracking detector, will cross 3 different pixel layers or planes for a wide η range. For very high pseudorapidities the particle crosses only two discs of FPIX.

Strip Tracking Detector

The strip tracker occupies the radial region between 20 cm and 116 cm (see Fig. 3.3). It consists of several parts: **T**racker **I**nner **B**arrel (TIB), **T**racker **I**nner **D**isc (TID), **T**racker **O**uter **B**arrel (TOB) and **T**racker **E**nd**C**aps (TEC+ and TEC-, where the sign corresponds to the sign of the z-coordinate of the detector position).

The TIB and the TID extend in radius to 55 cm. The TIB consists of 4 layers and the TID is composed of 3 disks on each side of the TIB. The silicon strips in the TIB and the TID are 320 μm thick. The TIB and the TID are surrounded by the TOB, which has an outer radius of 116 cm. The TOB consists of 6 barrel layers of 500 μm thick micro-strip sensors. The TOB, as well as the TIB and the TID together, extends in z from -118 cm to $+118$ cm.

The TEC+ and the TEC- discs cover the regions 124 cm $< |z| <$ 282 cm and 22.5 cm $< r <$ 113.5 cm. Each TEC consists of 9 discs with up to 7 rings of silicon micro-strip sensors each.

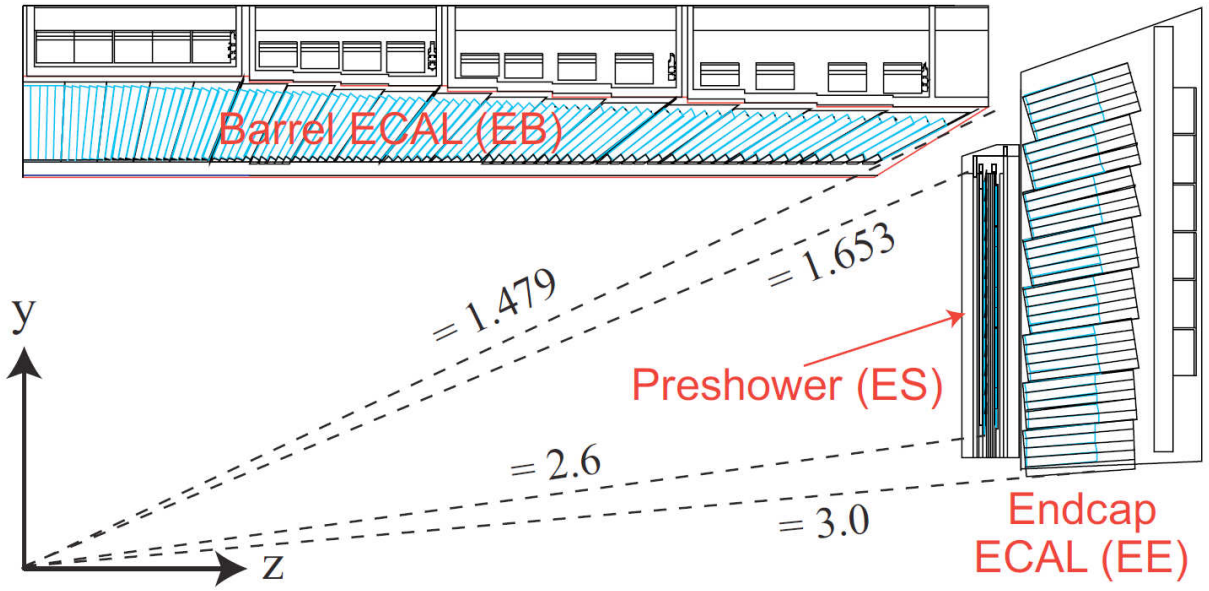

Figure 3.4: View of one quarter of the CMS ECAL. The plot is taken from [63].

3.2.2 Electromagnetic Calorimeter

The CMS **E**lectromagnetic **CAL**orimeter (ECAL) [56] (see Fig. 3.4) is a homogeneous calorimeter which consists of 61200 crystals in the barrel part and 7324 crystals in the endcaps. The crystals are made of lead tungstate ($PbWO_4$). The ECAL is complemented by a Preshower detector, which is located in front of the ECAL endcaps. The goal of the ECAL is identify leptons and photons, which enter the volume of this detector part, and measure their energies.

The material $PbWO_4$ was chosen for the ECAL because of its properties [64]: high density (8.28 g/cm^3) and short radiation length[4] (0.89 cm), which allows to make the calorimeter compact and granulated enough.

The barrel part of ECAL (EB) covers the pseudorapidity range $|\eta| < 1.479$. The length of each crystal in EB is 230 mm with an area of 22×22 mm^2 at the front side and 26×26 mm^2 at the back. Each crystal from EB is tilted by 3^o with respect to the direction from the nominal collision point to avoid particles traveling through the cracks between the crystals. The light, which the particle emits passing through the EB is detected by **A**valanche **P**hoto**D**iodes (APDs) [56].

The endcap part of ECAL (EE) covers the range $1.479 < |\eta| < 3.0$. The crystals from EE have a length of 220 mm with a surface area of 28.62×28.62 mm^2 at the front side and 30×30 mm^2 at the back side. The tilt angle of the EE crystals with respect to the direction from the nominal collision point varies from 2^o to 8^o. **V**acuum **P**hoto**T**riodes are exploited in the endcap calorimeter to collect light signals from the crystals.

The CMS Preshower aims to detect π^0 particles in the endcaps within $1.653 < |\eta| < 2.6$. In this region the two photons from π^0 decay are hard to distinguish from each other and they may be treated as one photon e.g. from the Higgs boson decay. The

[4]The *radiation length* is the distance, which a high energy electron has to travel through in matter to loose e^{-1} of its energy by breamsstrahlung.

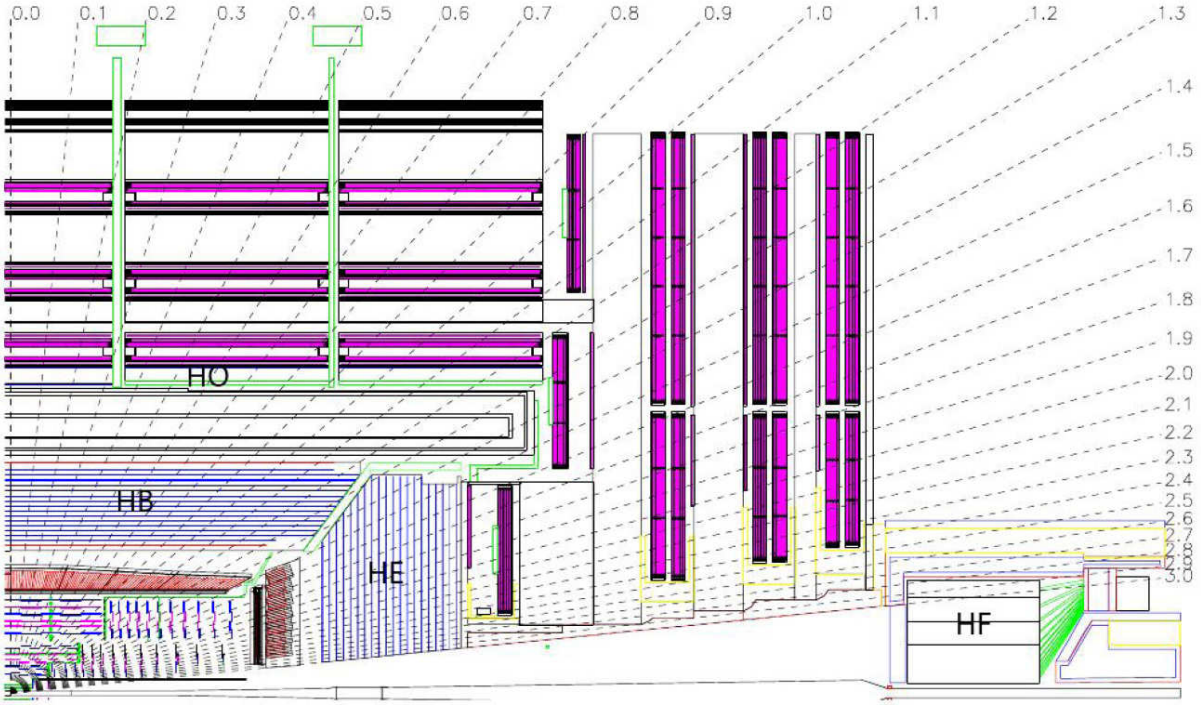

Figure 3.5: View of one quarter of the CMS HCAL in the $r - z$ plane. It shows different parts of the calorimeter. Dashed lines show the η coordinates. The plot is taken from [66].

Preshower is a finely granulated sampling calorimeter. It consists of lead radiators (where the electromagnetic shower from incoming photons is initiated) followed by silicon strip sensors to measure the energy and the profile of a shower.

The energy resolution of the ECAL for energies below 500 GeV can be parametrized as following [56]:

$$(\frac{\sigma}{E})^2 = (\frac{S}{\sqrt{E}})^2 + (\frac{N}{E})^2 + C^2, \tag{3.6}$$

where E is the measured energy, S is the stochastic term describing event-to-event fluctuations, N is the noise term describing electronics noise and C is the constant term which denotes the intercalibration errors. The energy resolution of ECAL was measured with the help of electron beams with momenta between 20 GeV and 250 GeV [65]:

$$(\frac{\sigma}{E})^2 = (\frac{2.8\%}{\sqrt{E}})^2 + (\frac{0.12\%}{E})^2 + (0.30\%)^2, \tag{3.7}$$

where E is in GeV.

3.2.3 Hadronic Calorimeter

The CMS **H**adronic **CAL**orimeter (HCAL) [56] (see Fig. 3.5) is a hermetic sampling calorimeter, which aims to absorb and measure the energy of hadronic particles, which enter its volume.

In the experimental environment of the LHC one expects that the sum of momenta in the transverse plane is zero, because the momenta of the colliding protons don't have

a transverse component in the CMS coordinate system (see Sec. 3.2). If the sum of the transverse momenta of all the reconstructed particles is not zero, it may point to the presence of undetected objects, like neutrinos. The negative sum of all the reconstructed transverse momenta ($-\sum \vec{p}_T$) is called the vector of *missing transverse energy*, or MET, $\vec{\not{E}}_T$. The HCAL plays an important role in the measurement of MET, as it aims to possibly reconstruct all hadronic particles not to loose any information. That ensures that with combining the HCAL information with the signals from other detector parts, the transverse momenta imbalance can be trustfully assigned to neutrinos and not to some particles, which were accidentally not detected.

The HCAL surrounds the ECAL. It consists of several components: the barrel part (HB) is placed between the ECAL and the solenoid magnet coil, with an inner radius of 1.77 m and outer radius of 2.95 m, covering the rapidity range of $|\eta| < 1.3$; the outer hadron calorimeter (HO), or the *tail catcher*, is placed around the magnet coil to complete the HB; the end cap HCAL (HE) covers the rapidity range of $1.3 < |\eta| < 3$; the forward part (HF) is placed at $|\eta| > 3$ and starts at 11.2 m far from the nominal interaction point to complete the coverage in pseudorapidity up to $|\eta| < 5$.

The HB consists of altering layers of absorbing material (two stainless steel plates at the innermost and outermost layer and 14 brass layers) and plastic scintillators. The particles passing through the detector volume initiate showers in the absorbers and the products of the showers emit light in the scintillators in the blue to violet wavelength region. The special wavelength-shifting fibers (they shift the light towards the green wavelength region) transfer the light from the scintillators to the hybrid photodiodes [56]. Due to the limited space between the ECAL and the magnet coil, the thickness of the HB is not enough to absorb all the expected particles. That is why an additional layer of the barrel HCAL right after the magnet (HO) was mounted. The HO utilizes the solenoid magnet bulk as an absorber, followed by a scintillator at $r = 3.82$ m. This scintillator is enclosed by a 19.5 cm thick steel absorber, which is followed by the second HO scintillator layer.

The HE consists of 18 layers of 79 mm absorbing brass planes each followed by scintillators.

The HF consists of two cylindrical structures on each side of the interaction point in the z coordinate. They are made of 5 mm thick grooved steel absorbers with quartz scintillating fibers located in the grooves. The HF has an inner radius of 12.5 cm and outer radius of 130 cm. It is azimuthally divided into 20^o wedges. The HF is operating under very harsh radiation conditions, coping with the largest fluxes of particles, as this detector part is located in the region where the proton remnants flow. It faces both electromagnetic and hadronic radiation. That is why quartz tubes located in the inner part of HF (electromagnetically interacting particles are not expected to travel deep into the detector volume) are also used to identify electromagnetic showers.

The energy resolution of the HCAL [67] was measured using charged pions with energies of 20-300 GeV resulting in the following (energy is taken in GeV):

$$(\frac{\sigma}{E})^2 = \frac{115^2}{E} + 5.5^2. \tag{3.8}$$

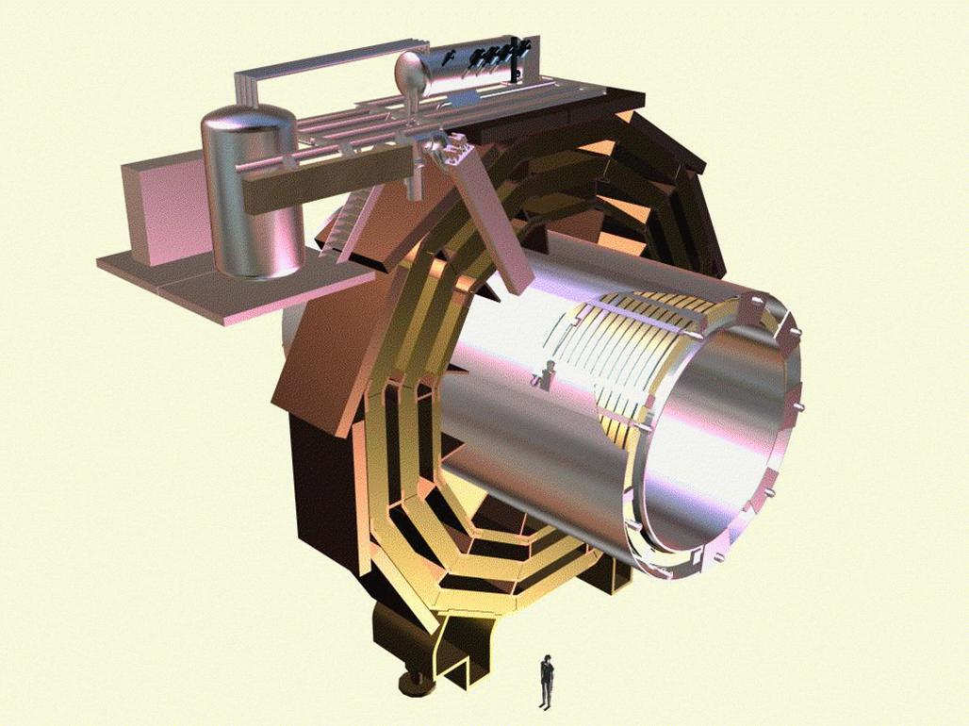

Figure 3.6: General view of the CMS solenoid magnet with the return yoke. The plot is taken from [68].

3.2.4 Superconducting Magnet

The CMS magnet [56] is the name-giving part of the detector, as it is very powerful (3.8 T) and having a compact size. The task of the magnets in the collider experiments is to bend the charged particles. The higher the momentum of the particle is, the less it bends under the influence of the magnetic field. Thus, measuring the curvature of the particle trajectory allows to measure its momentum. Moreover, the direction in which the particle bends gives the information about the charge of this particle.

The CMS magnet has a solenoid structure with a 10000 ton return yoke. It is the largest solenoid magnet ever built. The magnet is cooled to a temperature of -268.5 ^{o}C to maintain superconductivity. The length of the magnet is 12 m. The CMS detector parts described above (tracker, ECAL and HCAL) are surrounded by the magnet, while the muon detector is located outside of the solenoid. The muon detector is interleaved with three layers of the 12-sided return yoke of the magnet.

The solenoid magnet of the CMS has a strength, which is large enough to bend the high energetic particles emerging from the LHC collision and additionally to bend out the low energetic particles so that they don't reach the calorimeters. These particles are not likely to emerge from the hard processes or the decays of the rare heavy particles. The solenoid covers the central rapidity ranges, thus the magnetic field is getting non-uniform in the rapidities. That is one of the reasons why the particle reconstruction is worse in the forward direction.

The general view of the CMS magnet is presented in Fig. 3.6.

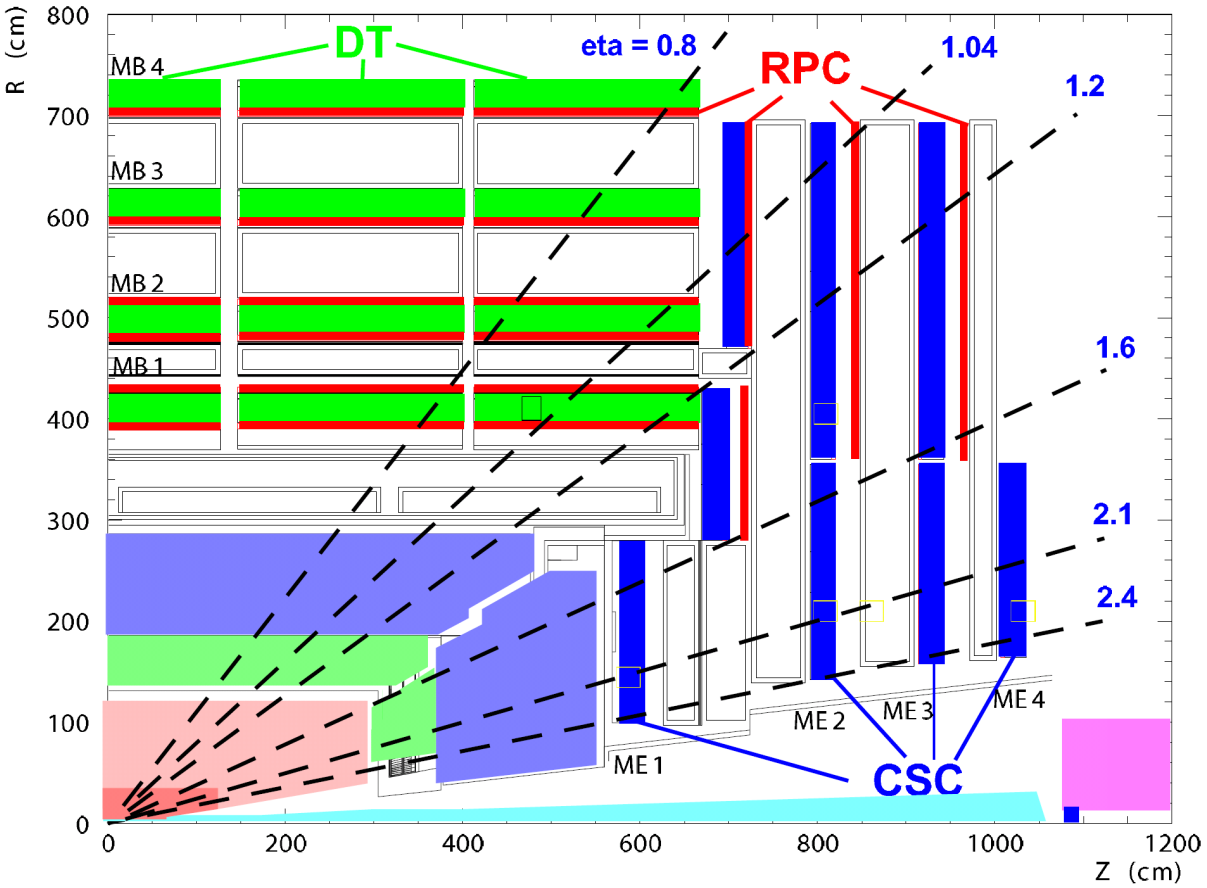

Figure 3.7: Longitudinal view of the muon system indicating the location of the three detector types contributing to the muon detector. The plot is taken from [63].

3.2.5 Muon Detector

Muons have a unique signature in the event. They travel through all the detector parts and are not stopped by the calorimeters and solenoid bulk. That is why muons can be unambiguously identified in the outer detector parts.

The precise measurement of the muon characteristics is one of the central tasks of the CMS detector. The muon detector system has to identify muons and precisely measure their momenta. The CMS magnet plays an important role in muon reconstruction as it is able to bend the high energy muons. The return yoke of the solenoid, which is placed in between the muon detector plates, acts as an additional hadron absorber in the outer detector levels to ensure that only muons are present in the muon detectors.

In total there are 1400 muon chambers at CMS: 250 **D**rift **T**ubes (DTs), 540 **C**athode **S**trip **C**hambers (CSCs) and 640 **R**esistive **P**late **C**hambers (RPCs) (see Fig. 3.7).

The DTs cover the rapidity range of $|\eta| < 1.2$ and are organized into four stations. The first three stations each contain 8 chambers that measure the muon coordinate in the $r - \phi$ plane and 4 chambers, which measure the z-coordinate of the muons. The fourth station doesn't have the z-coordinate measuring chambers. Each drift tube is 4 cm wide and consists of a positively charged stretched wire within a gas tube. The electrons, which appear due to muon ionizing the gas of the tube, are collected on the wire.

The CMS muon end cap system consists of six layers of in total 468 CSCs of trapezoidal shape, covering the pseudorapidity region $0.9 < |\eta| < 2.4$. The CSCs consists of positively charged wires crossed with negatively charged copper strips, all within the gas volume. The particles ionize the gas and the charges are collected by the strips and wires. The CSCs provide precision muon measurements and can be used for muon triggering.

The RPCs are gaseous parallel-plate detectors. The RPC is able to tag the time of an ionizing event in much smaller time than 25 ns (the designed bunch spacing at the LHC). That is why it can be used in the triggering system (see Sec. 3.2.6) to unambiguously identify the time stamp of the bunch crossing to which the muon track is associated. The RPCs are located in the barrel and in the end cap muon systems as presented in Fig. 3.7 – 6 coaxial sensitive cylinders in the barrel and 3 stations for both endcaps. Each RPC consist of two plastic plates (one positively charged and one negatively charged) with a gas volume between the plates. The electrons which appear in the gas volume due to ionization by a muon are collected by the external metallic strips with a precisely known time delay.

3.2.6 Triggering and Data Acquisition Systems

The nominal frequency of the LHC is 40 MHz, which corresponds to a bunch spacing of 25 ns. As was discussed before, there are on average 20 *pp* interactions per bunch crossing. This results in an enormous amount of data from each event. It is impossible to store all the information from the LHC performance period. That is why it is important to filter only the events which potentially have a hard interaction and interesting physical processes. The decision if the event has to be skipped or is stored is taken by the CMS trigger system.

The triggering system has two stages: **L**evel-**1** (L1) trigger and **H**igh **L**evel **T**rigger (HLT) [56].

The L1 trigger consists of custom-designed programmable electronics. It uses the coarsely segmented data from calorimeters and muon detectors to make a choice to keep or to skip the event. In the meanwhile the full event information is kept in pipelined memories in the front-end electronics. The L1 hardware is implemented using FPGA[5] technologies, ASICs[6] and programmable look-up-tables. The design total rate of the L1 trigger is 100 kHz. The architecture of the L1 trigger is presented in Fig.3.8.

The HLT exploits the full event information making a choice based on some primer analysis (similarly to the offline analysis algorithms). It reads the event information out of the front-end electronics memory, analyses it and forwards the accepted events.

The final rate after L1 and HLT triggers is expected to be 40 Hz.

[5] **F**ield-**P**rogrammable **G**ate **A**rray (FPGA) is an integrated circuit, which can be configured by a customer for multipurpose use.

[6] **A**pplication-**S**pecific **I**ntegrated **C**ircuit (ASIC) is an integrated circuit, which can be customized for particular use.

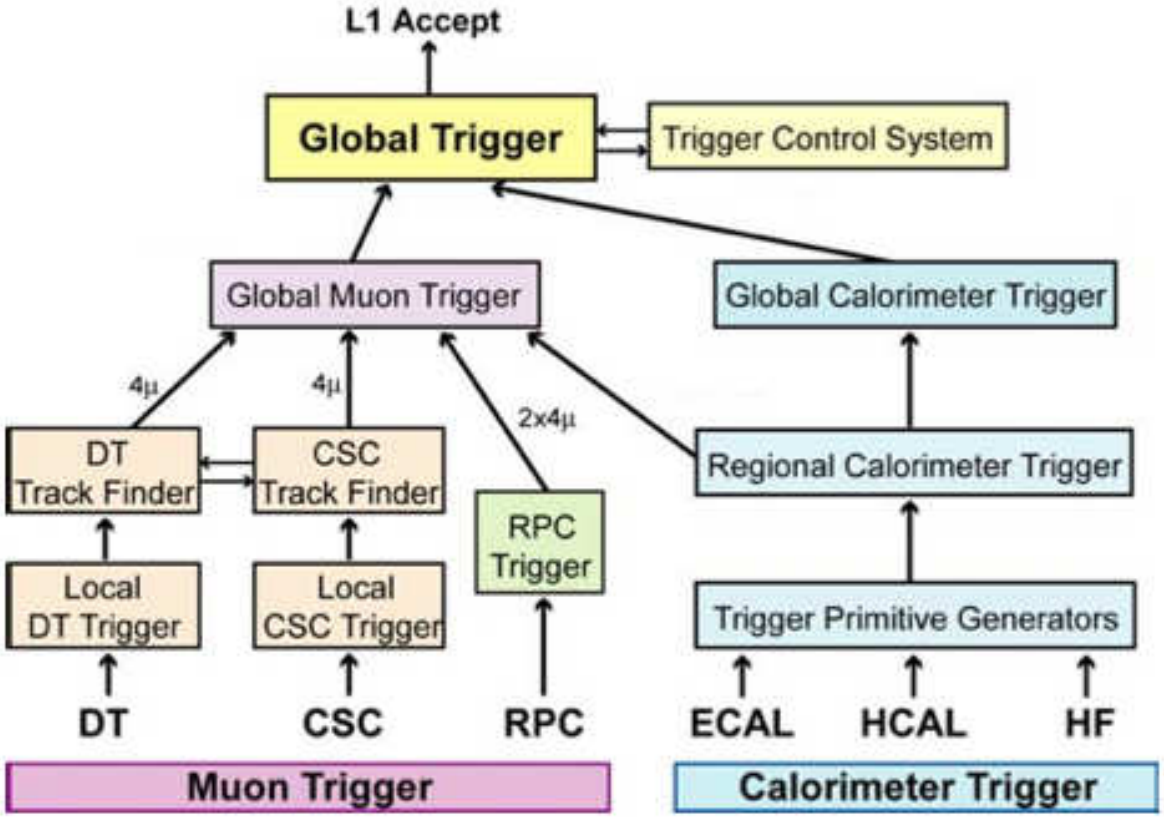

Figure 3.8: Architecture of the CMS L1 trigger. The plot is taken from [56].

4 | Simulation

In the *pp* collisions from the LHC a variety of SM processes of different complexity are produced. Not every detail of these processes can be (easily) theoretically calculated. The reason for that might be that the processes are not possible to be treated with the perturbative QCD approach, or the calculations are too complicated. Moreover, the physicists do not have a "pure" information about the products of the *pp* collisions. What they have are the signals from the particle detectors, like CMS (see sec. 3.2). Thus, to theoretically reproduce these processes, one also needs to take the interaction of the particles with the detector matter to account.

These complicated calculations at particle physics experiments are usually performed using Monte Carlo probabilistic approaches implemented in the particle physics generators and tools for the detector simulation:

1. **Event generation** is done in in the generators. They simulate the processes starting from the collisions until the collision products enter the detector volume. Some of the information, provided by the event generators is unavailable in the real particle physics experiments, thus allows to carry out various studies and checks.

2. **Detector simulation** includes a full detector modeling and describes the evolution of the particles, which were produced by the event generators, inside the detector volume, the interaction with the detector materials and the resulting creation and recording of signals (e.g. charge depositions, lights, etc.) in the active detector parts.

3. **Reconstruction**, which uses the same strategy as the reconstruction of the objects obtained from the real experimental facilities (will be described in Chapter 5).

The main directions of usage of the MC tools in this analysis are the following [69]:

- *Estimation of detector acceptance and resolution effects*: the simulated data has the information about the physical processes "before" it was detected, i.e. on the event generation level. This gives the possibility to estimate the detector efficiencies and corrections for migrations and resolution effects.

- *Comparison to experimental results*: there are certain (Standard Model) predictions implemented as a core of the MC generators. The comparison of the experimental results to the simulation gives an estimate of the consistency of the theory, on which a certain generator is based.

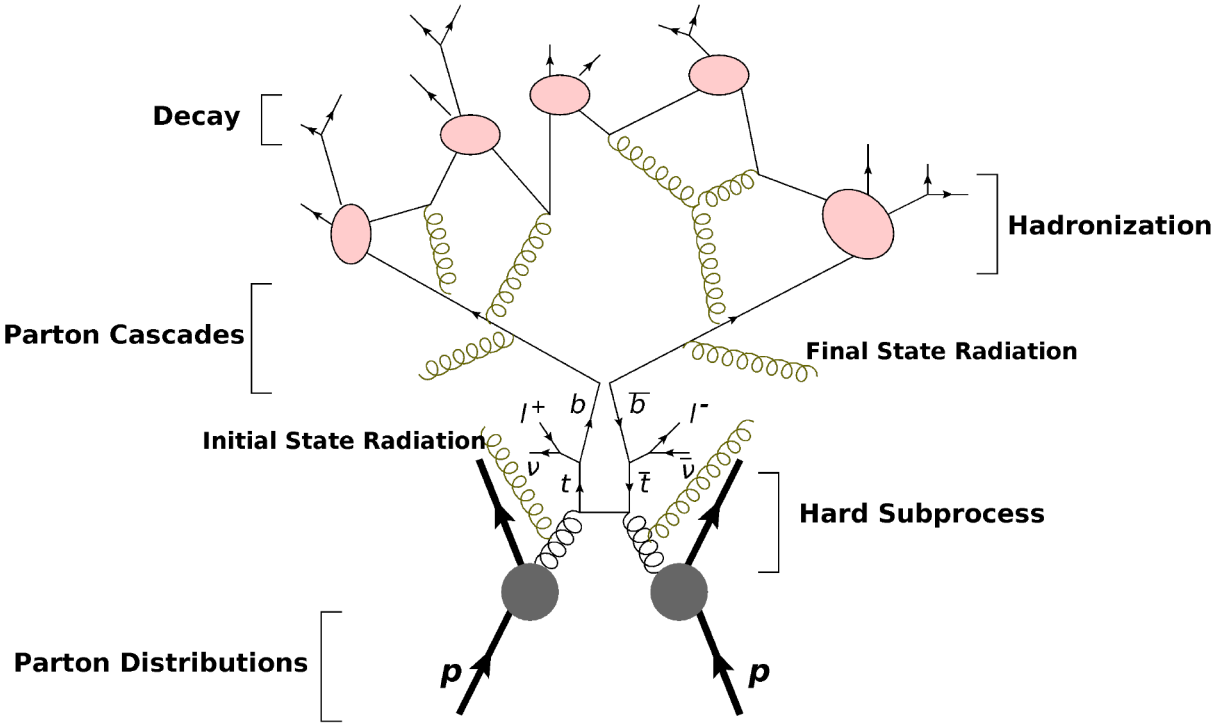

Figure 4.1: Illustration of the generation of the $t\bar{t}$ production and decay in the MC generators.

The MC generators model the events in several steps: hard interaction, showering (or parton cascades), hadronization and hadronic decays. All these subprocesses are illustrated in Fig. 4.1 (on the example of $t\bar{t}$ production in pp collision). The matrix elements (ME) in the hard process are calculated using perturbative QCD up to certain order of the α_s expansion (see Sec. 2.3.5). All the hard processes and initial and final state radiations, which could not be described in the ME perturbative calculations are modeled in the parton showering step. The non-perturbative transition from partons to hadrons is simulated in the hadronization part.

In this chapter a brief overview of the generators and tools to simulate the detector, which were used for this work, is given.

4.1 Pythia

Pythia [69] is a multipurpose event generator used in particle physics, which provides the possibility to generate all the processes in an event, starting from the initial hard interaction and ending up with the hadronization and decays.

Pythia is used to simulate e^+e^-, ep and pp collisions and simulates a large variety (over 300) of hard $2 \to 2$ processes, with two initial and two final state partons (including also many beyond Standard Model processes). However, they are mostly calculated with the LO QCD accuracy. The cross sections of the hard processes are calculated using the factorization theorem (see Sec. 2.5.1). The PDFs which are used for the hard process calculations are taken from external PDF sets.

The showering process describes the possible radiations, which are not taken to account

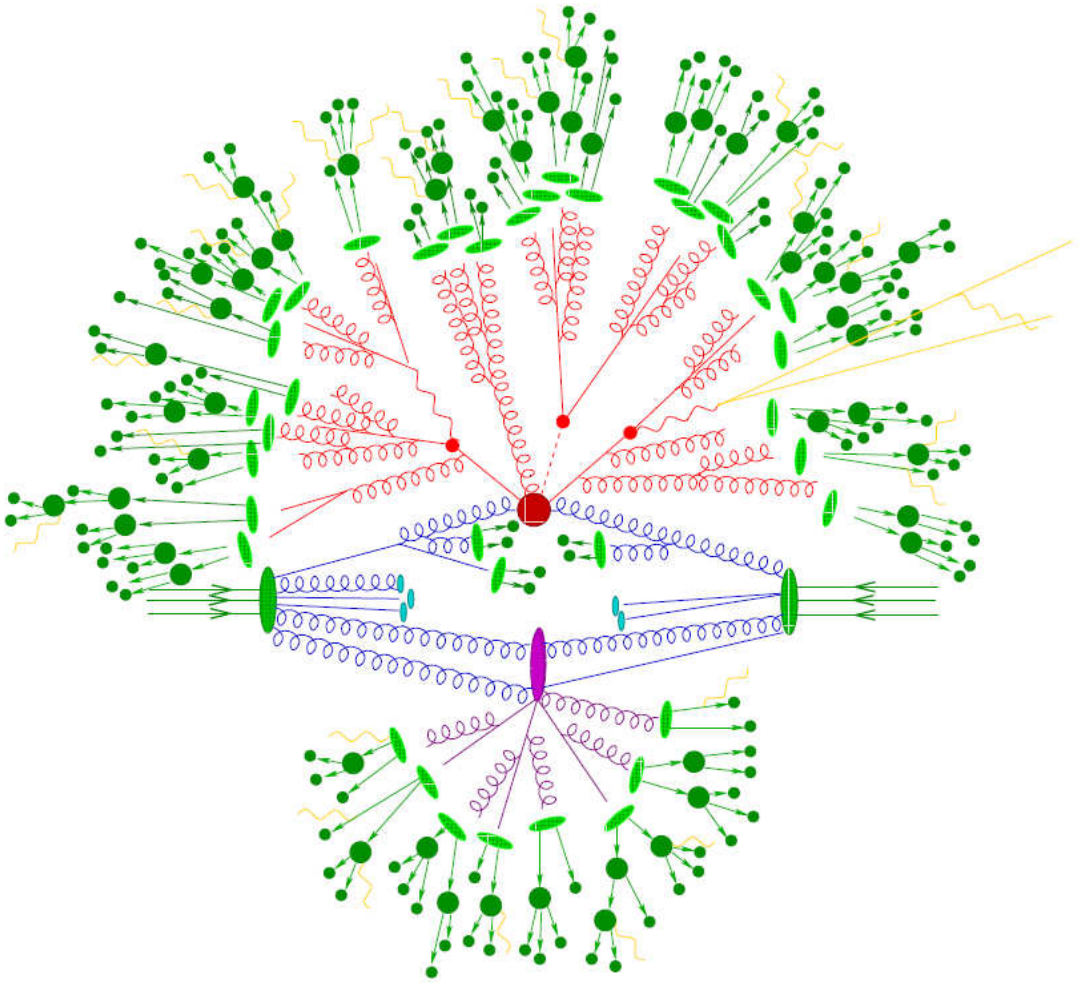

Figure 4.2: Illustration of the generation of an event in the MC generators. The hard interaction is represented with a red circular blob and the underlying event is illustrated with an elliptic magenta blob. The plot is taken from [71].

in the hard subprocess calculations. In the parton showering approach one allows splittings of one parton into two (or multiple) other partons until the arbitrary energy limit Q_0 is reached. In PYTHIA showering the partons are ordered by their p_T and $Q^2 = p_T^2$ - the closer the parton approaches the vicinity of the hard process, the higher p_T is assigned to it. This is called a *p_T-ordered showering*.

To describe the radiations from the beam particle remnants (in case of the LHC these are proton remnants) and from the multiple parton interactions in one collision, or *underlying event*, an additional $Z2^*$ tune [70] is used in this work for the samples generated with PYTHIA. In Fig. 4.2 the full event generation with underlying event is illustrated.

The *Lund string model* [72] is used in PYTHIA to describe hadronization. The main idea of the processes, which are assumed in this model, are illustrated in Fig. 2.6. A simplistic description of the model is the following: the field lines between partons are assumed to be strings, in which the tension grows the further the partons split from each other. When the critical tension is reached, the string breaks producing a quark-antiquark

pair or a pair of diquarks at the point of breaking. This process of parton-splittings is continued until there is not enough energy left to produce a quark-antiquark pair. All the connected pairs and triplets of quarks produced in this process are hadrons.

The history of the Lund hadronization model goes back to the JETSET generator, developed first to describe in detail the non-perturbative hadronization and hadronic decays. Later, PYTHIA was started to develop separately, interfacing to JETSET to describe hadronization and decays. Starting from the PYTHIA version 6, JETSET was merged to PYTHIA.

For this work, PYTHIA6.4 was exploited.

4.2 Herwig

HERWIG [73] is a multipurpose generator, which, as PYTHIA, can provide the description of all subprocesses of the event. However, HERWIG uses different approaches and algorithms than PYTHIA to generate different event parts. For instance:

- **Parton showering**: instead of using p_T-ordering, the HERWIG tool implements an *angular-ordered showering*, and the energy scale of the partons is expressed as $Q^2 = 2E_p^2(1-\cos\theta)$. Here, E_p denotes the energy of the parent parton and θ is the angle between the parent parton and the emitted parton.
- **Hadronization**: in HERWIG the *cluster hadronization model* is exploited [73]. The gluons are forced to split into quark-antiquark or to diquark-antidiquark pairs. The quarks and diquarks are then grouped into colorless clusters. Clusters with lower mass are preferred. An illustration of the Lund string model and the cluster model is provided in Fig. 4.3.

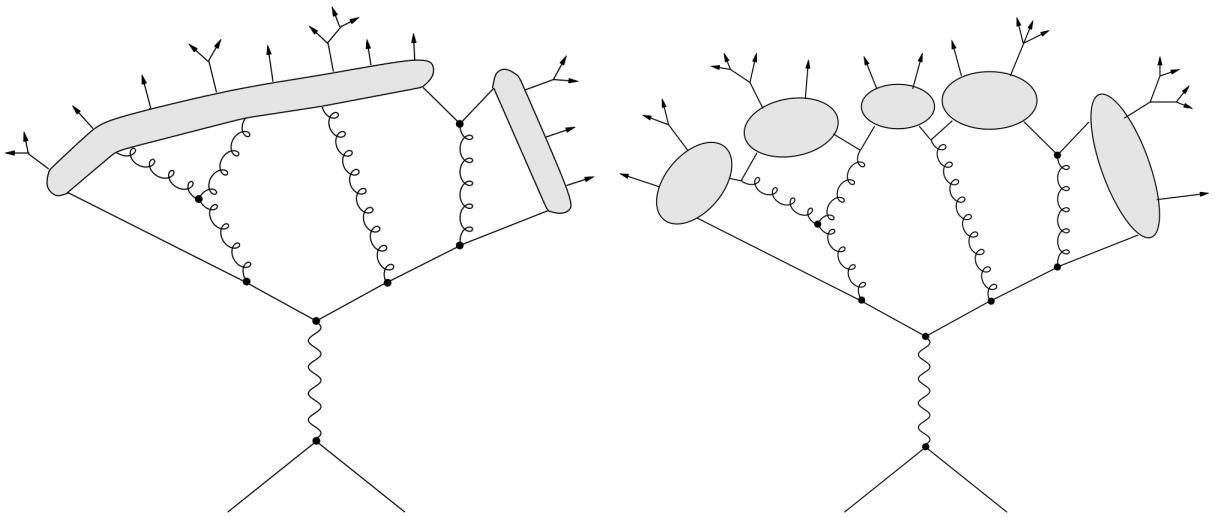

Figure 4.3: Illustrative comparison of the string (left) and cluster (right) hadronization models. The figures are taken from [74].

For this analysis the AUET2 tune [75] is exploited in HERWIG to simulate the underlying event.

4.3 MadGraph

The MadGraph generator [76] is used in this analysis to generate the $t\bar{t}$ signal sample. MadGraph is a general purpose event generator, which can generate only the hard process (*matrix element generator*) to a given order of pQCD. Only tree level diagrams (e.g. with no virtual loops) are simulated in MadGraph.

To describe the showering, hadronization, MadGraph is interfaced with Pythia or Herwig. To avoid the double counting of radiation from the hard process and the showering, the MLM matching scheme [77] is applied for MadGraph.

The CTEQ6L1 [29] proton PDF sets are used in this work for MadGraph simulation.

4.4 Powheg

The Powheg event generator [78] also generates events at matrix element level and needs to be interfaced with Pythia or Herwig for the parton showering and hadronization. Powheg provides the hard process at NLO QCD accuracy. To describe the structure of the proton the CT10 PDF sets [79] were used for the samples generated with Powheg.

4.5 MC@NLO

The MC@NLO [80] tool provides the hard process simulation with NLO QCD accuracy. For showering and hadronization, the MC@NLO tool is interfaced with Herwig. While exploiting the MC@NLO tool, one has to consider that it provides a small fraction of events with negative weights.

In this work the CTEQ6M PDF sets [29] are used for the samples generated with MC@NLO.

4.6 Detector Simulation with Geant4

A very powerful tool used to model the particle physics detectors is called Geant4 [81]. It is a software package generally used to accurately simulate the passage of particles through matter. The Geant4 toolkit includes a large variety of the models of particle interaction with matter.

The main applications of Geant4 are the simulation of the materials and geometry of the whole experimental facility, tracking of particles, their interaction with matter and the detector response. The visualization of the detector and the particle tracks is also possible.

Geant4 is used for the CMS detector simulation. The event display of the CMS uses the geometrical visualization of the detector (see Fig. 4.4).

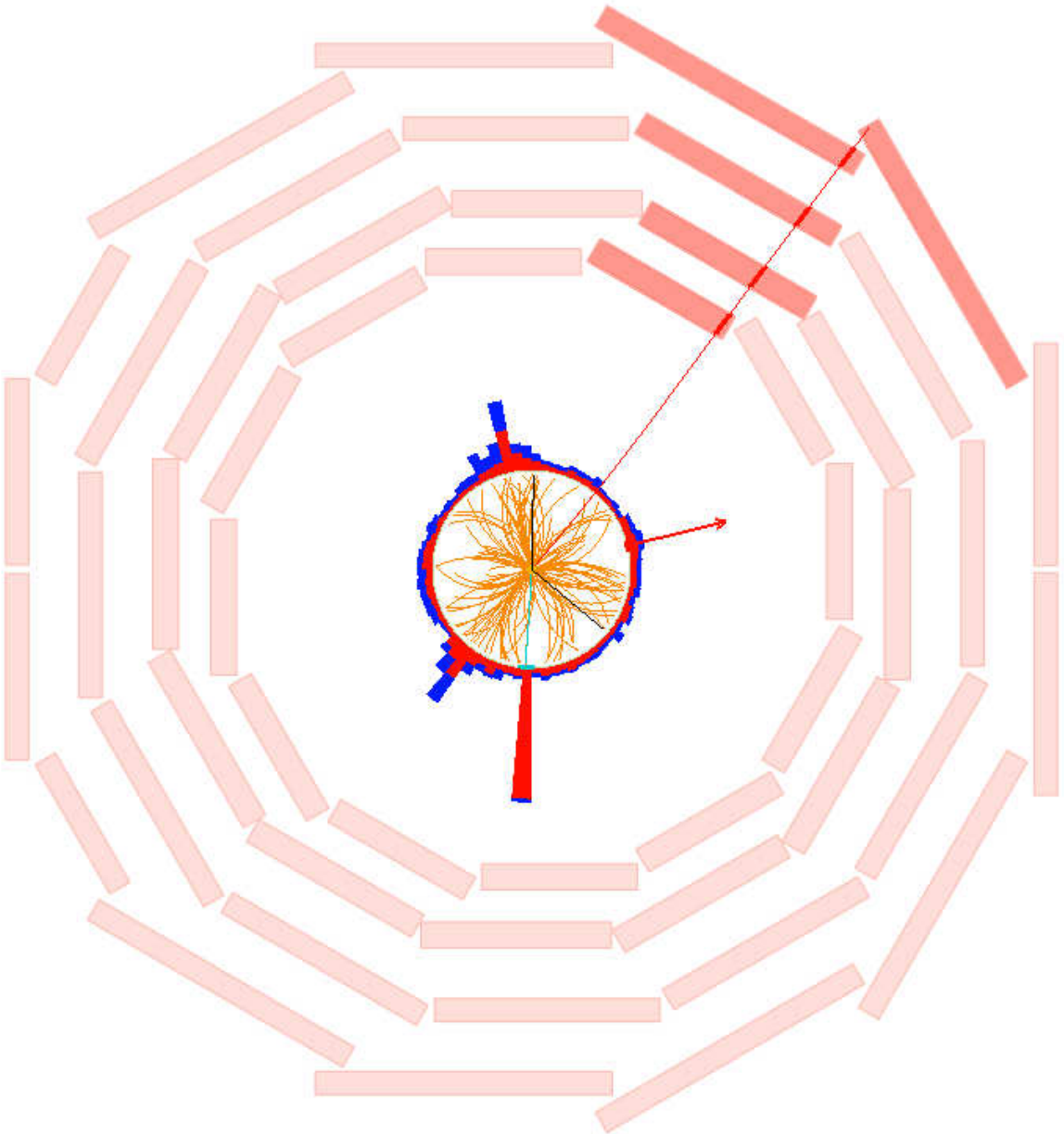

Figure 4.4: CMS event display: simulated geometry of the CMS detector (radial view) with the event where the $t\bar{t}$ pair is produced.

5 | Object Reconstruction

In this chapter the reconstruction of the physical objects used for the analysis is described.

5.1 Track Reconstruction

One of the basic objects that is reconstructed in the CMS detector is a *track*, or a curve which represents a trace of some particle in the detector volume. At the beginning the tracks are reconstructed as separate objects, which are not assigned to any particle. On the later steps of the reconstruction and identification, tracks which match certain physical requirements are assigned to a particular physical objects.

The sensors of the tracker (pixels and strips) have a signal threshold. All the signals above this threshold are accepted. Then the signals from neighboring sensors are clustered into *hits*. The information about hits is exploited to reconstruct tracks.

The track reconstruction software at CMS is called **C**ombinatorial **T**rack **F**inder (CTF) [62], which is an adaption of the combinatorial Kalman filter [82–84]. The CTF has four steps:

- **Seed generation**: provides initial track candidates using 2-3 tracker hits;
- **Track finding**: searches for additional hits which can be assigned to the track. Track finding is based on the Kalman filter [85].
- **Track fitting**: determines the track trajectory parameters using additional hits. The Kalman filter is used for track finding.
- **Track selection**: checks if the tracks satisfy the quality criteria (for example, the track has to intersect some minimum number of detector layers and the quality of the trajectory fit has to be high).

The CTF cycle is repeated six times. Each of these iterations differs one from another by the configurations for seed tracks and by the track selection requirements. On the first iteration the track quality requirements are the highest. After reconstruction of the high quality tracks, the hits used for these tracks are blinded. On each further iteration the configurations for seed tracks and the quality requirements are getting less tight.

5.2 Primary Vertex Reconstruction

The *primary vertex* is the point with given coordinates and uncertainties which denotes the location of a proton-proton interaction. The points of primary interaction are determined separately for each event.

For the primary vertex reconstruction the information about the *beam spot* is exploited. The beam spot is the 3D luminous region where the LHC beams collide inside the CMS detector. Its position is defined as the average collision point over many events.

The primary vertex reconstruction follows three steps [62]:

- **Track selection**: select the tracks which are consistent with being produced in the beam spot region[1].
- **Track clustering**: the z coordinates of the points of the closest approach of the selected tracks to the vertex are analyzed. It is being checked if the z position of the point of the closest approach of one track is not too far from the other track's z coordinates [86]. If it is the case, the track is not used further for this vertex reconstruction.
- **Fitting the vertex position** using the tracks selected in the previous two steps.

In this procedure all the coordinates of the primary vertices in the event are determined. In the LHC the average number of interaction vertices per bunch crossing is 25 (see Fig. 5.1). The effect of presence of multiple primary vertices in one event is generally called *pileup*.

5.3 Reconstruction Algorithms of the Physical Objects

The information about the particles can in principle be extracted from some particular dedicated part of the CMS detector: only tracker, only calorimeter or only muon system. However, combining information from different detector parts makes the reconstruction more precise.

In the CMS the *particle flow algorithm* [88] is widely exploited to gain a good particle reconstruction.

5.3.1 Particle Flow Algorithm

The **P**article **F**low (PF) algorithm aims to efficiently reconstruct particles optimally combining information from different detector parts. This algorithm is successfully deployed in the CMS experiment, profiting from the excellent track reconstruction and finely granulated calorimeters.

In the CMS PF algorithm [88] the muons are reconstructed first. For their reconstruction the tracks from the tracking detectors are extrapolated towards the calorimeters and

[1] In practice this is done by restricting the significance of the track transverse impact parameter relative to the beam spot to be less than 5. The impact parameter d is the shortest distance from the track to some object – in this case, to the beam spot. The significance s of the impact parameter denotes the impact parameter divided by its error: $s = \frac{d}{\sigma_d}$

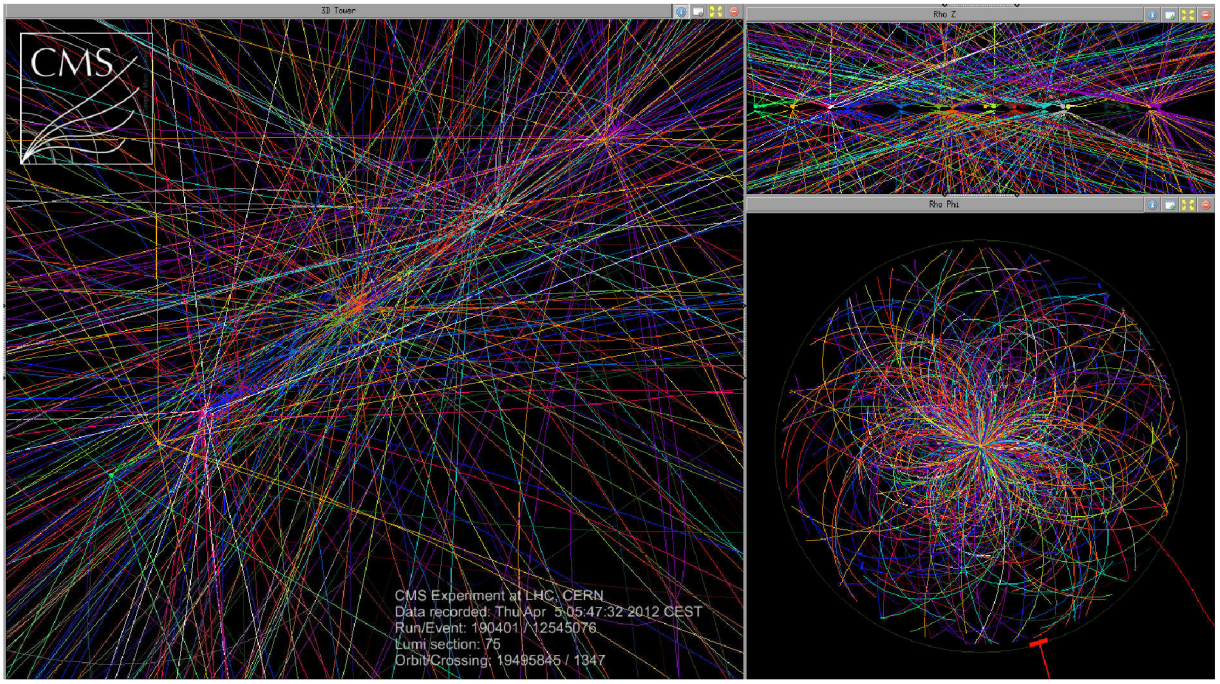

Figure 5.1: CMS event display. An event with 29 distinct vertices is shown. The plot is taken from [87].

combined with the calorimeter signals. Furthermore, the signals from muon detectors are also associated with those tracks to form the muon candidates. All the tracks, calorimeter and muon detector signals associated to the muons are blinded for the reconstruction of further particles.

The next step of the CMS PF algorithm after muon reconstruction is the electron reconstruction. In this case the tracks from tracker are also extrapolated and associated to the calorimeter signals. However, due to frequent Bremsstrahlung, one needs to properly attach the calorimeter signals from photons to the electron candidates.

After the tracks and calorimeter signals assigned to electrons and muons are blinded, the charged hadrons are reconstructed and then blinded.

Once all the tracks are treated, the remaining signals in the ECAL are assigned to the photons and the signals from HCAL – to the neutral hadrons.

After all the detector signals are associated to some particle, the energies and momenta are calculated. The information from different detector parts is also combined to obtain the best set of object characteristics. For example, if the energy from HCAL and ECAL associated to some track has a momentum excess compared to the momentum determined in the tracker, this excess is attributed to an overlapping neutral particle, which carries this extra momentum.

5.4 Reconstruction of Physical Objects

In this work the $t\bar{t}$ decaying in the dilepton channel are studied. Therefore the reconstruction of muons, electrons, jets originating from b-quarks and of neutrinos, represented as MET in the detector is very important. This section describes how these physical objects

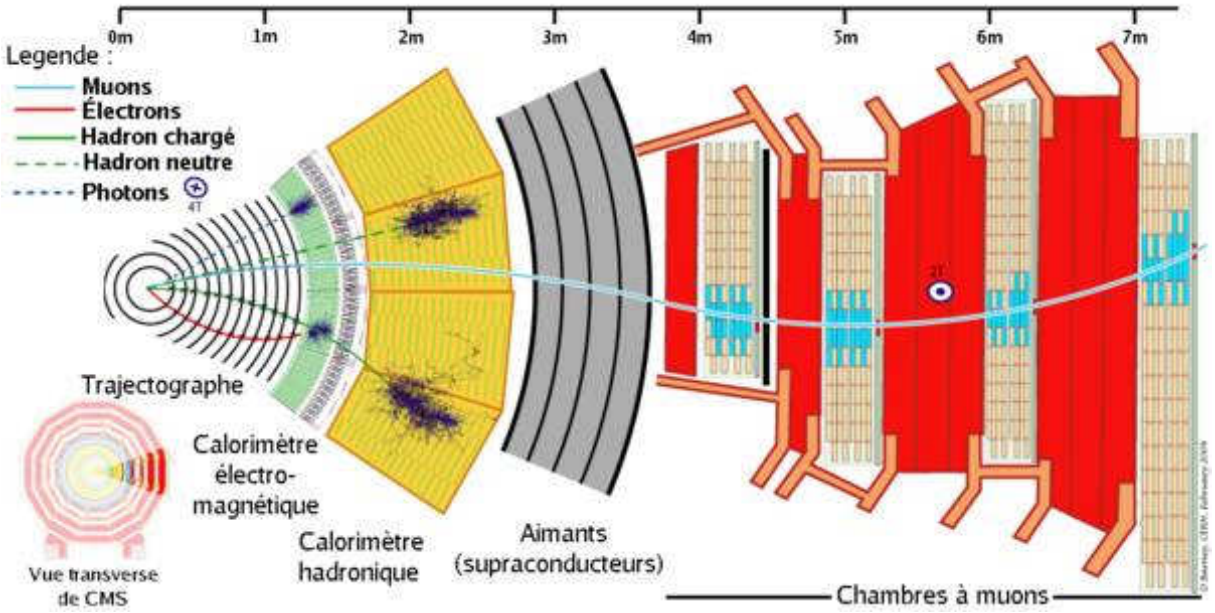

Figure 5.2: Transverse view of a sector of the CMS detector. The expected trajectories of different particles in the detector are marked with different colors. The plot is taken from [89].

are reconstructed at CMS.

5.4.1 Muon Reconstruction

The muons can be almost unambiguously identified in the CMS detector as they travel through the whole detector, unlike the other charged particles which mostly stop in the calorimeters or magnet bulk (see Fig. 5.2).

The muon reconstruction approaches used in CMS analyses follow two strategies [90]:

- **Global Muon Reconstruction** (outside-in): all the muon tracks reconstructed using muon detectors only (standalone muons) are matched to the tracker tracks. This is done by checking the consistency of the tracker track and a standalone muon track, propagated onto a common surface. The global muon tracks are fitted with the Kalman filter technique combining the hits from tracker tracks and standalone muon tracks.

- **Tracker Muon Reconstruction** (inside-out): the tracker tracks with $p_T > 0.5$ GeV and $p > 2.5$ GeV are treated as muon candidates and are extrapolated to the muon detectors region. If at least one muon segment is matched to the extrapolated track then this track is qualified as a tracker muon.

The tracker muon reconstruction is more efficient at lower momenta (approximately $p < 5$ GeV), while the global muon reconstruction is more efficient for reconstructing more energetic muons.

5.4.2 Electron Reconstruction

In the CMS PF algorithm, the electrons are reconstructed using the information from the tracking detector and from the ECAL [91].

There are two complementary algorithms to generate the electron track seeds by combining tracks the from tracker and the ECAL superclusters[2] [62]:

- **ECAL driven seeding** is performed by matching the ECAL superclusters to the tracks from the tracker. Each ECAL supercluster is assumed to originate from the electron and the track which can be matched to this supercluster is searched for. This strategy is exploited for the electron candidates with $p_T > 5$ GeV.
- **Tracker driven seeding** is performed by matching the tracks to the ECAL superclusters. If no supercluster for a track is found, the track is no longer regarded as an electron candidate. The tracker driven seeding approach is exploited for the electron candidates with $p_T \leq 5$ GeV.

After the seeds for electron candidates are selected by one of the methods, they are further treated following the steps for a regular track reconstruction (see Sec. 5.1). The final track fitting, however, is somewhat modified, as the electrons loose their energy in the tracker mostly via Bremsstrahlung, not by ionizing within the tracker volume. The track fitting is performed exploiting the **G**aussian-**S**um **F**ilter (GSM) [92] as the Kalman filter technique is not describing properly the electron behavior in the detector. The tracks reconstructed this way are called *GSM tracks*.

Finally, the track selection (see Sec. 5.1) is performed using a **M**ulti **V**ariate **A**nalysis (MVA), for the electron candidates based on the tracker seed [91]. For the electron candidates emerging from the ECAL, the seeds are selected by applying restrictions in η and ϕ [93].

5.4.3 Jet Reconstruction

Each quark and gluon, as a particle with non-zero color (see Sec. 2.3.5) starts to hadronize[3] and form bunches of particles, which generally flow in one direction. These bunches of particles are called *jets*. There are various algorithms of jet reconstruction, but all of them finally aim to correctly group together all the particles which emerged from one quark or gluon.

Jet Finding Algorithms

The jet reconstruction algorithms are generally divided into two subgroups – cone algorithms and sequential clustering algorithms.

The *cone algorithms* are based on the idea of reconstructing geometrically cone-shaped jets around some direction of the main energy flow [94]. Jets reconstructed using cone algorithms usually have the property that their cone axis coincides with the total four-momentum vector of all the particles which fall within this jet. Despite the relative simplicity and straightforward implementation of the cone algorithms, they have a strong

[2] *ECAL superclusters* are formed by merging ECAL clusters of similar η over some range of ϕ.

[3] Except for the t-quark, which decays before hadronizing (see Sec. 2.5).

disadvantage: most of the cone algorithms (excluding SISCone [94]) are not collinear[4] and infrared[5] safe.

The *sequential clustering algorithms* are represented by k_T [95], Cambridge-Aachen [96] and anti-k_T [97] algorithms. They are infrared and collinear safe. All these three algorithms have a common basic procedure as follows [97]. First they define a minimal distance d_{ij} between two particles (or pseudojets, as will be discussed further) i and j and a distance d_{iB} between particle (or pseudojet) i and the beam B. If $d_{ij} < d_{iB}$, where i and j runs over all the objects in the event, then the objects i and j are combined together to a pseudojet and the whole procedure is repeated. If $d_{ij} > d_{iB}$ for all the j, then the object i is assumed to be a jet and is excluded from the list of objects for further combinations.

One should also mention that the word "distance" doesn't have a purely geometrical meaning for the sequential clustering algorithms. It is defined the following way [97]:

$$d_{ij} = min(k_{Ti}^{2p}, k_{Tj}^{2p})\frac{\Delta_{ij}^2}{R^2} \tag{5.1}$$

and

$$d_{iB} = k_{Ti}^{2p}. \tag{5.2}$$

Here k_{Ti} is the transverse momentum of the object i; $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, where y and ϕ are the rapidity and azimuth angle of the corresponding object; R denotes the radius parameter. The parameter p is responsible for the relative power of energy and geometrical scales. Different sequential clustering algorithms denote basically different p. The $p > 0$ corresponds to the k_T algorithm, $p = 0$ represents the Cambridge-Aachen algorithm and $p < 0$ is chosen for the anti-k_T algorithm.

For the analysis described in the following the anti-k_T jet algorithm is used. It is the only sequential clustering algorithm which produces circular cone-shaped jets.

Tagging of the b-jets

It is not always clear what the nature of the particular reconstructed jet is. However, one can strive to find out from which particle a jet emerged. Jets, which were formed by b-quarks, are of particular interest for the studies of the top-quark (see Sec. 2.5.2).

b-hadrons are relatively long lived particles ($\sim 10^{-12}$ s). This property allows the reconstruction of a *secondary vertex* inside a jet emerging from a b-quark. The secondary vertices are reconstructed as points of common origin of sets of charged particles using the adaptive vertex fitter [98]. A schematic illustration of a secondary vertex and other related quantities is presented in Fig. 5.3. Additionally, a b-quark has a relatively large mass, which means that the secondary vertices from the b-hadrons have a high number of associated tracks with large impact parameters with respect to the primary vertex (see Fig. 5.3). All these properties are used for the tagging of the b-jets.

There are several b-tagging algorithms used for CMS data analysis [100]. All of them define a single discriminant value for the b-tagging. By restricting some minimal value

[4] The *collinear safe* jet finding algorithms provide the same set of reconstructed hard jets not depending on the possible presence of collinear splittings, which can occur as a part of the fragmentation process.

[5] If the resulting set of jets is not changed by adding or subtracting a soft particle to the bunch of particles over which a jet finding algorithm is applied, this jet finding algorithm is regarded as *infrared safe*.

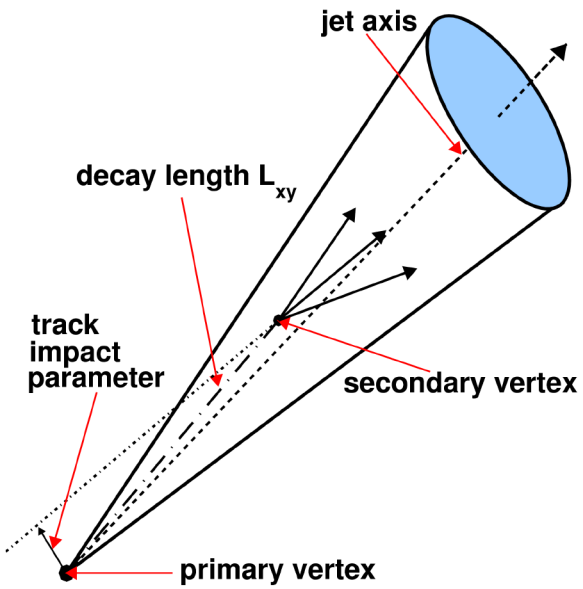

Figure 5.3: Schematic view of a b-hadron decay inside a jet resulting in a secondary vertex with three charged particle tracks. The track impact parameter, which is the distance of closest approach between the extrapolation of the track and the primary vertex, is shown in addition for one of the secondary tracks. The plot is taken from [99].

of the b-tagging discriminant for a jet, one gets a sample enriched in b-jets. Some of the b-tagging algorithms are listed below:

- **T**rack **C**ounting **H**igh **P**urity (TCHP). In this algorithm the impact parameter[6] significance[7] of the track with the third highest impact parameter significance in the jet plays the role of the b-tagging discriminant.
- **J**et **P**robability (JP). This method assigns to every jet an impact-parameters dependent discriminant.
- **C**ombined **S**econdary **V**ertex (CSV). In this algorithm track impact parameters and secondary vertices information is used to construct a discriminant for the b-tagging.

For the analyses one usually operates with three main minimum thresholds for the b-tagging discriminants determined in all of the methods: loose ("L"), medium ("M") and tight ("T") working points. For this work the CSV b-tagging algorithm is used. The corresponding minimum thresholds on the CSV discriminant values are 0.244 (L), 0.679 (M) and 0.898 (T) [100]. They correspond to misidentification[8] probabilities of 10%, 1% and 0.1% respectively. The b-tagging efficiency for the CSVL and CSVM is above 80%, while for the CSVT it is of the order of 60% [100].

[6] The *impact parameter* of the track is the distance of minimal approach of the track to the primary vertex.

[7] The *significance* of the impact parameter d_0 is defined as a ratio of d_0 over its measurement uncertainty σ_{d_0}.

[8] Misidentification of jets means tagging a true u-, d-, s- or gluon-jet as a b-jet.

Jet Energy Correction

The energy of every reconstructed jet has to be corrected for detector and systematic effects. The jet energy correction is a multiplicative correction [101]:

$$p_{corr} = C \cdot p_{raw}. \tag{5.3}$$

Here p_{corr} denotes the corrected jet four-momentum, p_{raw} is an input uncorrected four-momentum and C is the correction factor. This correction factor is a product of different corrections:

$$C = C^{offset}_{L1-FastJet} \cdot C^{MC} \cdot C^{rel}_{L2} \cdot C^{abs}_{L3} \cdot C^{res}_{L2L3}. \tag{5.4}$$

First, a general level 1 FastJet offset correction, $C^{offset}_{L1-FastJet}$, is determined and applied. It aims to correct for the additional energy in a jet which occurs from the noise and pileup. This correction is applied on both jets from reconstructed MC and real data.

The Monte Carlo correction (C^{MC}) is performed on top of the offset correction for the MC jets. Reconstructed jets (the reconstruction procedure is the same as for the data) are corrected towards the generated jets momenta in different bins of p_T and η.

The level 2 relative jet energy correction (C^{rel}_{L2}) is applied on the data jets only on top of the offset correction. It aims to produce a flat response as a function of η. Additionally, the level 3 absolute jet (C^{abs}_{L3}) energy correction is applied on the data jets to ensure flat response in p_T.

The L2 and L3 correction factors are applied on the data, but they are determined from MC. The additional corrections on the difference between data and MC behavior are called residual L2L3 corrections, C^{res}_{L2L3} and are applied on the data jets on top of level 2 and level 3 corrections.

5.4.4 Missing Transverse Energy Reconstruction

The MET was already introduced in Sec. 3.2.3. The $\not{E}_T$ may be a hint to the undetected particles – like the two neutrinos present in the $t\bar{t}$ dileptonic decay channel (see Sec. 2.5.2). Based on the MET definition, one can write down its determination:

$$\sum_{detected\ objects} \vec{p}_T + \sum_{undetected\ objects} \vec{p}_T = 0, \tag{5.5}$$

$$\not{E}_T = \sum_{undetected\ objects} \vec{p}_T = - \sum_{detected\ objects} \vec{p}_T. \tag{5.6}$$

The MET reconstruction is obviously the last step of the PF chain, as it uses information about all the reconstructed objects.

Additionally, the pileup corrections based on a trained Boosted Decision Tree (BDT) and recoil corrections are applied to the reconstructed value of the MET [102, 103]. This improves the resolution of the reconstructed $\not{E}_T$.

6 | Event Selection

The $pp(gg) \to t\bar{t} \to l^-\bar{\nu}\bar{b}\, l^+\nu b$ process is occurring only in a tiny fraction of events recorded by the CMS detector. Additionally, other physical processes with a similar signature occur. These kinds of processes are called *background processes* for the $t\bar{t}$ production in the dileptonic final state. The selection criteria described in this section are aiming to select the events, in which the $t\bar{t} \to l^-\bar{\nu}\bar{b}\, l^+\nu b$ process (*signal*) is present and to reduce the possibility to select the events with background.

The selection mostly follows the procedure described in [104]. However, some restrictions were optimised for the boosted top analysis presented in this thesis.

The selection criteria described in this chapter are applied on both, data and reconstructed level MC[1].

6.1 Selection of Good Runs

The CMS data sets exploited in this work are listed in the Table 6.1:

Samples	Events	Run Range
/DoubleElectron/Run2012A-22Jan2013-v1/AOD	13M	190456 - 193621
/DoubleElectron/Run2012B-22Jan2013-v1/AOD	23.5M	193834 - 196531
/DoubleElectron/Run2012C-22Jan2013-v1/AOD	39M	198022 - 203742
/DoubleElectron/Run2012D-22Jan2013-v1/AOD	34.5M	203777 - 208686
/DoubleMu/Run2012A-22Jan2013-v1/AOD	5.6M	190456 - 193621
/DoubleMuParked/Run2012B-22Jan2013-v1/AOD	29M	193834 - 196531
/DoubleMuParked/Run2012C-22Jan2013-v1/AOD	37M	198022 - 203742
/DoubleMuParked/Run2012D-22Jan2013-v1/AOD	38M	203777 - 208686
/MuEG/Run2012A-22Jan2013-v1/AOD	2.5M	190456 - 193621
/MuEG/Run2012B-22Jan2013-v1/AOD	15M	193834 - 196531
/MuEG/Run2012C-22Jan2013-v1/AOD	21M	198022 - 203742
/MuEG/Run2012D-22Jan2013-v1/AOD	22M	203777 - 208686

Not all the events recorded by the detector have a high quality - the detector might have been performing not as expected [105]. Only the events which are present in the list of certified good LHC runs [106] were used for the analysis. The resulting luminosity of the samples accepted for the analysis is 19.7 fb^{-1}.

[1] The simulated events undergo the same reconstruction criteria as the real data recorded by the detector.

6.2 Trigger Selection

Each event recorded by the CMS detector was selected by one or more triggers (see Sec. 3.2.6). As in this analysis one searches for a dileptonic final state, the events which were selected by the dilepton triggers are accepted. These triggers require the presence of at least two leptons (electrons or muons) in the event with a minimum transverse momenta of 17 GeV for muons and 8 GeV for electrons. The efficiencies of the triggers as function of lepton momenta are presented in Fig. 6.1. The plots show that the efficiency plateau is stable with $p_T(l) > 20$ GeV.

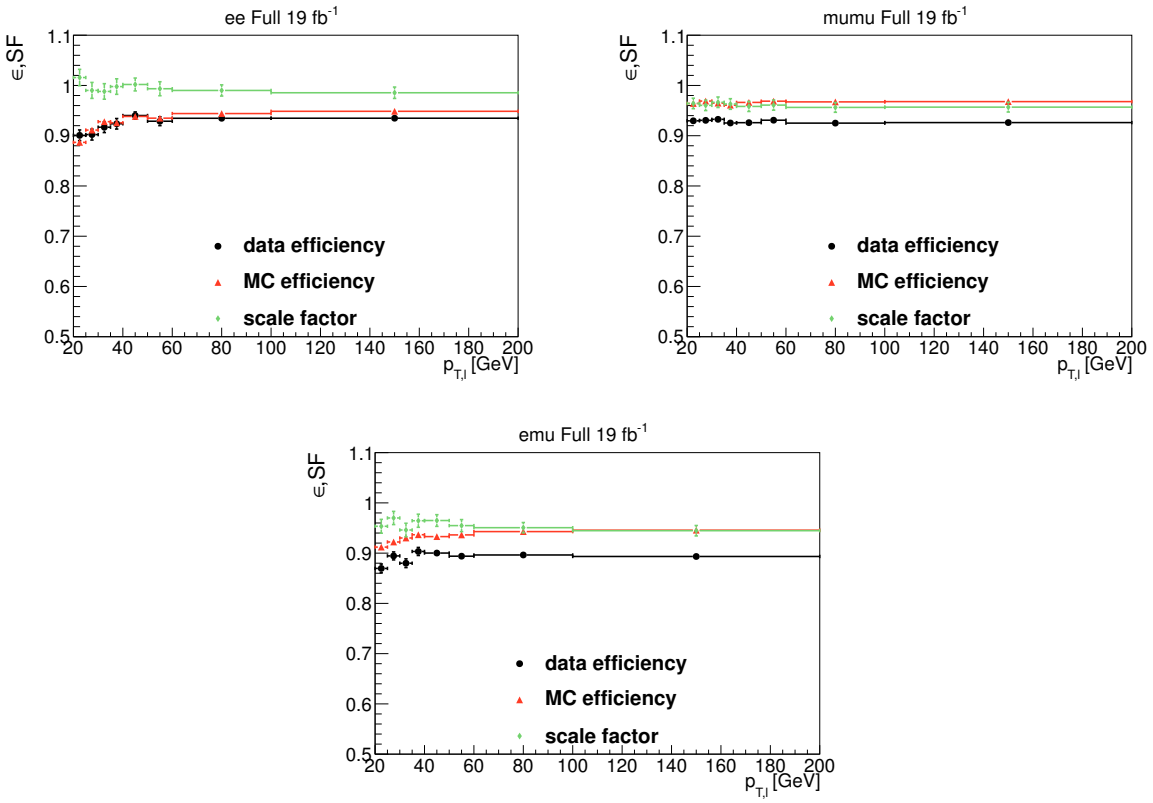

Figure 6.1: Dielectron trigger efficiency as a function of lepton p_T in the ee (top left), $\mu\mu$ (top right) and $e\mu$ (bottom) channels. Efficiencies in data (black circles), MC (red triangles) and their ratio (green diamonds), which serves as a correction scale factors for the MC are presented. The error bars on efficiencies include the statistical uncertainties only, while the scale factors include both, statistical and systematic uncertainties. The plots are taken from [107].

As there is no data before trigger selection recorded by CMS, the usual definition of efficiency[2] for the trigger selection doesn't work. The efficiency for one trigger can be only determined relatively to some other trigger. The reference trigger is usually chosen to be as much as possible independent from the tested trigger. The efficiency is then determined the following way:

[2] The *efficiency* of some selection criteria or procedure $\mathcal{R}$ is the ratio of the number of events left after $\mathcal{R}$ over the number of events entering $\mathcal{R}$: $\epsilon = \frac{N_{after\ \mathcal{R}}}{N_{before\ \mathcal{R}}}$

$$\epsilon^{Trigger\ under\ Test} = \frac{N^{Trigger\ under\ Test \& Reference\ Trigger}}{N^{Reference\ Trigger}}. \tag{6.1}$$

Here $N^{Trigger\ under\ Test \& Reference\ Trigger}$ is the number of events selected simultaneously by both, trigger under test and reference trigger, and $N^{Reference\ Trigger}$ denotes the number of events selected by the reference trigger. In this analysis the data sample selected by a MET based trigger was used as a reference [104].

The MC events are corrected for differences in the efficiency of the trigger selection in simulation and in data. The correction factors are called the **S**cale **F**actors (SF) and are determined as SF $= \frac{\epsilon(MC)}{\epsilon(data)}$. The SFs for the trigger selection are determined in bins of the pseudorapidity of the two selected leptons [104].

6.3 Event Cleaning

The events with a clean signature are selected for the analysis. The following was required:

- **Beam Scrapping**: accept only events with not more than 10 reconstructed tracks, or events, where at least a quarter of the tracks has a high quality. This helps to remove the events with a high fraction of background from the beam - the events contain tracks from the beam remnants.
- **High quality primary vertex**: accept the events where at least one primary vertex is reconstructed. Additionally, the vertex has to be positioned in the central detector region ($|z| < 24$cm) and the radial coordinate r counted from the center of the CMS coordinate system (see Sec. 3.2) should be less than 2 cm.
- **HCAL Noise Removal**: accept only the events without anomalous HCAL noise [108].

6.4 Pileup Removal

It is very important to analyze only the objects related to the hard process one aims to study and not to take some extra particles from the pileup vertices. Only the objects arising from the hardest primary vertex[3] are considered for the further analysis. This criterion provides in most of the cases a vertex with hard interaction, like $t\bar{t}$ production.

A pileup correction is applied event-by-event on the MC level. This is a weight correction which is aiming to match the MC distribution of the primary vertex multiplicity to the data distribution. For this correction the primary vertex multiplicity probability density is determined as a product of the estimated luminosity [109] and total pp inelastic cross section σ_{pp} [110]. The vertex multiplicity distributions with and without applying the scale factors are shown in Fig. 6.2. The figure shows a tangible improvement of the MC description of the data after applying the scale factors.

[3] The hardest primary vertex is taken as the one with the largest sum of transverse momenta squared (p_T^2) of all the charged particles emerging from it.

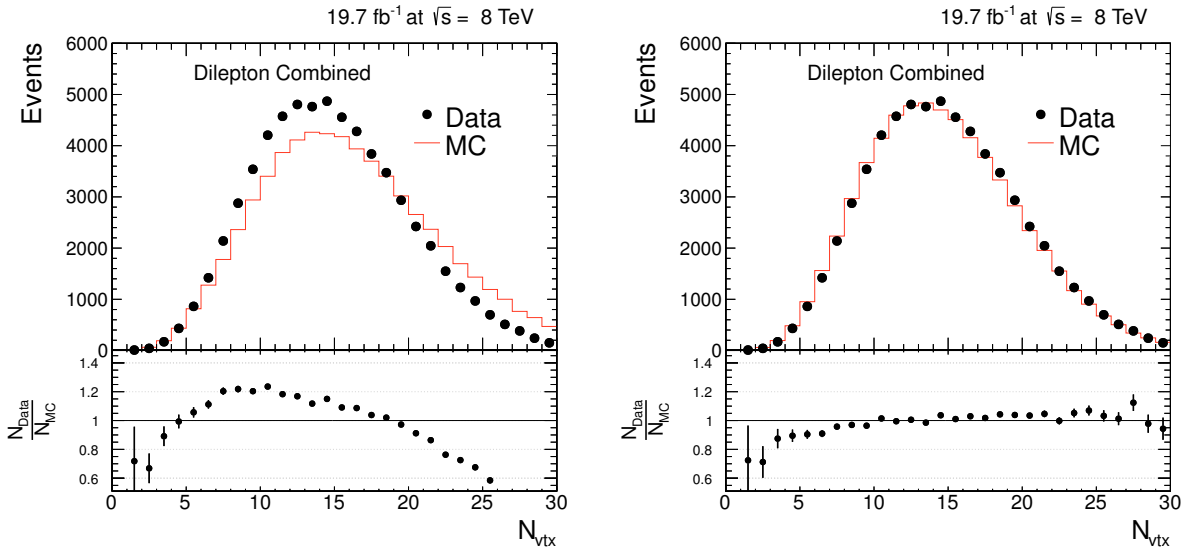

Figure 6.2: Control distribution of the number of primary vertices before applying the pileup scale factors (left) and after applying them (right). Black dots show the data points and red line represents the MC. The ratio of data over simulation is presented in the bottom panel of each distribution. The error bars represent the statistical uncertainties of the data.

6.5 Background Processes

As already discussed in this chapter, the background processes can be wrongly treated as the signal process. The yields of background production are estimated using the MC simulation [35]. This is done because the background level in the $t\bar{t}$ dileptonic final state is expected to be very small. To make the simulated background yields more reliable, they are scaled to the total cross section predicted in higher order calculations [111].

The contribution of the following sources of backgrounds are considered as non-negligible in this work:

- $t\bar{t} \rightarrow$ *other* **production**: production of the $t\bar{t}$ pair, which decays in a different decay channel (see Sec. 2.5.2), or also dileptonically, but through τ leptons[4]. The latter can further decay to electrons or muons plus neutrinos, thus they can be very likely mistreated as a signal event.

- **Single top production**: events with a single top produced in association with a W-boson (see Sec. 2.5.1) can be mistreated as a $t\bar{t}$, as the associated W may be considered as the one which arises from the second top decay. The sample for this background was simulated in POWHEG + PYTHIA.

- $Z/\gamma^* \rightarrow \tau^+\tau^-$, $Z/\gamma^* \rightarrow e^+e^-/\mu^+\mu^-$, **or the Drell-Yan production**: the signature of this process is the presence of two leptons of the same flavour in the final state. The $t\bar{t}$, which decays to two electrons or two muons has the same signature. That

[4]The dileptonic $t\bar{t}$ decay via intermediate τ leptons is considered as a background for this analysis.

is why the Drell-Yan process with associated jet produced can be mistreated as a $t\bar{t}$ signal. The Drell-Yan samples are generated using PYTHIA. These samples are normalized to the data with data driven factors [112].

- **Diboson production**: ZZ, WW or WZ can also have a dileptonic final state and thus, if produced in the association with additional jets, can be wrongly reconstructed as signal. The corresponding background samples are generated in PYTHIA.

- **Associated $t\bar{t}$ and $W/Z/\gamma$ production**: the bosons decay leptonically and can be combined with the jets from the $t\bar{t}$ decay. This can mimic a dileptonic final state with two jets. This background sample is generated utilizing MADGRAPH+PYTHIA.

- W/Z + **jets production**. Only one genuine lepton is produced in this process from the boson decay. This background is simulated in MADGRAPH + PYTHIA.

- **QCD multijet production**: for this process zero prompt leptons are produced. However, their production rate at the LHC is high enough, so that there will be cases when two leptons can be associated to these processes. The same applies to the W/Z + jets production process. The additional leptons can either occur from hadron decays or some hadrons may be occasionally misidentified as leptons. The QCD background is generated using the PYTHIA event generator.

The simulated background samples are finally normalized to the luminosity of the data sample used for the analysis (19.7 fb^{-1}).

The selection requirements which will be further discussed in this chapter are aiming to suppress different background contributions to the final yield of events.

6.6 Lepton Selection

6.6.1 Lepton Isolation

This analysis aims to measure the cross sections in the boosted regime in which the momenta of t-quarks are large. In Fig. 6.3 schematic decays of a top quark with low and high momentum are presented. An example of an event from signal simulation with one boosted top out of the $t\bar{t}$ pair is presented in Fiq. 6.4. As one can see, the decay products of the boosted top are geometrically much closer than for the unboosted case and it is not that easy to isolate the lepton from the jet. In the hadronic decay channel the decay products of a boosted t-quark are not well separated from each other and reconstructed as one top-jet using a *fat jet* technique [113]. This analysis aims to reconstruct all the top decay products separately.

One would be tempted not to use any lepton isolation for the reconstruction of the decay products of a boosted top quark. However, the lepton isolation is important to reject leptons from hadronic decays inside a jet. This largely reduces the fraction of QCD background in the selected events. As the QCD multijet production in the events from the LHC collisions is very likely to happen, one should apply some isolation criterion for the leptons not to get overwhelmed by the QCD background.

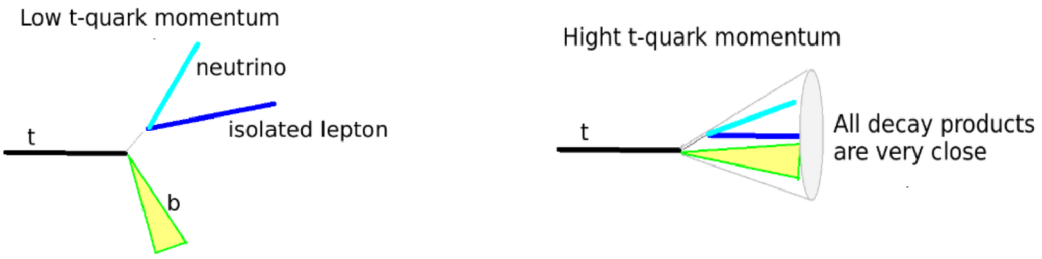

Figure 6.3: Illustration of the t-quark decay in unboosted (left) and boosted (right) regimes.

In the other CMS analyses, which measure the $t\bar{t}$ production cross sections differentially [35] and double differentially [9] in the dilepton channel the *relative isolation* is used. This isolation requires that for each lepton a condition $I_{\text{rel}} \leq 0.15$ in a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.3$ around the lepton track is fulfilled. The relative isolation I_{rel} is defined as the sum of transverse energy deposits from charged and neutral hadrons and photons, relative to the transverse momentum of the lepton.

For this analysis a very loose *2D isolation* criteria for leptons is used. The criteria considered for the 2D isolation are:

- the lepton has a so called relative transverse (or the momentum projection on the closest in (η, ϕ) jet axis in the event) $p_T^{rel}(l,\ any\,jet) > 15$ GeV. p_T^{rel} is a Lorentz invariant variable and thus it is largely independent of the boost. The p_T^{rel} provides a nice separation between signal and QCD background events, as (due to the large mass of the top quark) it is usually much larger for the lepton and jet produced from a top quark decay than for the background objects;

- the lepton has a cone separation to the closest jet in the event $\Delta R(l,\ any\,jet) > 0.5$ where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

For both criteria only to the jets with a minimum transverse momentum of 30 GeV are considered.

Only if a lepton fulfills at least one of the two criteria, it is accepted for the further analysis.

The studies of the performance of the 2D isolation are presented in Sec. 8.1.

6.6.2 Lepton and Lepton Pair Selection

It is required that at least two opposite-signed leptons are present in the event after isolation. Moreover, these leptons must have transverse momentum of at least 20 GeV (to avoid QCD background) and $|\eta| < 2.4$.

In case there are more then two opposite signed leptons in the event, the two opposite signed leptons with the highest transverse momenta are chosen for the further analysis. At this point the event is assigned to one of the three decay channels of the $t\bar{t}$: ee, $\mu\mu$ or $e\mu$.

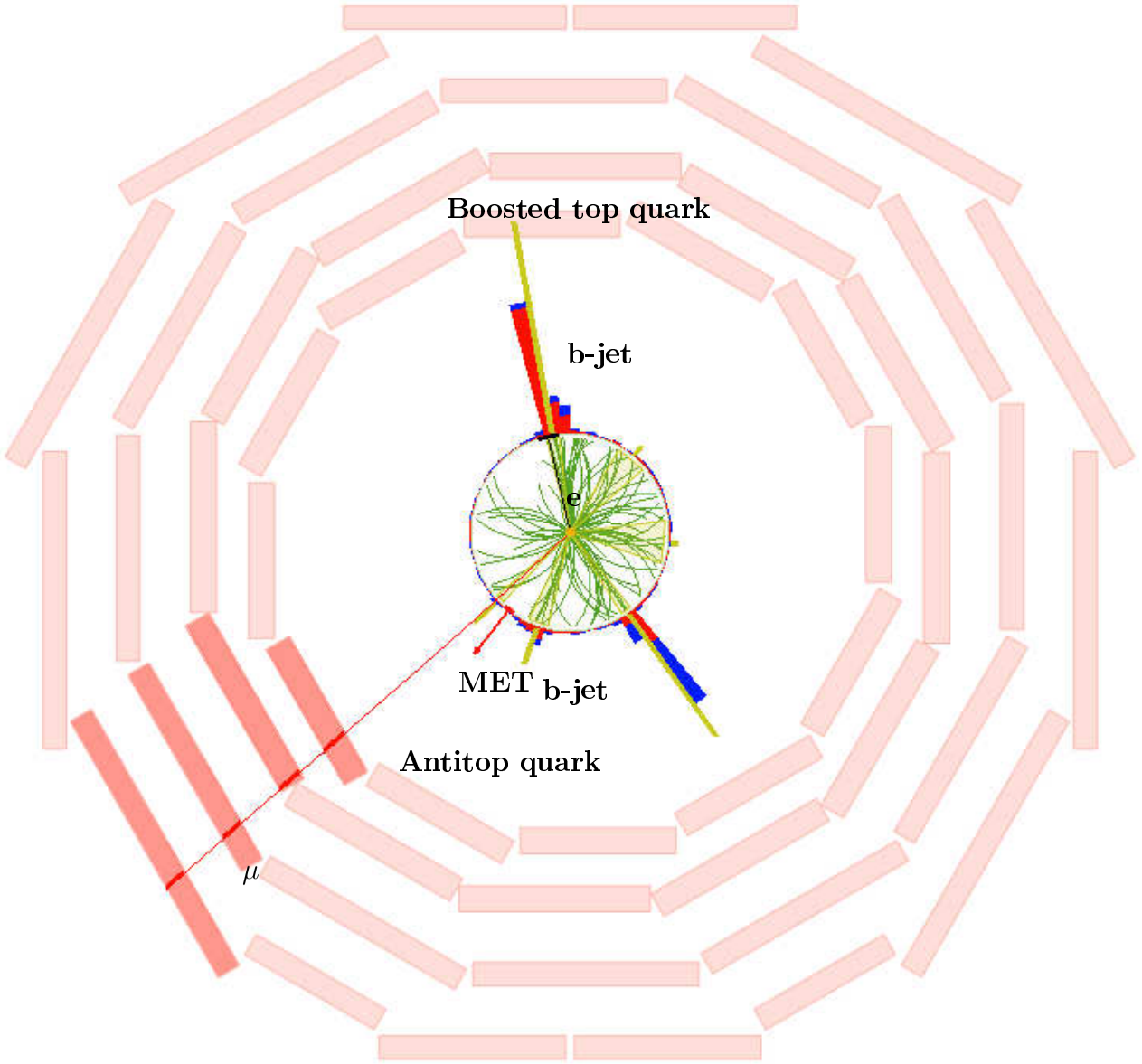

Figure 6.4: CMS event display: radial view of a fully simulated event where a $t\bar{t}$ pair with one boosted top was produced. Here the boosted top decays to an electron (black track inside the tracker) and a high energy b-jet (thick green line). The second top from the $t\bar{t}$ pair decays to a muon (red line) and a b-jet. Neutrinos contribute to the missing transverse energy vector (marked with a red arrow).

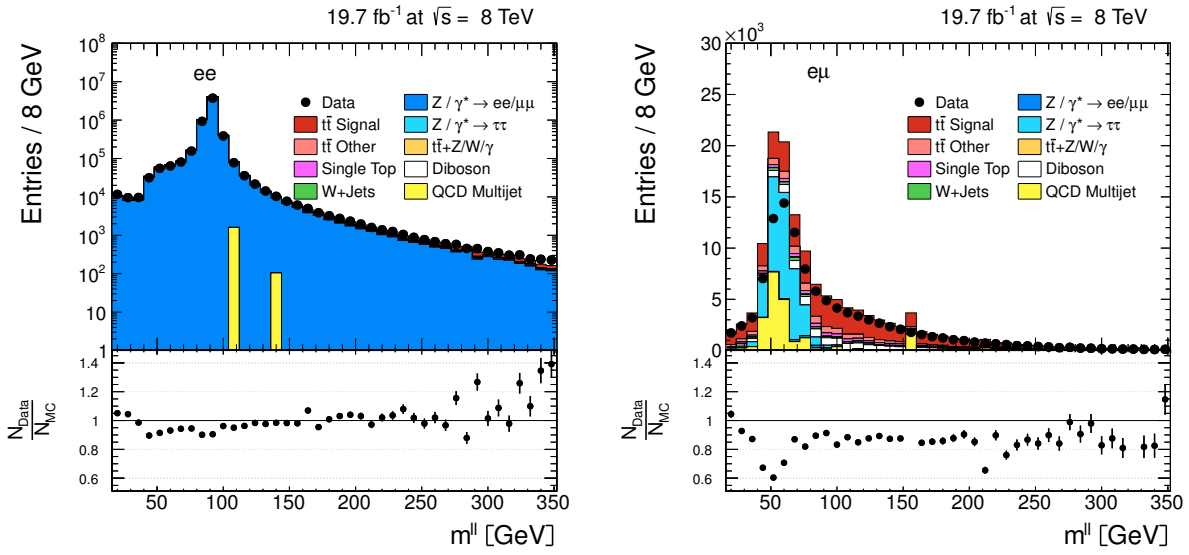

Figure 6.5: Control distribution of the dilepton mass before applying any $m(ll)$ restriction in the ee (left) and $e\mu$ (right) decay channels. Black dots show the data points and filled histograms represent the MC. Histograms with different colors represent contributions from different background processes. The ratio of data over simulation is presented in the bottom panel of each distribution. The error bars represent the statistical uncertainties of the data.

To reduce the fraction of the $Z/\gamma^* \to ee/\mu\mu$ Drell-Yan background events, a restriction on $m(ll)$ is applied. The minimum threshold on the $m(ll)$ is set to 20 GeV.

Additionally, events with 76 GeV $\leq m(ll) \leq$ 106 GeV, which corresponds to the region around the Z-resonance mass, are rejected. As the ee and $\mu\mu$ $t\bar{t}$ decay channels are much stronger affected by the Drell-Yan background compared to the $e\mu$ channel (see the control distributions[5] in Fig. 6.5), the restriction on the dilepton mass is applied only for the same flavour channels. The poor description of the data by Monte Carlo observed in this plot is fixed in the further selection steps.

6.7 Jet Selection

All the jets in the event are required to have transverse momentum of at least 30 GeV and $|\eta| < 2.4$.

It is required that every event selected for the analysis contains at least two jets and at least one of the jets has to be b-tagged. Jets which fulfills the CSVL criterion (see Sec. 5.4.3) are assumed to be b-tagged in this analysis. Fig. 6.6 shows the jet and b-jet multiplicity and the threshold values. It shows that these threshold requirements are efficiently reducing the number of events with QCD multijet and Drell-Yan production.

[5] The **control distribution** contains data and MC distributions of a certain variable on one plot to see the behavior of both and to compare them. For this analysis, the simulated background sources are also plotted on the control distributions, which allows to see their expected fraction in the final selected sample.

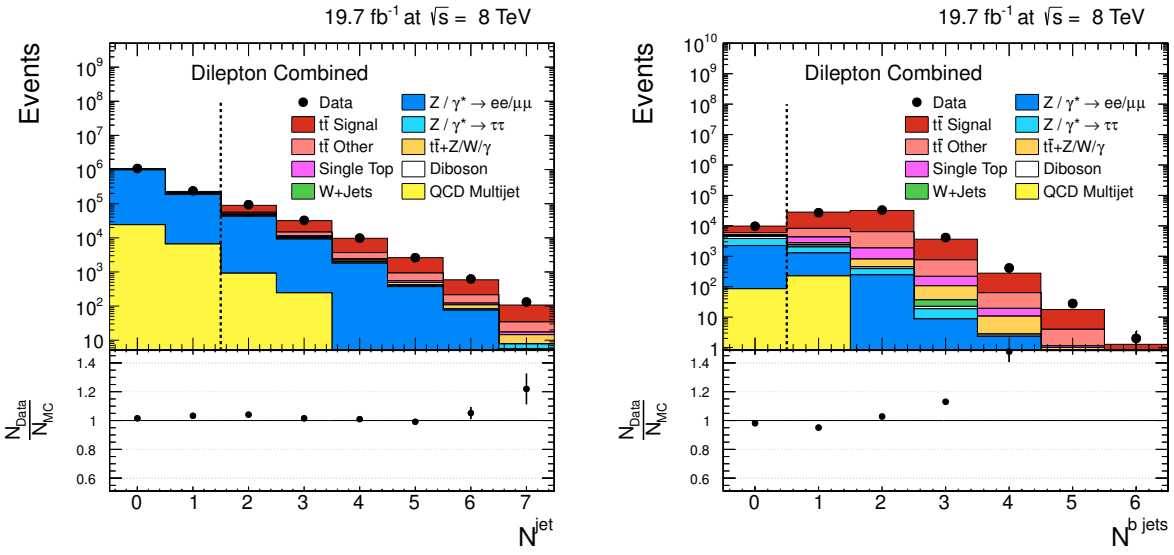

Figure 6.6: Control distribution of jet multiplicity (left) and b-jet multiplicity (right) for the combination of all three dacay channels. The distribution of the jet multiplicity is presented after applying trigger selection, event cleaning, pileup removal and lepton selection. The distribution of the b-jet multiplicity is presented after the same selection steps and MET cut, which is described in Sec. 6.8. The vertical lines represent the minimum threshold applied in the analysis. Other details as in Fig. 6.5.

The scale factors related to the b-tagging procedure are applied for MC events. The efficiency of the CSV tagging is estimated in a $Z \to b\bar{b}$ data sample using the tag-and-probe method for data and in a QCD-multijet simulation sample for MC [100]. The determination of the scale factors was following the procedure described in [114]. The efficiencies and scale factors are determined for different jet flavours in bins of the jet p_T and η.

6.8 MET Selection

To reduce the number of the Drell-Yan background events, a requirement on the minimum value of the missing transverse energy of 40 GeV for the ee and $\mu\mu$ channels was applied. The MET spectrum before applying the cut is shown in Fig. 6.7. The plot shows that the minimum threshold applied for MET is effectively rejecting a large amount of the events with Drell-Yan production. The effect of this restriction is shown in Fig. 6.8 for the ee channel.

6.9 Selection Summary

To display the result of the selection described above, different control distributions are plotted and analyzed.

The control distributions of the dilepton mass, lepton pseudorapidity and lepton p_T are

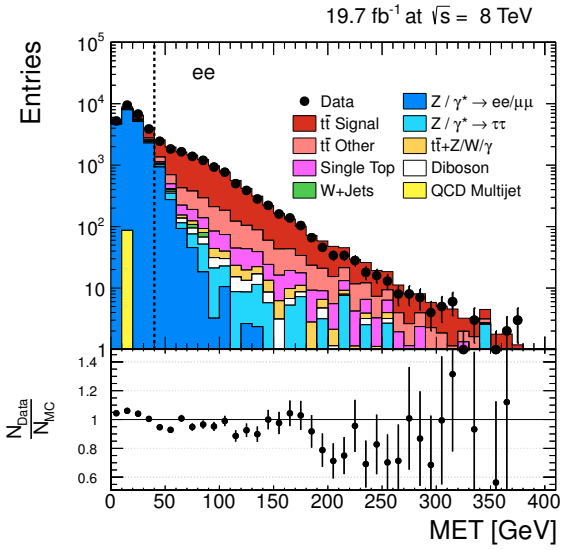

Figure 6.7: Control distribution of the missing transverse energy in the ee channel before MET requirement. The vertical dashed line represents the minimum threshold required for the MET in the ee and $\mu\mu$ channels in this analysis. Other details as in Fig. 6.5.

presented in Fig. 6.9 and the control distributions of the jet multiplicity, jet pseudorapidity and jet p_T are shown in Fig. 6.10. From these plots one can conclude an overall reasonable agreement of data and simulation. The $t\bar{t}$ signal fraction is highly dominating over the background contributions. All the control plots, mentioned above, are presented for the combination (sum) of all the three top decay channels - ee, $e\mu$ and $\mu\mu$.

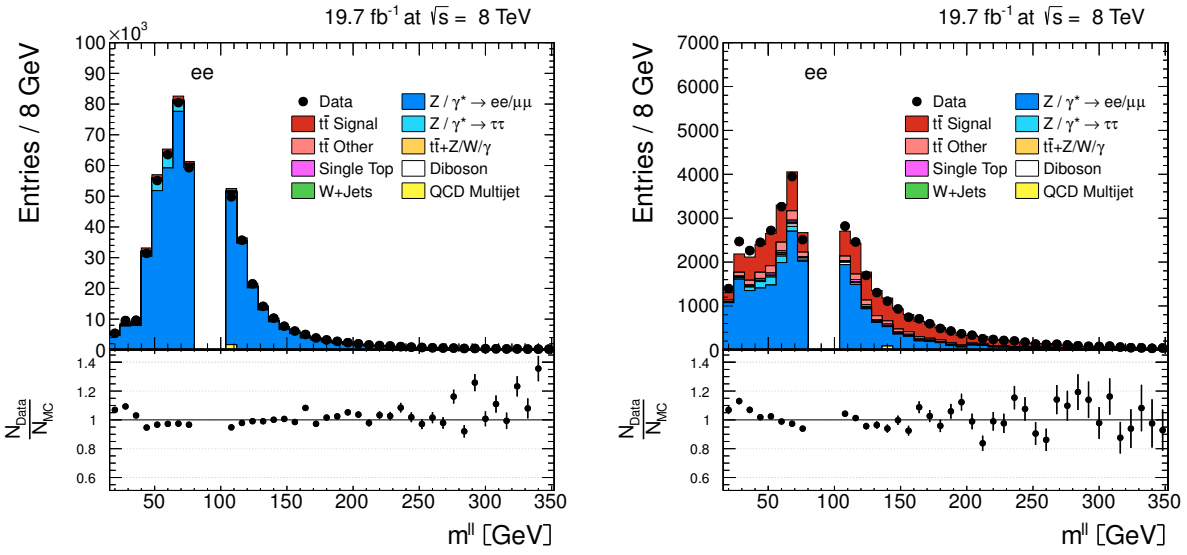

Figure 6.8: Control distribution of the dilepton mass in the *ee* channel before (left) and after (right) the minimum MET requirement. Other details as in Fig. 6.5.

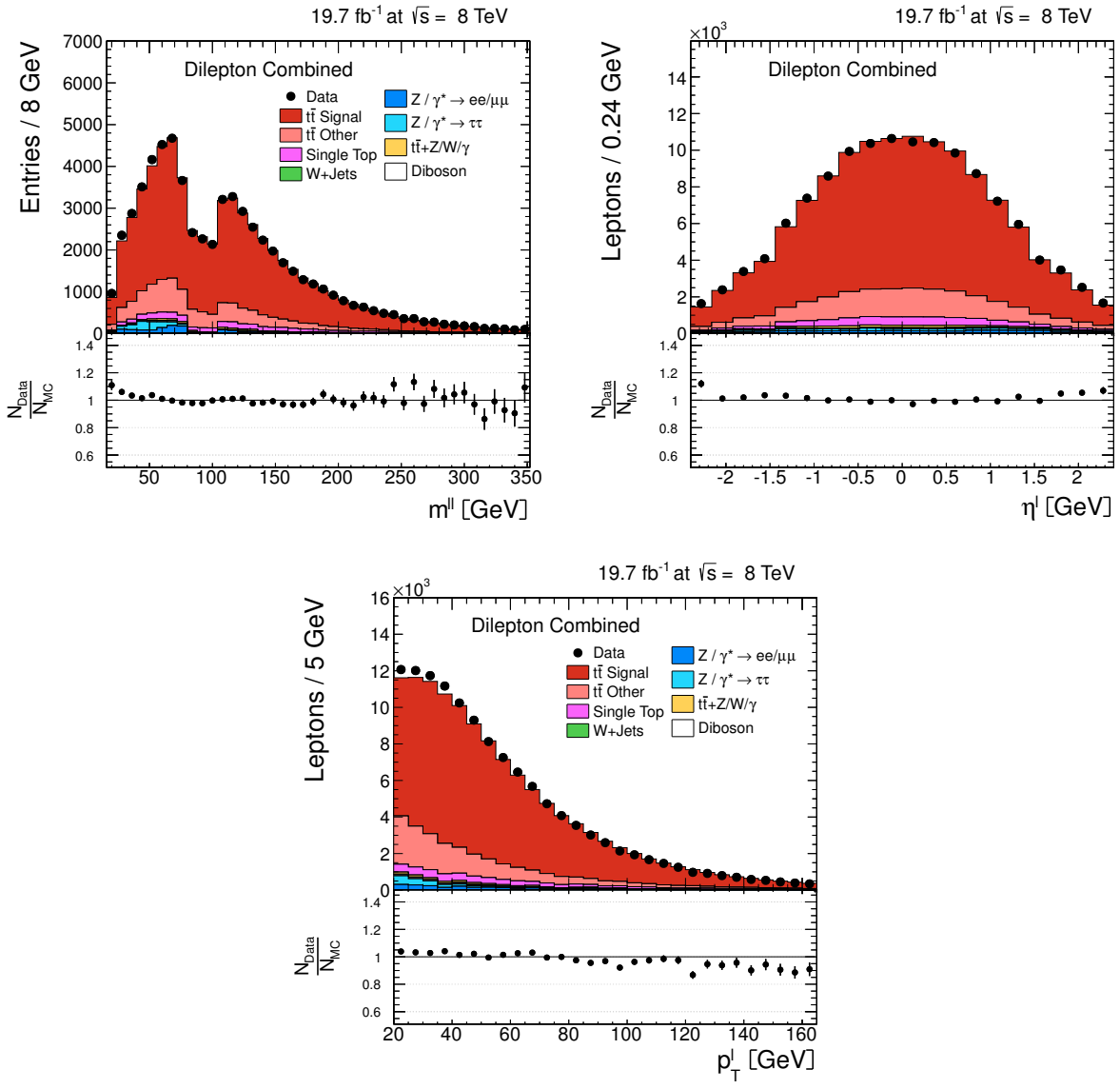

Figure 6.9: Control distribution of the dilepton mass (top left), lepton pseudorapidity (top right) and lepton p_T (bottom) after applying all the selection criteria. Other details as in Fig. 6.5.

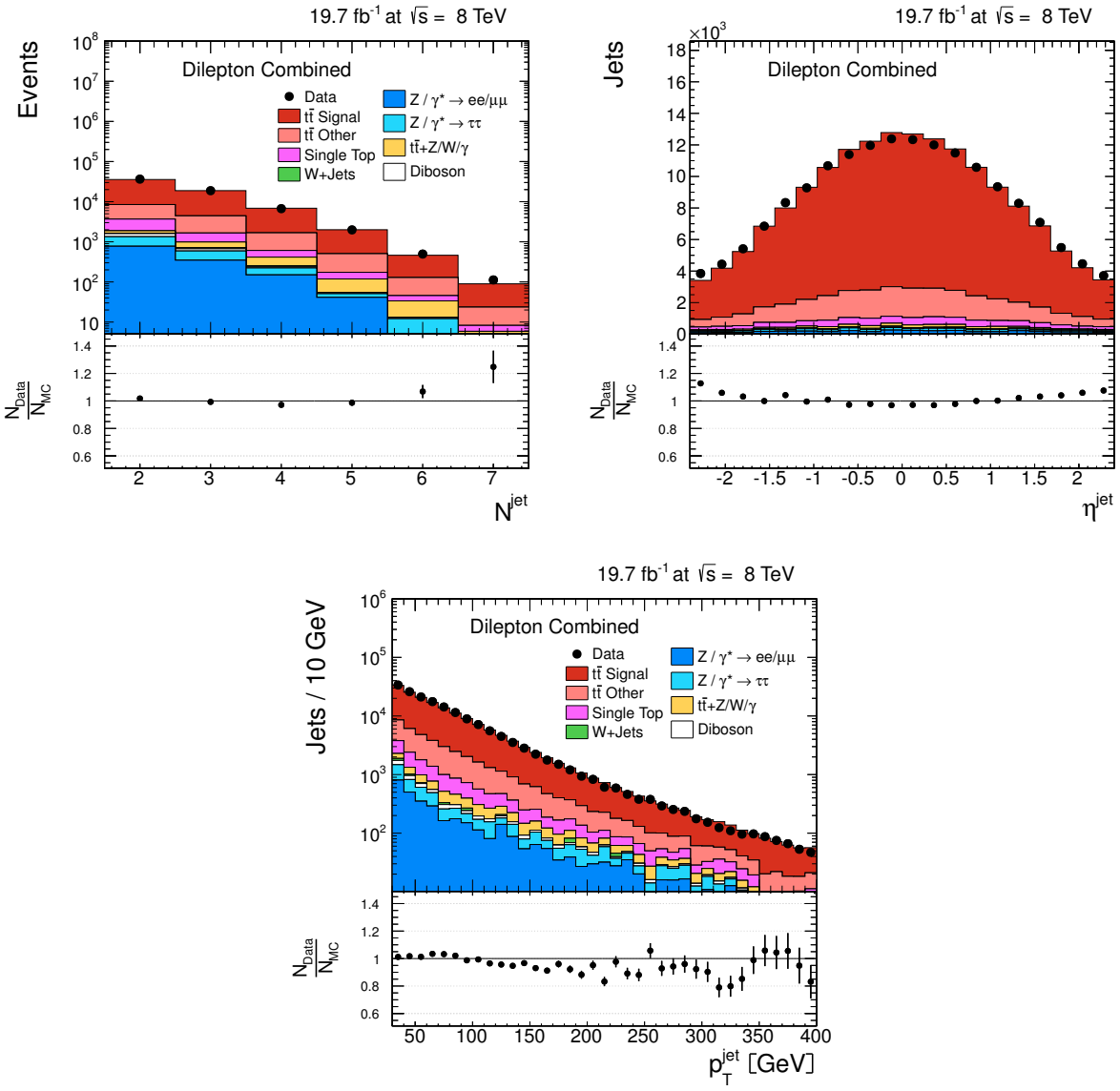

Figure 6.10: Control distribution of the jet multiplicity (top left), jet pseudorapidity (top right) and jet p_T (bottom) after applying all the selection criteria. Other details as in Fig. 6.5.

7 | Kinematic Reconstruction of the $t\bar{t}$ System

The selection described in Chapter 6 provides a data sample with large amount of events with $t\bar{t} \to l\bar{\nu}\bar{b}\ \bar{l}\nu b$. The decay diagram of the studied decay chain is presented in Fig. 1.1. This would be enough to calculate the total cross section of this process or to make differential measurement in bins of lepton and jet kinematics. However, there is the full information about the kinematics of the top quarks is not directly available. The reason for this indeterminacy is the presence of two neutrinos in the final state of the $t\bar{t}$ dilepton decay. Neutrinos do not interact with the detector material and thus can't be identified directly by any means. Two neutrinos are the missing information to complete the final state and extract the top-quark kinematics.

The kinematic reconstruction is an algorithm which allows to recover the full kinematics of the $t\bar{t}$ system by means of making physical assumptions and applying corresponding constraints. In this chapter an algorithm for reconstruction of the $t\bar{t}$ in dilepton final state is described (see also [9]). The chapter includes the description of the main assumptions, algorithm details and studies of the performance.

7.1 Kinematic Constraints

Two undetected neutrinos in the $t\bar{t}$ final state provide six unknowns to the system (assuming the neutrino mass zero) – three momentum components for each of the neutrinos. To make the system solvable, six kinematic constraints are introduced:

- **Top mass constraints** (2 constraints). It is assumed that $m(t) = m(\bar{t}) = 172.5$ GeV (this number is taken from the Particle Data Group particle listings [17]).

- **Constraints on the transverse momentum of the neutrino** (2 constraints). It is assumed that all the missing transverse energy of the event is originating only from the two neutrinos from the W decays in the $t\bar{t}$ decay chain.

- **Constraints on the $W^{\pm}$ mass** (2 constraints). The masses of W^+ and W^- are assumed to be equal. The actual value of the $W^{\pm}$ mass is taken from the generated mass (will be discussed later in Sec. 7.3).

These constraints result in an algebraic system of six equations with six unknowns:

$$m_t^2 = (E_b + E_{\bar{l}} + E_\nu)^2 - (p_{b_x} + p_{\bar{l}_x} + p_{\nu_x})^2 - (p_{b_y} + p_{\bar{l}_y} + p_{\nu_y})^2 - (p_{b_z} + p_{\bar{l}_z} + p_{\nu_z})^2, \tag{7.1}$$

$$m_t^2 = (E_{\bar{b}} + E_l + E_{\bar{\nu}})^2 - (p_{\bar{b}_x} + p_{l_x} + p_{\bar{\nu}_x})^2 - (p_{\bar{b}_y} + p_{l_y} + p_{\bar{\nu}_y})^2 - (p_{\bar{b}_z} + p_{l_z} + p_{\bar{\nu}_z})^2, \tag{7.2}$$

$$\not{E}_x = p_{\nu_x} + p_{\bar{\nu}_x}, \tag{7.3}$$

$$\not{E}_y = p_{\nu_y} + p_{\bar{\nu}_y}, \tag{7.4}$$

$$m_{W^+}^2 = (E_{\bar{l}} + E_\nu)^2 - (p_{\bar{l}_x} + p_{\nu_x})^2 - (p_{\bar{l}_y} + p_{\nu_y})^2 - (p_{\bar{l}_z} + p_{\nu_z})^2, \tag{7.5}$$

$$m_{W^-}^2 = (E_l + E_{\bar{\nu}})^2 - (p_{l_x} + p_{\bar{\nu}_x})^2 - (p_{l_y} + p_{\bar{\nu}_y})^2 - (p_{l_z} + p_{\bar{\nu}_z})^2, \tag{7.6}$$

where

$$E_{\nu,\bar{\nu}}^2 = p_{\nu,\bar{\nu}}^2 = p_{\nu_x,\bar{\nu}_x}^2 + p_{\nu_y,\bar{\nu}_y}^2 + p_{\nu_z,\bar{\nu}_z}^2. \tag{7.7}$$

All the momenta and energies of the leptons and jets are reconstructed from the detector signals (see Chapter 5).

7.2 Solution of the Kinematic Equations

The system of equations 7.1-7.6 can be solved algebraically. The solution followed here was proposed by L. Sonnenschein in [115]. After several transformation steps, the kinematic equations 7.1-7.6 result in a fourth order polynomial equation for one of the neutrino (or antineutrino) momenta components. Following the solution algorithm from [115], the equation is obtained for the neutrino x momentum component:

$$h_4 p_{\nu_x}^4 + h_3 p_{\nu_x}^3 + h_2 p_{\nu_x}^2 + h_1 p_{\nu_x} + h_0. \tag{7.8}$$

The coefficients h_i from the Eq. 7.8 are expressed in terms of kinematic characteristics of the leptons and jets. The actual dependency can be found in [115, 116].

The method how to solve the quartic equation 7.8 is given in [115] and is followed in this work. The other neutrino and antineutrino momenta as well as the top and antitop kinematics can then be successively calculated.

The Eq. 7.8 can have up to four solutions. However, only one neutrino momentum in the $t\bar{t}$ dileptonic decay has to be reconstructed. Thus, one has to either combine the multiple solutions into one, or take only one single solution to reconstruct the $t\bar{t}$ system. In this work the single solution which provides the smallest $M(t\bar{t})$ is taken[1]. This criterion works for low and for high $p_T(t)$ regions [9]. As will be discussed later (see Sec. 7.4.1) there is a problem of a large number of event migrations from true low $p_T(t)$ to reconstructed high $p_T(t)$ and also from the true low $M(t\bar{t})$ to high $M(t\bar{t})$. However, it is clear that the choice of the minimal reconstructed $M(t\bar{t})$ for the solutions of the eq. 7.8 can not enhance this effect.

[1] The solution with the smallest $M(t\bar{t})$ was proven to be more often correct compared to the other solutions, the corresponding studies are presented in [9].

7.3 Treatment of the Detector Effects and Combinatorics

In the section above it was described how to get a single solution for the $t\bar{t}$ kinematics in case the leptons and jets are perfectly reconstructed and it is unambiguously known, which jets and which leptons are arising from top and antitop decays. In reality both of the conditions are not fulfilled.

First of all, the objects reconstructed from the detector signals are not perfect. This is related to detector imperfections and to possible features of the reconstruction algorithms and results in uncertainties for every kinematic characteristics of jets and leptons, as well as MET, reconstructed from the detector. However, the quartic equation 7.8 is very sensitive to the correct input (kinematics of jets and leptons and MET) and it is possible that no solution is found in case the input parameters are even slightly distorted.

To overcome this problem the whole kinematic reconstruction is repeated 100 times, each time altering simultaneously the energies and directions of jets and leptons within their uncertainties. This procedure will be further called *smearing*. For each of the 100 smearings, the energies of leptons and jets are scaled by a factor, which is determined from the simulated $t\bar{t}$ signal sample. This smearing factor is determined from the distribution of

$$f(E) = \frac{E^{true}}{E^{reco}}, \tag{7.9}$$

where E^{true} is the "true" energy of the jet or lepton, which means that this energy is taken on the generator level, and E^{reco} is the energy of the same jet or lepton on the reconstruction level.

The smearing factors $f(E)$ are each time taken randomly according to the distributions shown in Fig. 7.1. These distributions were determined using the events in which jets and leptons on the reconstruction level are matched to the generator level leptons and jets. The correction factors for leptons are determined not distinguishing electrons and muons, as the shapes of the distributions for electrons and muons do not differ much. The difference in the mean value of the distributions is less than a percent and the RMS for the distribution of muon energy correction factors is 0.047 and for electrons - 0.052.

The directional smearing is performed as shown in the sketch Fig. 7.2. The angle ω is taken randomly and the angle α is taken randomly according to the distribution of α (see Fig. 7.3). These distributions are determined in the signal MC as the angle between the true and the matched reconstructed jet or lepton direction. No distinction between electrons and mouns is made in the distributions, as the mean and the RMS of the distributions for muons and for electrons separately doesn't differ within less than a percent.

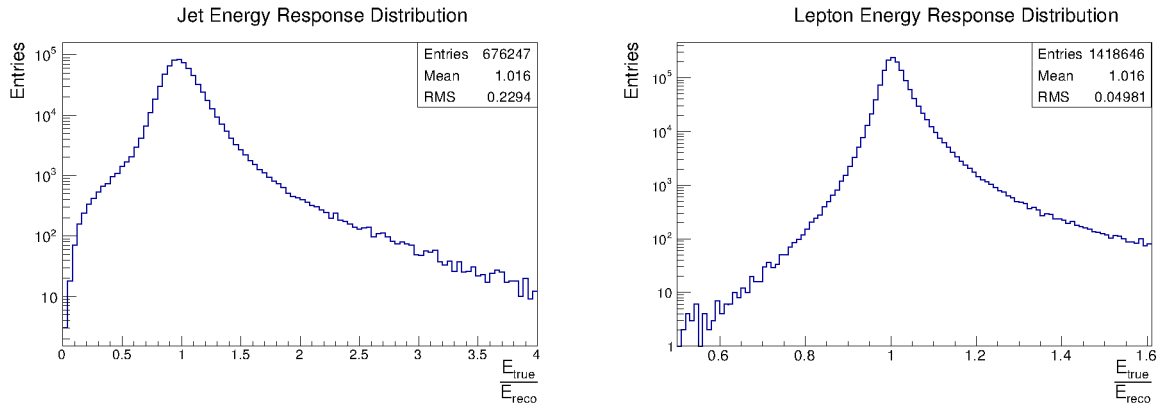

Figure 7.1: The distributions of the smearing factors $f(E)$ for the jets (left) and leptons, both electrons and muons (right) determined from the $t\bar{t}$ signal MC after applying selection as described in Chapter 6.

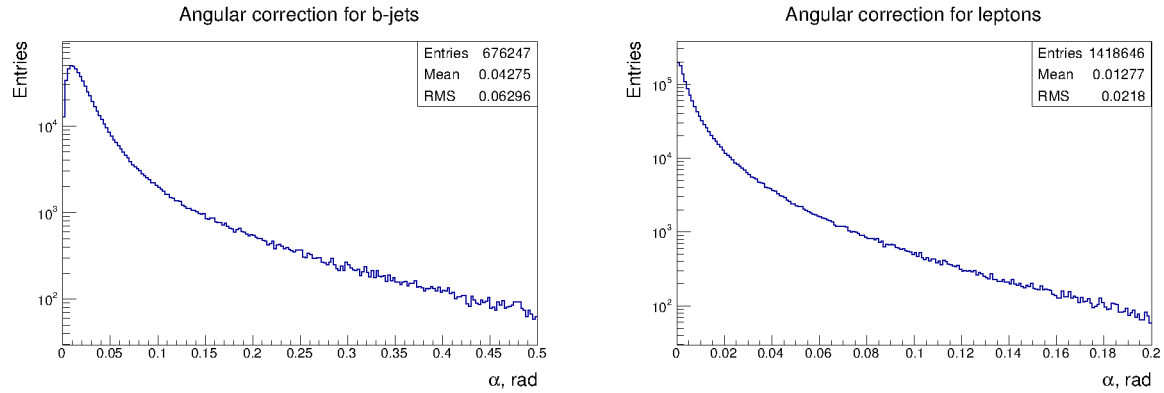

Figure 7.3: The distributions of the smearing angles α for the jets (left) and leptons (right).

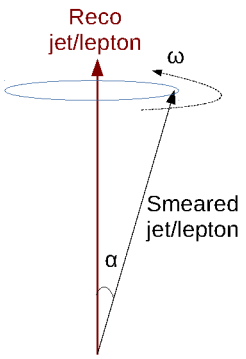

Figure 7.2: Sketch of the directional smearing applied for leptons and jets.

For each smearing the $W^{\pm}$ mass is randomly taken according to the generated distri-

bution of W mass[2].

The missing transverse energy is also recalculated in every smearing to preserve the transverse energy balance in the event [9].

The approach of smearing the characteristics of the objects reconstructed from the detector allows to increase the fraction of events in which a solution of the kinematic equation 7.8 is found. However, it provides another complication, and that is the presence of up to 100 different solutions for the $t\bar{t}$ kinematics in one event.

Not only smearing is increasing the multiplicity of different solutions for $t\bar{t}$ kinematics in one event. The other effect is *combinatorics*. If one had reconstructed a very pure event with a $t\bar{t}$ production with two leptons of different electric charge (l and $\bar{l}$) and two jets ($jet1$ and $jet2$) in the final state, one can not distinguish, which jet originated from the top-quark decay and which from the antitop. Thus, both, the combination ($jet1 - l$) and ($jet2 - l$) can be treated as the antitop candidate (the same with the $\bar{l}$ and the top candidate). However, in reality there are many events with more than two jets. In the general case, the combinatorics problem will give rise to $N_{jets}!$ $t\bar{t}$ candidates, where N_{jets} is the jet multiplicity.

The problem of appearing of up to $N_{jets}! \cdot 100$ different $t\bar{t}$ candidates is solved in two steps:

1. *Choice of single lepton-jet combination.* All the combinations of leptons and jets are tried out. For each of the combinations, the smearing procedure is applied. A weight according to the $m(lb)$ distribution (see Fig. 7.4) is assigned to every smearing of each lepton-jet combination, for which a physical solution of eq. 7.8 is found. Only the combination with the largest sum of weights over all the smearings ($\sum \omega(l\bar{b}) \times \omega(\bar{l}b)$, where ω is a corresponding weight and the sum runs over smearings in which a solution of the kinematic equations 7.1-7.6 is found) is taken for further analysis.

   However, a combination where both jets are tagged as b-jets, is preferred even if it provides a lower sum of the weights of the smearings as the combination, where only one jet is b-tagged.

2. *Averaging out the effect of 100 smearings.* After one lepton-jet combination is chosen the weighted average is taken for every momentum component of the top-quark:

$$p_i(t/\bar{t}) = \frac{\sum_{j=1}^{100} p_i^j(t/\bar{t})\omega_j}{\sum_{j=1}^{100} \omega_j}, \tag{7.10}$$

   where i is a component index of the momentum vector ($i = x, y, z$) and ω_j denotes the weight related with the invariant mass of lepton-jet pair, $\omega_i = \omega_i(m(l\bar{b})) \cdot \omega_i(m(\bar{l}b))$. In case there was no solution found for a specific smearing, the weight and the momentum are assumed to be zero.

   After the unique three-vector of the top and antitop momentum is determined this way, it is completed by assuming a top mass of 172.5 GeV to obtain the full kinematics of the top and antitop quark.

[2]The generated $W^{\pm}$ mass distribution has a Breit-Wigner shape. World average mass of the $W^{\pm}$ boson is ~ 80.4 GeV [32].

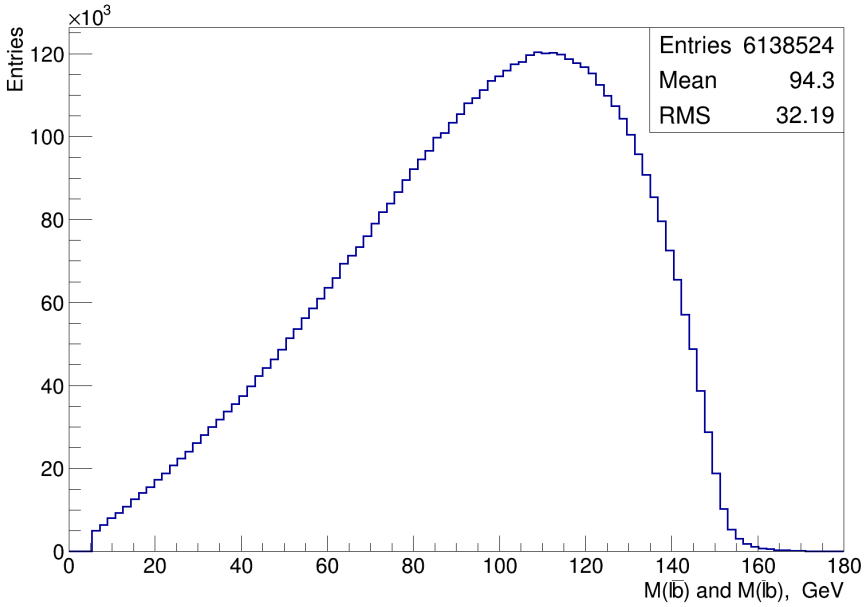

Figure 7.4: The area-normalized distribution of the invariant mass of the lepton – b-jet system originating from one $t/\bar{t}$-quark. This distribution is obtained from the signal $t\bar{t}$ MC on the generator level

7.4 Efficiency Studies

The events, where the kinematic reconstruction provides no solution, have to be discarded for the cross section measurements. The efficiency of the kinematic reconstruction is given as follows:

$$\epsilon^{kin.\ reco} = \frac{N^{with\ solution}}{N^{selected}}, \tag{7.11}$$

where $N^{selected}$ denotes the number of events, which fulfill all the selection criteria described in Chapter 6, and $N^{with\ solution}$ are the events, which also fulfill the selection from Chapter 6, but at the same time have a solution of the kinematic reconstruction.

The smearing procedure, employed in the kinematic reconstruction, increases the reconstruction efficiency, compared to the kinematic reconstruction previously applied for the single-differential $t\bar{t}$ production cross section measurements at 7 TeV at CMS [117], where no smearing of the reconstructed energies and directions of the objects objects was applied. This effect is presented in Fig. 7.5. More detailed studies are presented in [9].

The efficiency of the kinematic reconstruction was studied as function of several kinematic observables, such as the p_T and η of the leptons (see Fig. 7.6). The ratio of the data and MC efficiencies is in agreement with being a constant number of 0.9898 ± 0.0024 for the $e\mu$ channel. For the ee channel this number is 0.9871 ± 0.0050 and for the $\mu\mu$ - 0.9934 ± 0.0042. These scale factors were then applied as a weight for the reconstructed MC events.

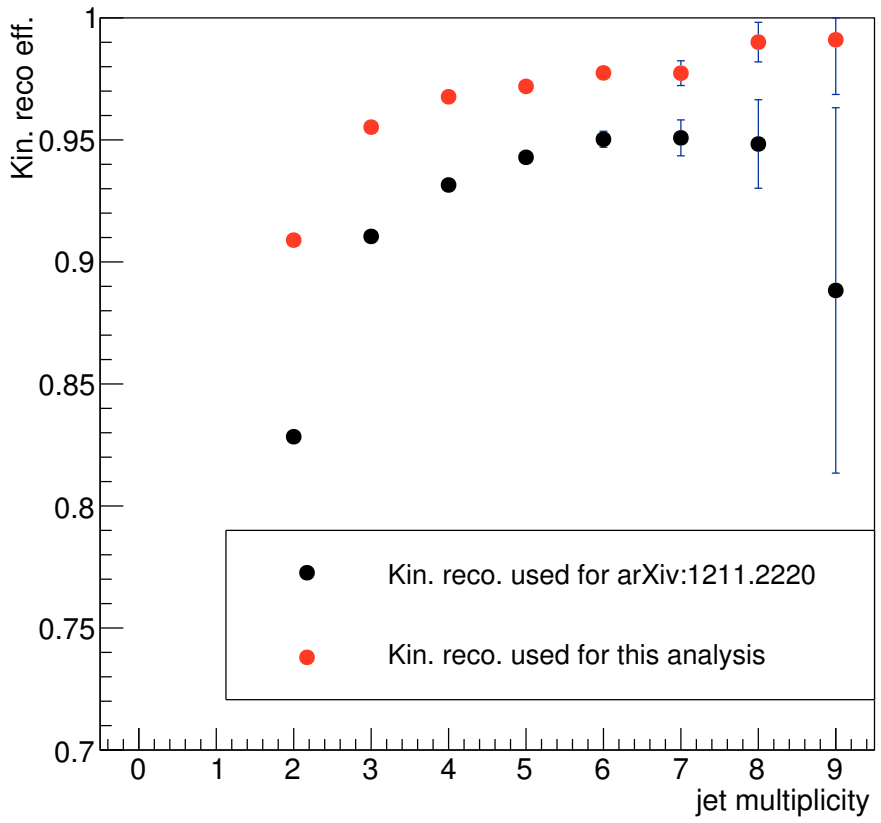

Figure 7.5: Efficiency of the kinematic reconstruction applied in this analysis compared to the one from [117] as a function of the jet multiplicities.

7.4.1 Studies of Inefficiencies and Migrations

In order to find possible ways of improving the kinematic reconstruction described in this chapter, one first needs to understand the reason of inefficiencies of the method. A number of event-wise studies has been performed in this context. The events, which were not reconstructed by the kinematic reconstruction or rejected by the cleaning were manually analyzed to understand the reason of fail. Only the simulated $t\bar{t}$ samples were analyzed, as they represent the signal process.

There are several reasons, which lead to failing of the kinematic reconstruction procedure or a poor reconstruction quality:

- *"Neutrino explosion"*: this name was introduced to describe an effect when neutrinos are reconstructed with a momentum, which is much higher than the true one. Even if the MET, which restricts the sum of the neutrino transverse momentum vectors (see Eq. 7.3 - 7.4), is small, the reconstructed neutrino momenta are large and back-to-back, not violating the MET restrictions. In the following an example of an event in the MC signal sample, which contains "exploded" neutrinos is presented:

```
Jets-reco info:
```

```
jet 0: (Pt,eta,Phi)=(158.478043 1.047849 -1.153090) btagDisc=0.448506
jet 1: (Pt,eta,Phi)=(112.374299 -0.263787 1.845721) btagDisc=0.002744

Lep-reco info:
lep 0: (Pt,eta,Phi)=(277.170469 -0.250428 2.469482) lepPdgId=-11
lep 1: (Pt,eta,Phi)=(125.735207 1.634679 -0.760010) lepPdgId=13


True info:
tt     -> (Mtt,Pt,eta,Phi)=(1165.230517 17.417528 4.405981 -2.642429)
  top      -> (Pt,eta,Phi)=(398.782867 -0.246215 2.294534)
    B      -> (Pt,eta,Phi)=(102.745514 -0.357465 1.664987)
      neutrino    -> (Pt,eta,Phi)=(10.949193 1.044668 2.508135)
      Antilepton  -> (Pt,eta,Phi)=(310.531647 -0.240176 2.482983)
  topbar   -> (Pt,eta,Phi)=(395.269012 1.468396 -0.890030)
    Bbar   -> (Pt,eta,Phi)=(152.874176 1.042002 -1.147113)
      Antineutrino -> (Pt,eta,Phi)=(6.874473 0.578765 1.082482)
      lepton       -> (Pt,eta,Phi)=(252.215057 1.629710 -0.760642)

Reco info:
tt     -> (Mtt,Pt,eta,Phi)=(1611.608396 25.791283 4.341296 -1.108135)
  top      -> (Pt,eta,Phi)=(585.895660 -0.188187 2.316067)
    B      -> (Pt,eta,Phi)=(112.374299 -0.263787 1.845721)
      neutrino    -> (Pt,eta,Phi)=(207.440300 -0.049013 2.362920)
      Antilepton  -> (Pt,eta,Phi)=(277.170469 -0.250428 2.469482)
  topbar   -> (Pt,eta,Phi)=(610.706178 1.351988 -0.837303)
    Bbar   -> (Pt,eta,Phi)=(158.478043 1.047849 -1.153090)
      Antineutrino -> (Pt,eta,Phi)=(339.963619 1.324092 -0.741431)
      lepton       -> (Pt,eta,Phi)=(125.735207 1.634679 -0.760010)
```

Here the kinematics of all the jets and leptons in the event is listed. The generated and reconstructed kinematics of all the objects in the $t\bar{t}$ decay chain are presented. Additionally, the information about the PDG particle codes for the reconstructed leptons (lepPdgId) and b-tagging discriminator value (as described in Sec. 5.4.3) for the reconstructed jets (btagDisc) is also present. The neutrinos after reconstruction appear to have more than ten times larger transverse momenta than the generated neutrinos.

Exploded neutrinos result in artificially large reconstructed top momenta, thus they cause migrations from the lower $p_T(t)$ to higher ones (see the scatter plot[3] in Fig. 7.7).

- *"Neutrino implosions"*: an opposite effect to the neutrino explosions, which results in migrations from higher to lower $p_T(t)$. However, the fraction of the events with

[3] In this thesis the two-dimensional distribution of one variable on the reconstructed versus the generator MC level are called *scatter plot*. These distributions illustrate the achieved reconstruction resolution and might point to possible biases in the reconstruction.

"imploded neutrinos" is smaller than the fraction of events with neutrino explosions. An example of an event with "neutrino implosions" is presented below:

```
Jets-reco info:
jet 0: (Pt,eta,Phi)=(278.364911 1.352950 2.742735) btagDisc=0.345707
jet 1: (Pt,eta,Phi)=(30.913907 0.746062 0.512795) btagDisc=0.662316

Lep-reco info:
lep 0: (Pt,eta,Phi)=(237.248785 1.055224 3.078619) lepPdgId=11
lep 1: (Pt,eta,Phi)=(177.654330 0.216169 -0.524885) lepPdgId=-13

True info:
tt     -> (Mtt,Pt,eta,Phi)=(1699.451492 5.646468 6.108470 -2.212514)
  top     -> (Pt,eta,Phi)=(747.064026 0.236955 -0.309609)
    B     -> (Pt,eta,Phi)=(21.369852 0.585819 0.856567)
      neutrino    -> (Pt,eta,Phi)=(558.018311 0.222328 -0.274407)
      Antilepton  -> (Pt,eta,Phi)=(185.193146 0.216175 -0.523356)
  topbar  -> (Pt,eta,Phi)=(748.924011 1.170368 2.839111)
    Bbar  -> (Pt,eta,Phi)=(281.758209 1.358503 2.733925)
      Antineutrino -> (Pt,eta,Phi)=(246.832703 0.997289 2.738154)
      lepton       -> (Pt,eta,Phi)=(229.696869 1.055060 3.078484)

Reco info:
tt     -> (Mtt,Pt,eta,Phi)=(1293.248465 111.897830 2.885152 -2.142542)
  top     -> (Pt,eta,Phi)=(549.404804 0.257593 -0.452094)
    B     -> (Pt,eta,Phi)=(30.913907 0.746062 0.512795)
      neutrino    -> (Pt,eta,Phi)=(347.745049 0.219277 -0.488359)
      Antilepton  -> (Pt,eta,Phi)=(177.654330 0.216169 -0.524885)
  topbar  -> (Pt,eta,Phi)=(573.623015 1.190099 2.884407)
    Bbar  -> (Pt,eta,Phi)=(278.364911 1.352950 2.742735)
      Antineutrino -> (Pt,eta,Phi)=(70.812822 0.833765 2.750936)
      lepton       -> (Pt,eta,Phi)=(237.248785 1.055224 3.078619)
```

- *Poor reconstruction of the jets*: the jets are often reconstructed wrongly and not only due to the detector resolution. For example, a large jet can be split into two geometrically close smaller jets in the case of a gluon radiation from a jet. Thus, the kinematic reconstruction fails due to the wrong input.

- *Choice of the wrong jet*: in some cases in the events with higher jet multiplicities the kinematic reconstruction prefers a $t\bar{t}$ candidate with a wrong jet taken as a b- or $\bar{b}$-jet, resulting in an invalid $t\bar{t}$ kinematics. In the following output an event with a wrongly chosen jet is presented:

```
Jets-reco info:
jet 0: (Pt,eta,Phi)=(202.019867 -0.543808 1.120093) btagDisc=0.055552
```

```
jet 1: (Pt,eta,Phi)=(167.175595 1.106433 -2.431321) btagDisc=0.116493
jet 2: (Pt,eta,Phi)=(145.248271 0.063589 -0.081827) btagDisc=0.854965
jet 3: (Pt,eta,Phi)=(54.716715 -1.742395 2.247415) btagDisc=0.173753
jet 4: (Pt,eta,Phi)=(39.340558 1.547696 0.035663) btagDisc=0.308361
jet 5: (Pt,eta,Phi)=(37.873044 -0.235022 0.548796) btagDisc=0.089997
jet 6: (Pt,eta,Phi)=(20.614101 2.298694 2.523000) btagDisc=-1.000000
jet 7: (Pt,eta,Phi)=(14.516475 2.277035 -0.018240) btagDisc=0.688632

Lep-reco info:
lep 0: (Pt,eta,Phi)=(207.247646 0.699888 -2.406876) lepPdgId=-11
lep 1: (Pt,eta,Phi)=(116.719016 -0.028368 0.250066) lepPdgId=13

 True info:
 tt     -> (Mtt,Pt,eta,Phi)=(1240.697371 295.969731 1.577808 -2.182287)
   top      -> (Pt,eta,Phi)=(687.846252 0.886633 -2.536537)
     B      -> (Pt,eta,Phi)=(211.556152 1.146382 -2.399150)
       neutrino    -> (Pt,eta,Phi)=(269.830505 0.774717 -2.742505)
       Antilepton  -> (Pt,eta,Phi)=(215.754349 0.698036 -2.414751)
   topbar   -> (Pt,eta,Phi)=(422.905640 -0.015684 0.359838)
     Bbar   -> (Pt,eta,Phi)=(149.453568 0.062657 -0.089313)
       Antineutrino -> (Pt,eta,Phi)=(186.666763 -0.069928 0.792180)
       lepton       -> (Pt,eta,Phi)=(119.529053 -0.024594 0.248166)

 Reco info:
 tt     -> (Mtt,Pt,eta,Phi)=(749.847640 339.594236 1.210644 -2.399249)
   top      -> (Pt,eta,Phi)=(518.528691 0.795921 -2.512406)
     B      -> (Pt,eta,Phi)=(167.175595 1.106433 -2.431321)
       neutrino    -> (Pt,eta,Phi)=(146.433487 0.430315 -2.741130)
       Antilepton  -> (Pt,eta,Phi)=(207.247646 0.699888 -2.406876)
   topbar   -> (Pt,eta,Phi)=(185.121256 0.326364 0.420538)
     Bbar   -> (Pt,eta,Phi)=(39.340558 1.547696 0.035663)
       Antineutrino -> (Pt,eta,Phi)=(47.949441 -0.306881 1.155024)
       lepton       -> (Pt,eta,Phi)=(116.719016 -0.028368 0.250066)
```

Here a wrong jet was taken as a $\bar{b}$-jet on the reconstruction level: jet 4 was chosen, although the correct jet is the jet 2. Both jets fulfill the CSVL tagging criterion (see Sec. 5.4.3).

- *Resolution effects*: there are cases, when the leptons or jets are poorly reconstructed due to the detector resolution effects and the kinematic reconstruction does not fix their kinematics with the smearing procedure. In these cases, the kinematic reconstruction provides no solution and the event is not analyzed further.

For a limited number of events the true and reconstructed kinematics was checked to conclude on the frequencies of the effects described above. The most frequently observed effect is the neutrino explosion. Another sizable effect is the choice of the wrong jet. It

significantly contributes to the fraction of the wrongly reconstructed events and is not treated so far. The poor jet reconstruction appears less frequent. The resolution effects are not large.

7.5 Control Distributions and Event Yields

The kinematic reconstruction provides the information about the full kinematics of the $t\bar{t}$ system. The results of the kinematic reconstruction can be illustrated by control distributions for different variables describing the $t\bar{t}$ kinematics (see Fig. 7.8). The data distributions are compared to the MC estimated signal and background contributions. An overall reasonable agreement is observed. However, some trends are visible. The distribution of the $p_T(t)$ is slightly harder in the MC. This trend is also observed in other measurements, like [35] or [9]. The rapidity of the top quarks are more central in the MC compared to the data, while for the $y(t\bar{t})$ one can observe that the data is more central. The $p_T(t\bar{t})$ and $M(t\bar{t})$ distributions are overall well described by the MC. Since the estimated background contributions are small, the aforementioned differences between data and MC must be coming mainly from deficiencies of the MadGraph + Pythia signal MC.

The event yields in data and MC after the selection (see Chapter 6) and kinematic reconstruction are presented in Table 7.1.

e^+e^- channel	lepton selection	jet selection	MET	b-tag	kin. reco.
$t\bar{t}$ signal	15699.8	11495.6	8741.5	8100.4	7469.8
$t\bar{t}$ other	2731.8	2020.3	1564.1	1433.2	1339.2
$t\bar{t}+Z/W/\gamma$	209.6	183.8	145.4	131.5	113.7
tW	1812.8	705.7	533.1	460.9	332
diboson	4695.4	534.4	166.5	62.2	38.8
W	1596.9	96.9	66.6	27.4	27.1
DY$\to\tau^+\tau^-$	12692.6	746.6	330.8	126.6	108
DY$\to\ell^+\ell^-$	461750	21309.3	1039.3	376.7	249.2
QCD	1780.1	86.8	0	0	0
Sum MC	502970	37179.7	12587.6	10719.3	9678
Data	492406	37581	12299	10369	9360
$\mu^+\mu^-$ channel	lepton selection	jet selection	MET	b-tag	kin. reco.
$t\bar{t}$ signal	20289.7	14956.2	11387	10560.4	9817.3
$t\bar{t}$ other	3882.5	2870	2202.7	2006.6	1893
$t\bar{t}+Z/W/\gamma$	282.1	247.2	196.6	179.3	156.1
tW	2375.7	902.6	681.1	591.1	442.5
diboson	6538	748.5	229	86.6	53.9
W	418.7	101.5	44.9	7.2	7.2
DY$\to\tau^+\tau^-$	20708.8	1061.6	439.2	143.8	97.8
DY$\to\ell^+\ell^-$	721995	31571.9	1804.5	696.5	441.7
QCD	9567.5	763.8	0	0	0
Sum MC	786058	53223.5	16985.2	14271.9	12909.7
Data	843514	57548	18206	15168	13721
$e^\pm\mu^\mp$ channel	lepton selection	jet selection	MET	b-tag	kin. reco.
$t\bar{t}$ signal	43904.9	32357.2	32357.2	30049.2	28082.2
$t\bar{t}$ other	8348.4	6175.1	6175.1	5641.1	5339.9
$t\bar{t}+Z/W/\gamma$	565.4	494.5	494.5	450.8	399.2
tW	5133.2	1925.5	1925.5	1679.4	1275.8
diboson	11006	596.6	596.6	209.8	147.8
W	2568.1	213.2	213.2	73.5	55.9
DY$\to\tau^+\tau^-$	32837.8	1719.8	1719.8	626	499.3
DY$\to\ell^+\ell^-$	1180.5	85.4	70.1	32.8	19.5
QCD	20652.6	311.9	311.9	226.3	224
Sum MC	126197	43879.6	43864.3	38989.3	36044
Data	107469	43435	43435	38774	35940
combined sample	lepton selection	jet selection	MET	b-tag	kin. reco.
$t\bar{t}$ signal	79894.3	58809	52485.7	48710.1	45369.3
$t\bar{t}$ other	14962.8	11065.6	9942	9081	8572.1
$t\bar{t}+Z/W/\gamma$	1057.2	925.5	836.5	761.7	669.2
tW	9321.8	3533.9	3139.8	2731.5	2050.4
diboson	22239.5	1879.6	992.2	358.7	240.5
W	4583.8	411.7	324.8	108.2	90.2
DY$\to\tau^+\tau^-$	66239.2	3528.1	2489.8	896.6	705.2
DY$\to\ell^+\ell^-$	1.18493e+06	52966.6	2864.6	1084.7	698
QCD	32000.3	1162.7	311.9	226.3	224
Sum MC	1.41523e+06	134283	73387.8	63959.1	58619.3
Data	1.44339e+06	138564	73940	64311	59021

Table 7.1: Number of selected events in data and different MC samples after different selection steps. The numbers are shown in the ee, $e\mu$, $\mu\mu$ and combined channels.

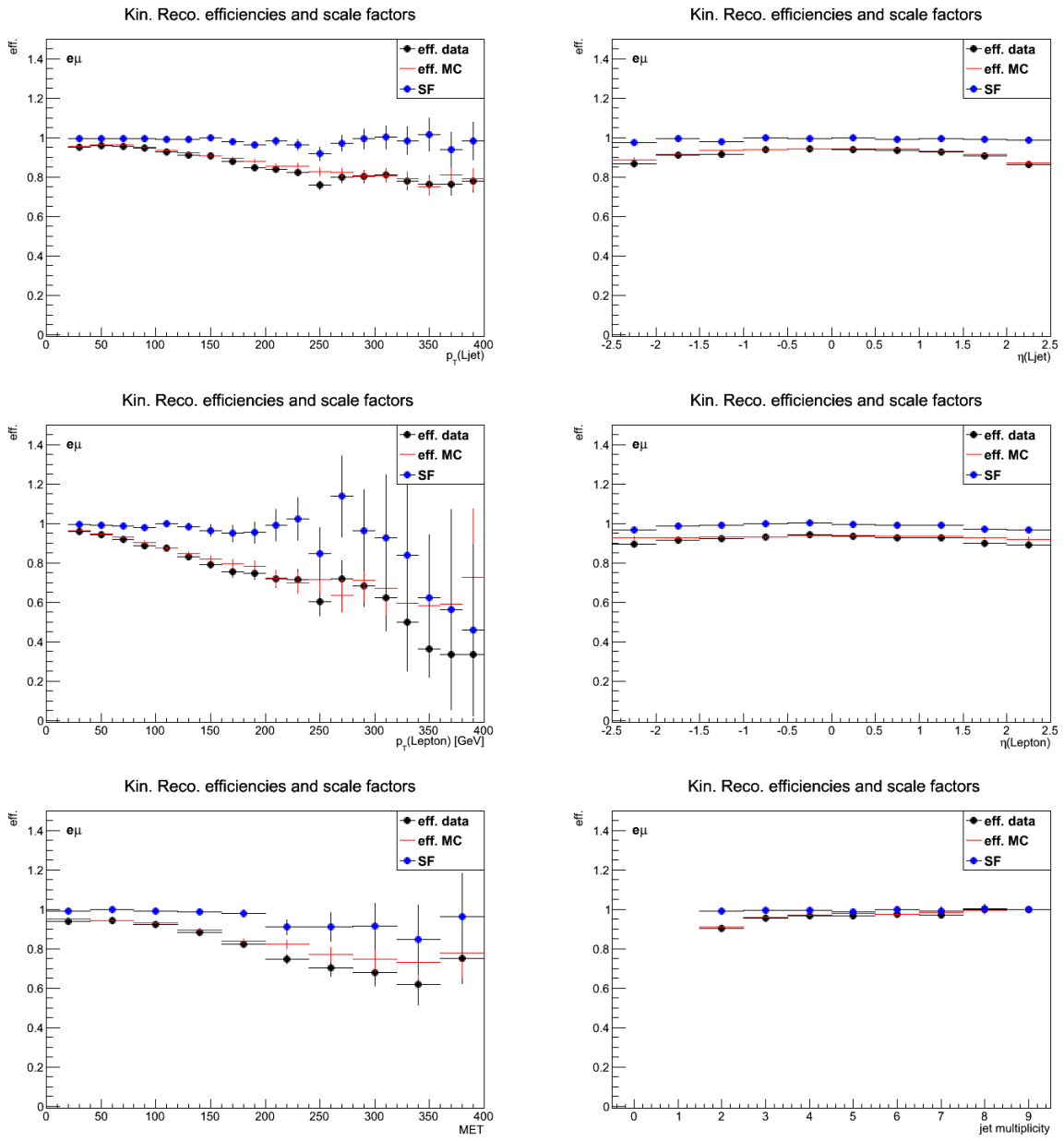

Figure 7.6: The efficiencies and scale factors related to the $t\bar{t}$ dilepton kinematic reconstruction in bins of b-jet p_T (top left) and η (top right), lepton p_T (middle left) and η (middle right), MET (bottom left) and jet multiplicity (bottom right) in the events in the $e\mu$ channel. The efficiencies in data are presented with black dots, efficiencies in MC are marked with red dots and the scale factors, determined from the data-to-MC ratio, are represented with blue dots. In the MC simulation presented in this plot both signal and background processes are taken into account.

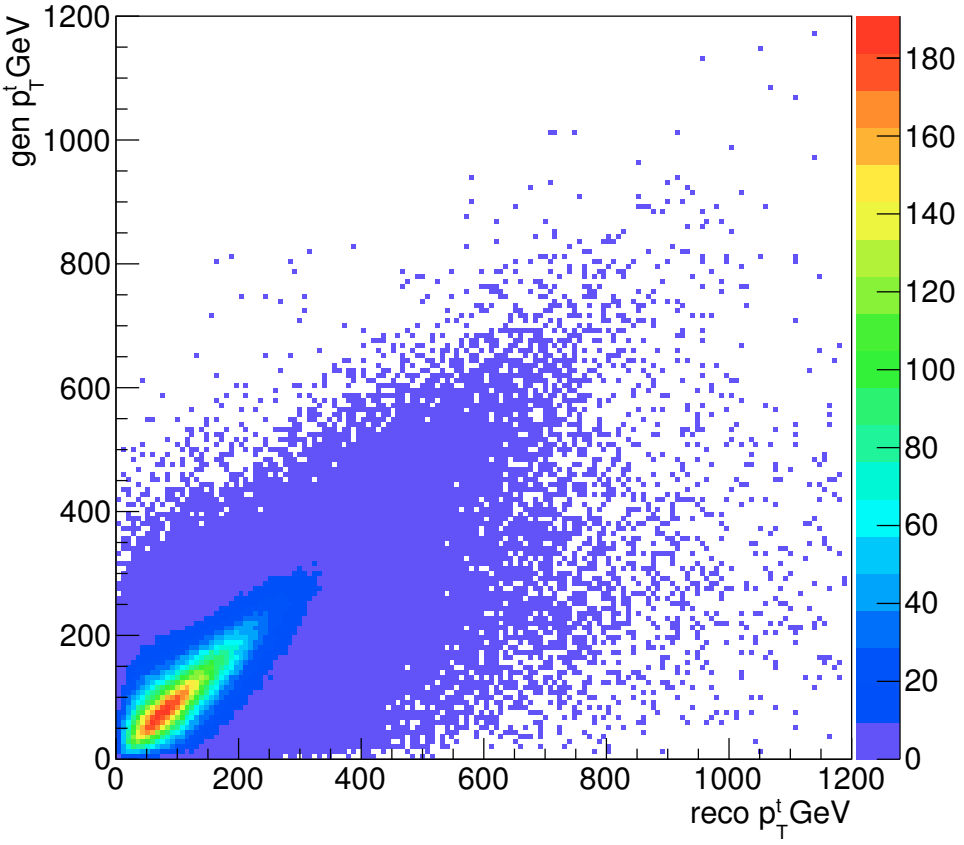

Figure 7.7: Scatter plot of the $p_T(t)$ in the combined dilepton channel.

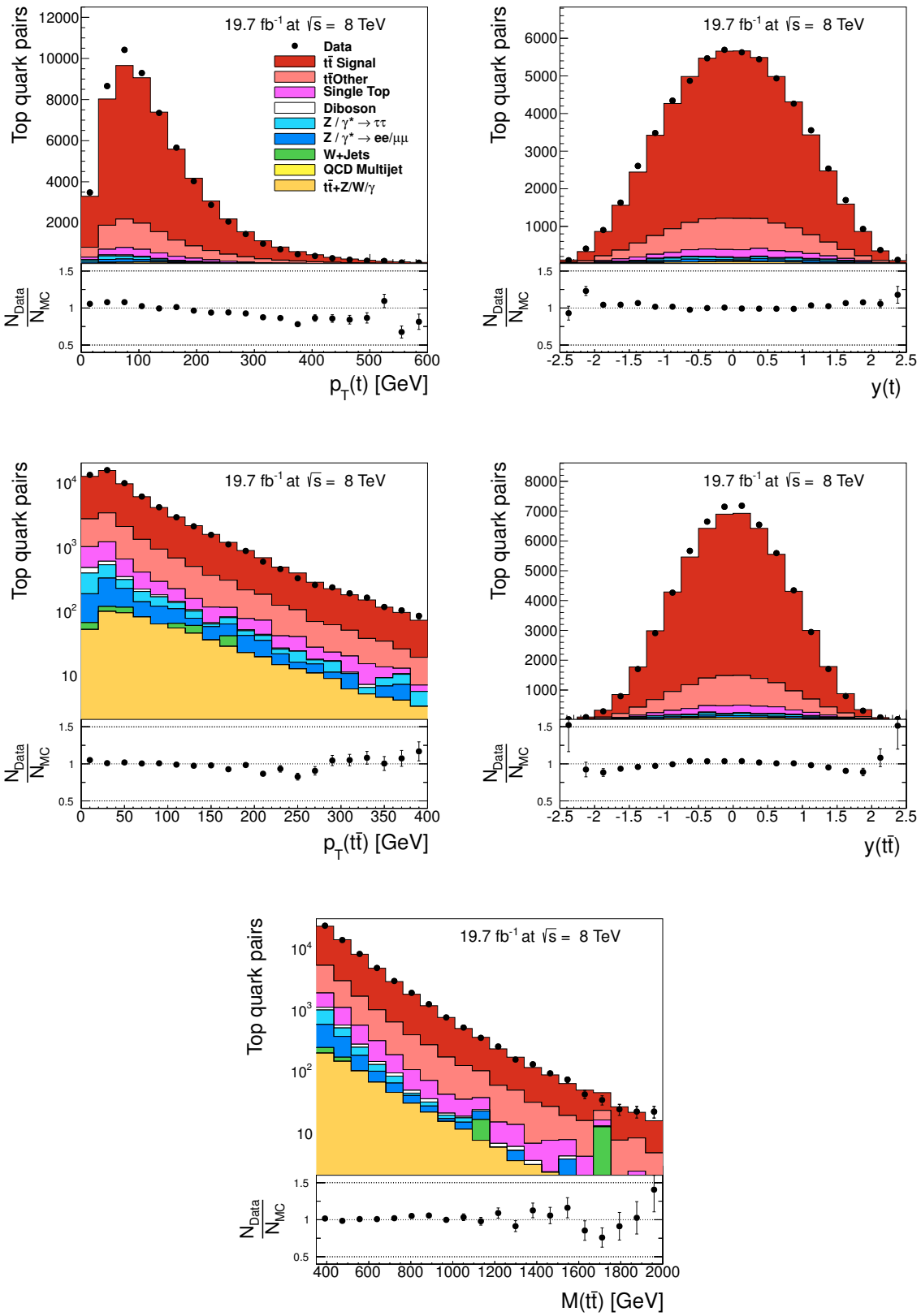

Figure 7.8: Control distributions of the $p_T(t)$ (top left), $y(t)$ (top right), $p_T(t\bar{t})$ (middle left), $y(t\bar{t})$ (middle right) and $M(t\bar{t})$ (bottom). Other details as in Fig. 6.5.

8 | Optimization for the Boosted Topologies

In order to obtain optimal results for the cross sections for the boosted topologies (see e.g. Sec. 6.6.1 and Sec. 7.4), some tuning is required for the lepton isolation requirements and for the event selection. The specific criteria (the 2D isolation and the cleaning conditions) applied in this analysis specifically for a better reconstruction of the boosted topologies helped to increase the efficiency of the analysis procedure and the purity of the signal in the region with high $p_T(t)$.

The tunes for the boosted topologies reconstruction are based on a number of studies and observations, the results of which are presented in this chapter.

8.1 Studies of the Lepton Isolation

The principal 2D isolation criteria applied for the leptons in this analysis (see Sec. 6.6.1) is taken over from a similar $t\bar{t}$ production analysis in the semi-leptonic decay channel [118]. The performance of this criterion was checked in detail in the present analysis.

In Fig. 8.1 efficiencies for different isolation criteria in bins of $p_T(t)$ are presented. One can see that the harder lepton isolation used for the $p_T(t)$ inclusive measurements has a low efficiency at high $p_T(t)$ and in general constantly and fastly drops with increasing $p_T(t)$, which is not desired for the analyses with boosted top quarks. The efficiencies of the two variants of 2D isolation is high in the boosted region. The first variant is a standard 2D isolation introduced in Sec. 6.6.1 and the second one is defined between leptons and b-jets in the event[1]. The efficiency of the isolation, which associates leptons and b-jets only, is higher and one is tempted to use it for the analysis. However, one needs to check this criterion in more detail.

In Fig. 8.2 the distribution of p_T^{rel} between leptons and jets (see Sec. 6.6.1) is shown. In case when the 2D isolation is defined between leptons and closest b-jets only, a large spike at small p_T^{rel} values is observed, for which the MC doesn't describe data well. In the region of this spike all the background sources are peaking, which indicates that there is background in the data, which is not properly described in the MC in this region. For the 2D isolation defined between leptons and any closest jets this peak is not present. Thus, the 2D isolation, which is determined between leptons and the closest jets (any jets, not necessarily b-jets) in the events provides a cleaner signal and a more reliable modeling by

[1] Here association of leptons to b-jets only means that the ΔR and p_T^{rel} in the 2D isolation (as described in Sec. 6.6.1) is determined between the lepton and the b-jet closest in ΔR.

the MC. That is why it was decided to take the 2D isolation between leptons and jets of any origin for the analysis.

8.2 Cleaning Conditions

8.2.1 Specification of the Kinematic Reconstruction for the Boosted Top Quarks

The method of the kinematic reconstruction of the $t\bar{t}$ system in the dileptonic final state, as described in Chapter 7, has been employed for the measurements of the single- [35,104] and double-differential $t\bar{t}$ production cross sections [9]. These measurements were done at low and medium $p_T(t)$ and the kinematic reconstruction showed a good performance over all the tested kinematic region.

In this work the measurement at high top p_T (boosted region) is performed. Thus, the kinematic reconstruction has to be tested in this region.

Fig. 7.7 shows the scatter plot for the $p_T(t)$. This plot shows that there is a large fraction of events reconstructed in the boosted region, which were actually generated with low momentum. This problem is partially caused by a very large number of events with small true transverse momenta of the top quarks for which the $t\bar{t}$ production cross section is rather large.

The problem of migrating events from the lower $p_T(t)$ to the higher ones is solved by not taking them to account in the boosted analysis. There are two ways considered in this work to deal with such events:

- not take these events to account for the further analysis. For this, a requirement $M(b\bar{b}l\text{MET})/M(t\bar{t}) > 0.5$ is applied. If the event doesn't fulfill this requirement, it is not counted further.

  The motivation for this requirement is the following. The value $M(b\bar{b}l\text{MET})/M(t\bar{t})$ is a relation between the invariant mass of the system of the objects, which enter the reconstruction and the invariant mass of the objects resulting from the reconstruction procedure. If the reconstruction results in a very high top momentum compared to the real momentum, the $M(t\bar{t})$ will be much larger than the $M(b\bar{b}l\text{MET})$, resulting in a very low value of the ratio $M(b\bar{b}l\text{MET})/M(t\bar{t})$. Thus, putting a minimum threshold on this ratio reduces the number of migrations from lower to higher $p_T(t)$ due to the kinematic reconstruction procedure. The variations of the minimum threshold values for the ratio $M(b\bar{b}l\text{MET})/M(t\bar{t})$ is presented in the Sec. 8.2. The applying of the minimum threshold on the ratio $M(b\bar{b}l\text{MET})/M(t\bar{t})$ will be further called *cleaning*.

  In the Fig. 8.3, the same scatter plot, as in Fig. 7.7, is presented, but after applying the cleaning. It shows a significant reduction of the events, which migrate from the lower to the higher $p_T(t)$ with minor influence on the correctly reconstructed events.

- use the cross section results obtained with the events with different ranges of $M(b\bar{b}l\text{MET})/M(t\bar{t})$, assuming lower weights for the bins with lower values of $M(b\bar{b}l\text{MET})/M(t\bar{t})$. This is done by the simultaneous unfolding of the studied distribution and the distribution

of the $M(b\bar{b}l\text{MET})/M(t\bar{t})$ quantity, where the latter is unfolded in one generated bin and four reconstructed bins. The unfolding procedure is described in Sec. 9.2.

The studies of the influence of the cleaning conditions on the $M(t\bar{t})$ distribution are presented in the Appendix E.

8.2.2 Variations of the Cleaning Condition

The cleaning conditions described in sec. 8.2.1 is $M(b\bar{b}l\text{MET})/M(t\bar{t}) > 0.5$. The distribution of this variable on the generator level is shown in Fig. 8.4. From the plots one can observe that the cleaning condition does not reject a large number of events, when requiring a minimum threshold at 0.5.

The minimum threshold of $M(b\bar{b}l\text{MET})/M(t\bar{t})$ was varied down to 0.4 and up to 0.6 and 0.7. The distribution of the reconstructed $M(b\bar{b}l\text{MET})/M(t\bar{t})$ with the various thresholds indicated by vertical lines are shown in Fig. 8.5. It is observed that in the events with high-p_T top quarks for the higher $M(b\bar{b}l\text{MET})/M(t\bar{t})$ values the agreement between data and simulation is getting worse - the MC overestimates the data yields. This is likely related to the effect, which is also observed in Fig. 7.8, that the composition of signal and background MCs overshoots the data at high p_T. As the total number of events with high p_T is small, this effect is not pronounced in the distribution shown on the top of Fig. 8.5 for the $p_T(t)$ inclusive sample.

One other observation from Fig. 8.5 is that some background samples are having larger contribution at low values of the $M(b\bar{b}l\text{MET})/M(t\bar{t})$. Thus, setting a threshold on the minimum $M(b\bar{b}l\text{MET})/M(t\bar{t})$ value may help to slightly reduce the fractions of the $t\bar{t} \rightarrow other$, single top and Drell-Yan background sources. To compare the performance of each of the minimum $M(b\bar{b}l\text{MET})/M(t\bar{t})$ threshold choices, the following was considered:

- **Reconstruction vs. generator scatter plots**: the scatter plots for the $p_T(t)$ for various thresholds for the cleaning conditions are compared between each other. These plots are presented in Fig. 8.6. From the visual comparison of the plots it is clearly seen that setting the threshold of $M(b\bar{b}l\text{MET})/M(t\bar{t})$ at 0.4 doesn't provide the cleaning effect as efficient as for the nominal threshold at 0.5. There is a significant fraction of the events migrating from lower true to higher reconstructed $p_T(t)$ ($p_T(t) > 400$ GeV). For the tighter thresholds at 0.6 and 0.7 a better cleaning effect is observed. This effect is the better the tighter the threshold is set. However, one should also account for the number of the correctly reconstructed events rejected by the threshold (efficiency loss).

- **Efficiency-Purity-Stability plots**: in the Fig.8.7 the efficiencies, purities[2] and stabilities[3] are shown for the different thresholds for $M(b\bar{b}l\text{MET})/M(t\bar{t})$ mentioned above and for the case, where no cleaning is applied. The purity for the loosest threshold of 0.4 is the lowest in the boosted region. For the other thresholds of 0.5, 0.6, 0.7 the purity is much better.

[2] The *purity* is related to the fraction of all reconstructed events in a bin which were also originating from the same bin on the true level. A more detailed discussion can be found in Sec. 9.4.

[3] The *stability* is related to the fraction of reconstructed events which were generated in a specific bin and were also reconstructed in the same bin. A more detailed discussion can be found in Sec. 9.4.

It is clear that without the cleaning conditions the purity in the boosted region collapses to very low values. Thus, applying even a loose cleaning condition is helpful. The most dramatic improvement in purity is observed when changing from the threshold of 0.4 to the threshold of 0.5 (from ~ 0.2 to ~ 0.3).

From these studies one can conclude that the minimal threshold of the quantity $M(b\bar{l}\bar{b}l\text{MET})/M(t\bar{t})$ is currently optimally set to 0.5 (0.6 is also acceptable). Looser and tighter thresholds are not expedient.

The cleaning condition brings a major improvement to the analysis of the top quarks with boosted topologies.

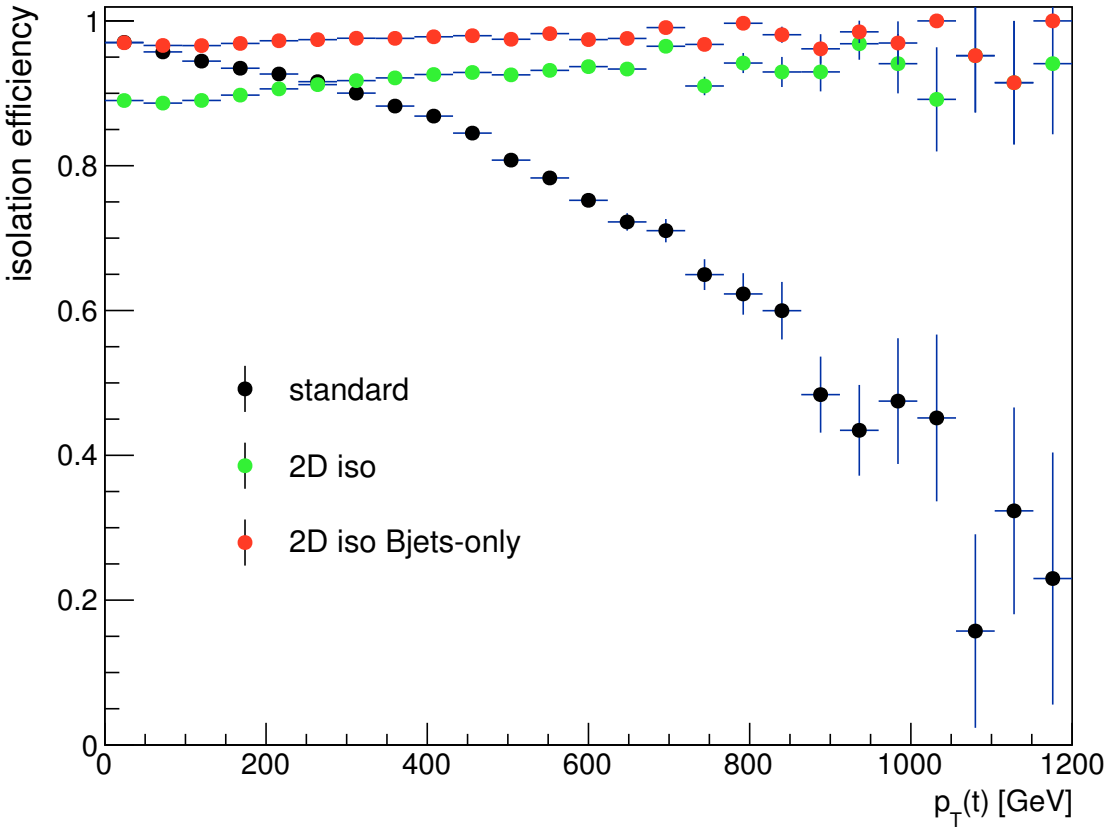

Figure 8.1: Lepton isolation efficiency in bins of the true $p_T(t)$. The efficiency of the relative isolation used for the unboosted measurement of the $t\bar{t}$ production cross section [35, 104] is marked with the black dots, the efficiency of the 2D isolation as described in Sec. 6.6.1 which is defined between lepton and any jet in the event, which is the closest one in ΔR to the lepton, is marked with the green dots and the efficiency of the 2D isolation as described in Sec. 6.6.1 which is defined between lepton and closest b-jets is marked with the red dots. The error bars represent the statistical uncertainties. All the efficiencies are obtained using the $t\bar{t}$ signal MadGraph + Pythia MC sample.

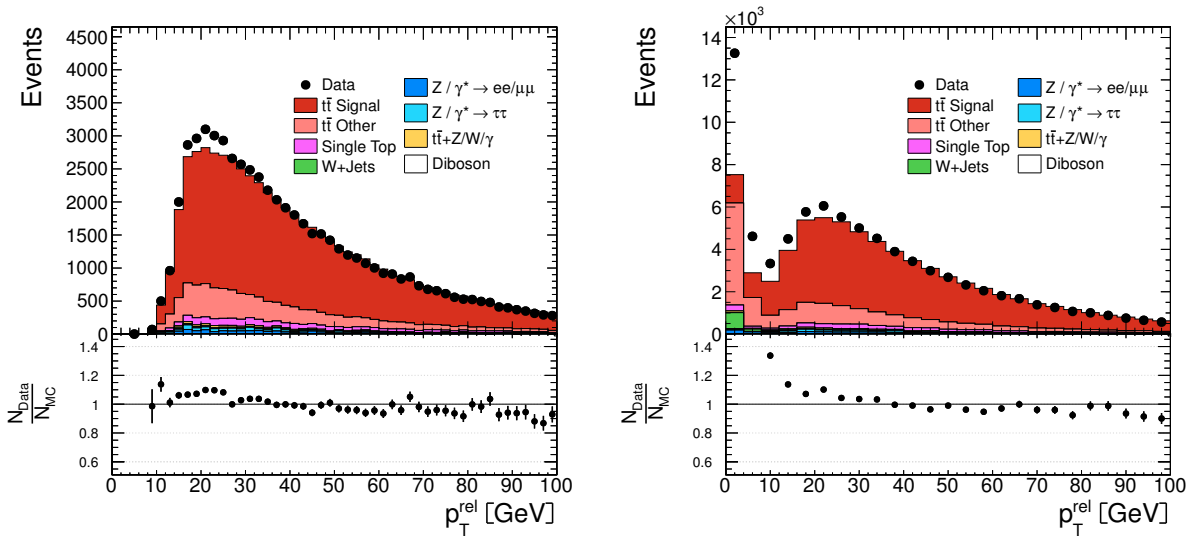

Figure 8.2: Control distributions of the p_T^{rel} between the leptons and the nearest (in (η, ϕ)) jets in the event with the 2D lepton isolation applied. On the left distribution the 2D isolation, which is defined for the leptons and all the jets in the event is applied and on the right distribution the 2D isolation which is defined as the minimum distance $\sqrt{\Delta\eta^2 + \Delta\phi^2}$ between a lepton and any (b-) jet is applied. The black dots are the data points with statistical uncertainties and the red histogram shows the distribution in the $t\bar{t}$ signal MC. The histograms with the other colors represent the different background contributions as estimated by the MC simulation.

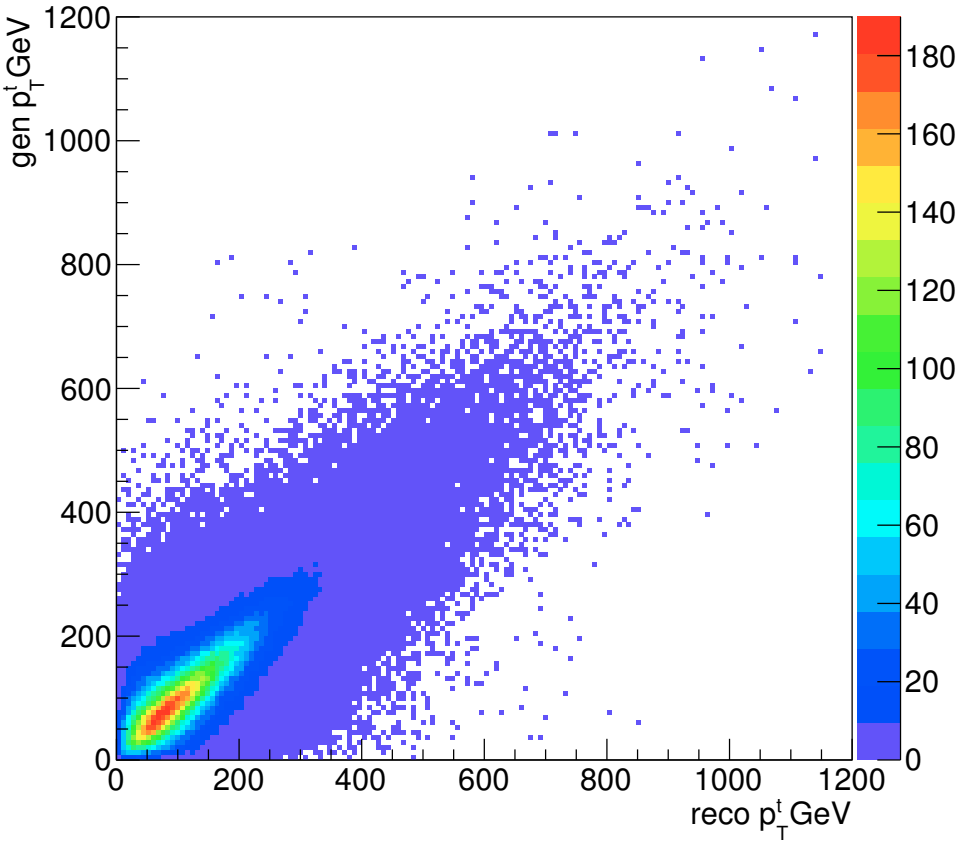

Figure 8.3: Scatter plot of the $p_T(t)$ in the combined dilepton channel after cleaning condition $M(b\bar{l}\bar{b}l\text{MET})/M(t\bar{t}) > 0.5$ is applied.

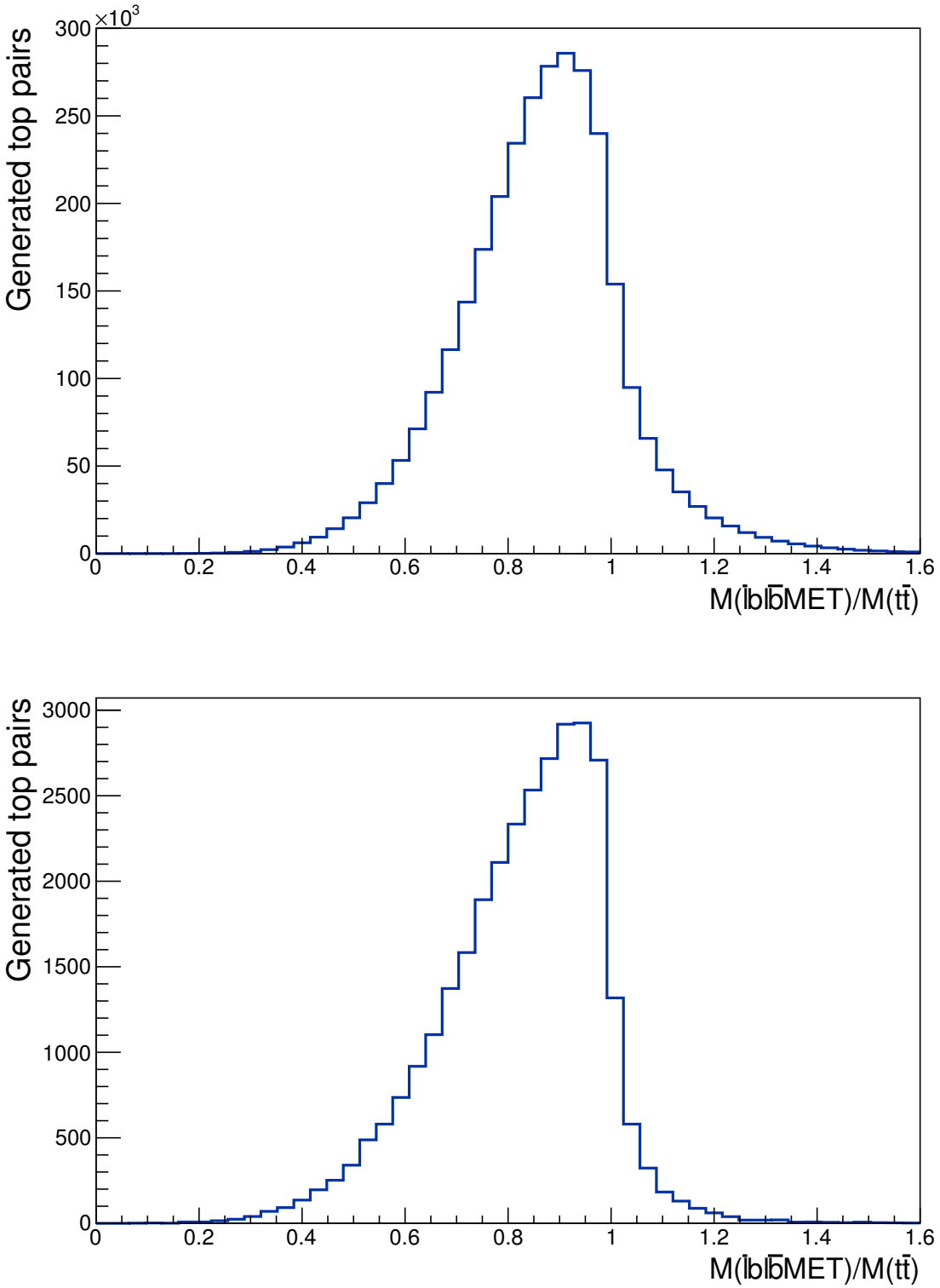

Figure 8.4: The distribution of $M(b\bar{l}\bar{b}l\text{MET})/M(t\bar{t})$ on the generator level for the whole $p_T(t)$ range (top) and for the large top transverse momenta (bottom). The plots were obtained using $t\bar{t}$ signal MC sample.

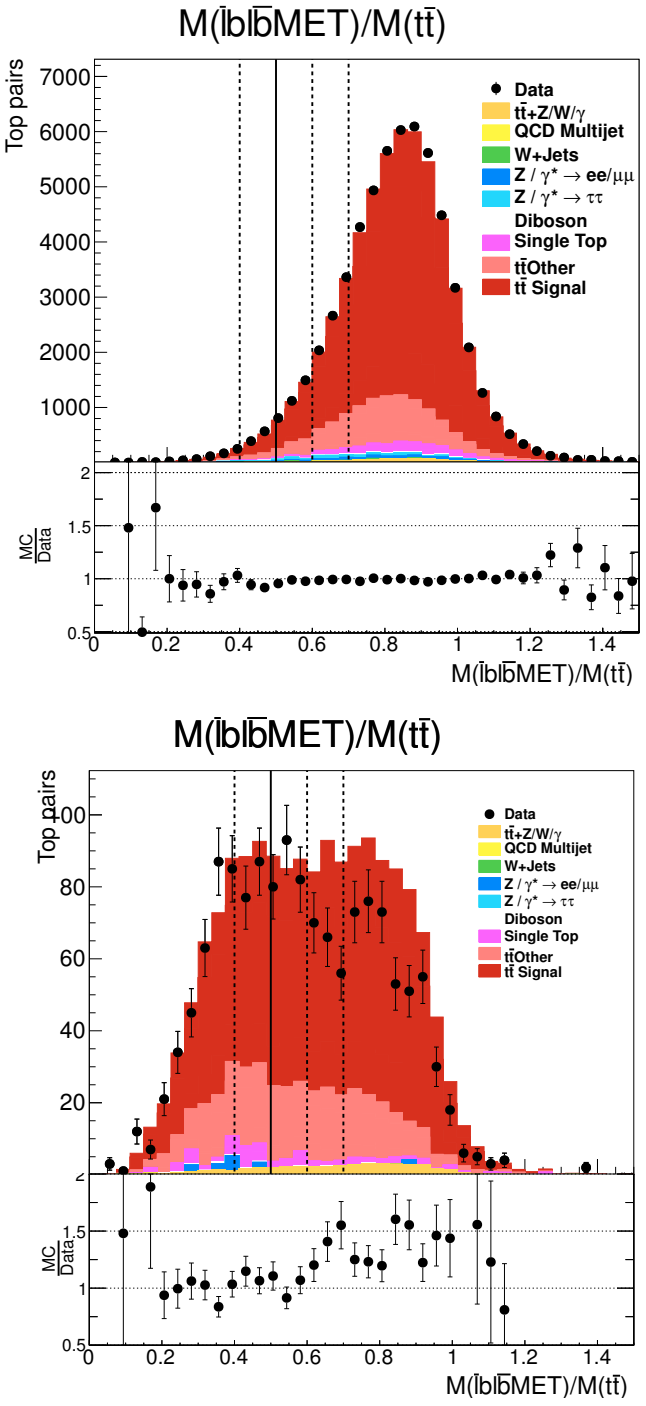

Figure 8.5: Control distribution of the $M(b\bar{l}\bar{b}l\text{MET})/M(t\bar{t})$ in total (top) and for the region with large top transverse momenta - $p_T(t) > 400$ (bottom). The vertical lines represent the minimum threshold values tested in the analysis. The solid line represents the nominal threshold value and the dashed lines correspond to the variations of the threshold.

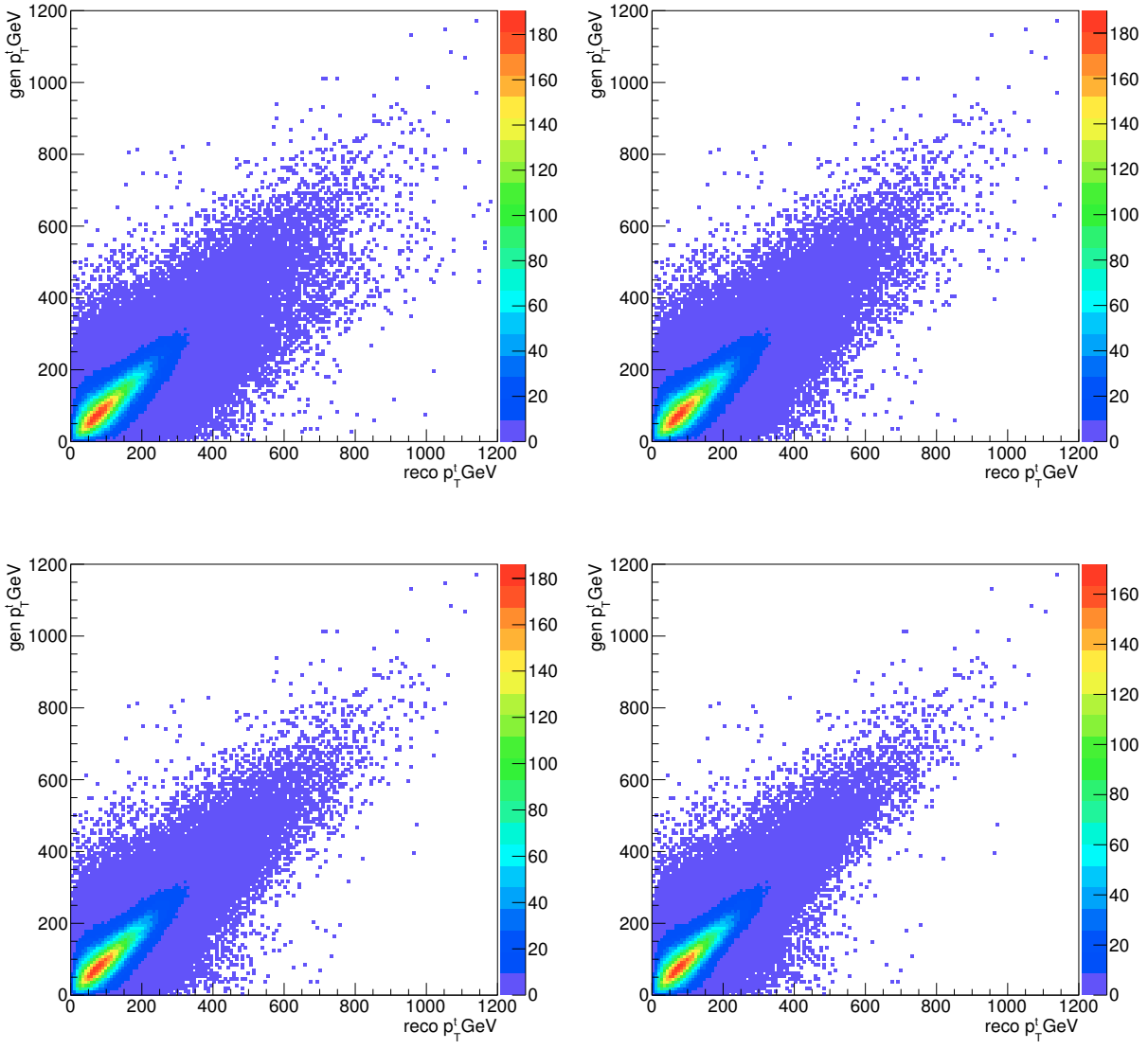

Figure 8.6: Scatter plots for $p_T(t)$ with a minimum threshold for $M(b\bar{l}\bar{b}l\text{MET})/M(t\bar{t})$ of 0.4 (top left), 0.5 (top right), 0.6 (bottom left) and 0.7 (bottom right).

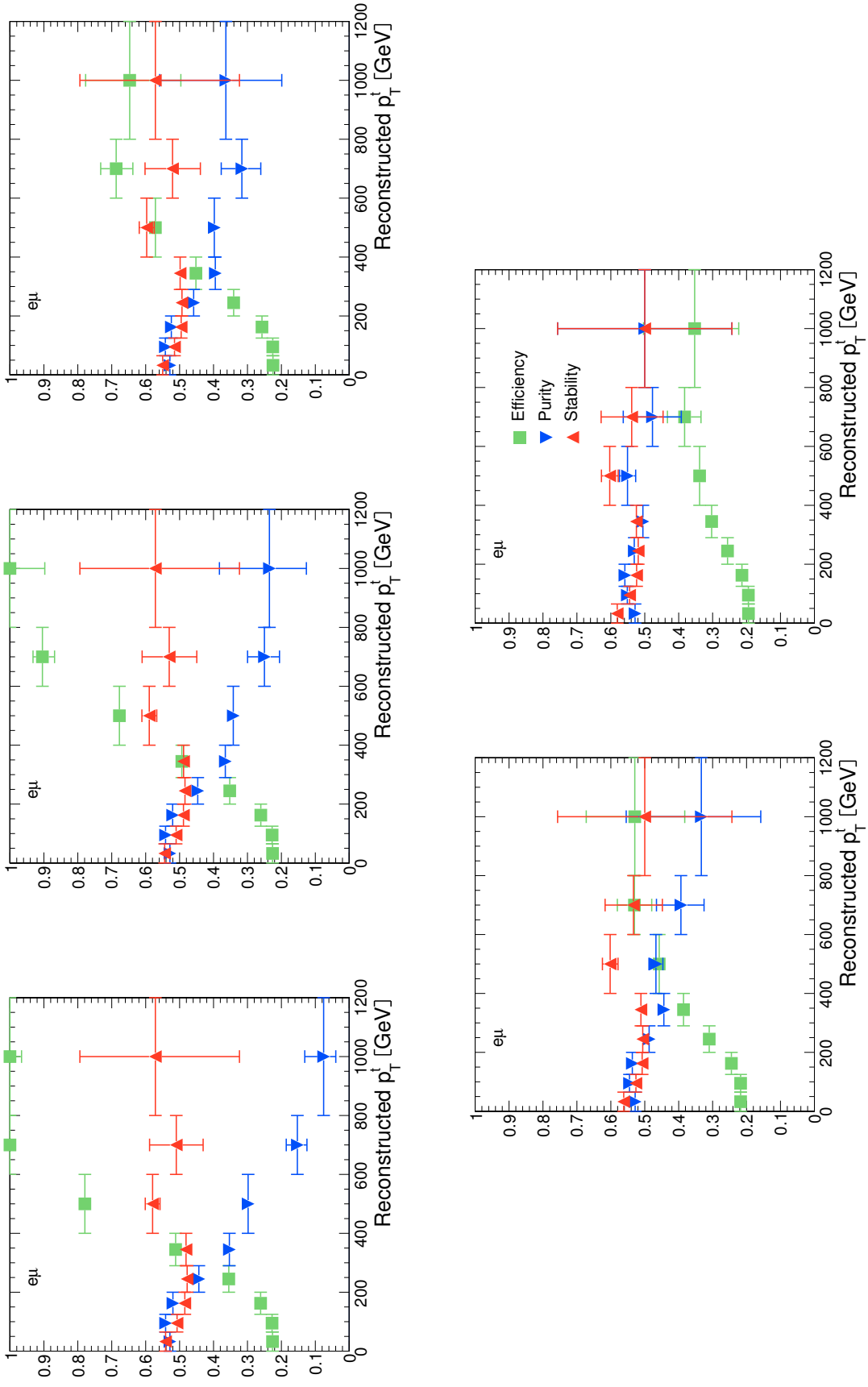

Figure 8.7: Efficiency, purity and stability in bins of $p_T(t)$ for the events with no cleaning (top left) and with the minimum threshold for $M(b\bar{b}l\bar{l}\text{MET})/M(t\bar{t})$ of 0.4 (top middle), 0.5 (top right), 0.6 (bottom left) and 0.7 (bottom right).

9 | Measurement of the Cross Sections

After selecting the events in which most likely the $t\bar{t}$ decaying in the dileptonic final state was produced, one should count them to determine the $t\bar{t}$ production cross sections. However, some further corrections have to be made to arrive at a proper measurement.

This chapter will give an overview of the sequential steps which are performed in this analysis to obtain the final cross sections results. It describes the background subtraction procedure, the unfolding of the data and the approaches used to determine the $t\bar{t}$ production cross sections in the boosted region.

9.1 Background Subtraction

The background processes are described in Sec. 6.5. Their yields which are estimated from the simulation (normalized to the data luminosity) are subtracted from the measured number of data events the following way:

$$N_{data}^{signal} = (N_{data}^{selected} - N_{backgr.}) \cdot \frac{N_{MC\ reco}^{signal\ t\bar{t}}}{N_{MC\ reco}^{signal\ t\bar{t}} + N_{MC\ reco}^{other\ t\bar{t}}}, \tag{9.1}$$

where $N_{data}^{selected}$ is the number of the data events, which fulfill the selection criteria described in Chapter 6 and have a solution of the kinematic reconstruction presented in Chapter 7 and $N_{bckgr.}$ is the number of all simulated backgrounds (properly normalized), except for the $t\bar{t} \rightarrow other$ background. The latter is corrected by the factor $\frac{N_{MC\ reco}^{signal\ t\bar{t}}}{N_{MC\ reco}^{signal\ t\bar{t}}+N_{MC\ reco}^{other\ t\bar{t}}}$ which is determined using MC simulation.

The separate treatment of the $t\bar{t} \rightarrow other$ is related to the fact that the $t\bar{t} \rightarrow other$ sample depends on the $t\bar{t}$ production cross section. As this measurement aims to measure the $t\bar{t}$ production cross section, no dependence on the cross sections in the samples, which are used for this measurement, is desired. In the factor $\frac{N_{MC\ reco}^{signal\ t\bar{t}}}{N_{MC\ reco}^{signal\ t\bar{t}}+N_{MC\ reco}^{other\ t\bar{t}}}$ the dependence on the $t\bar{t}$ production cross section assumed in the $t\bar{t}$ MC cancels.

The background subtraction is done separately for the three different $t\bar{t}$ dilepton decay channels (ee, $e\mu$ and $\mu\mu$) and the corrected event yields are afterwards added together for further calculations.

9.2 Unfolding

The measurements presented in the following are the integrated and differential cross sections in the top quark boosted region. For the differential measurements the events

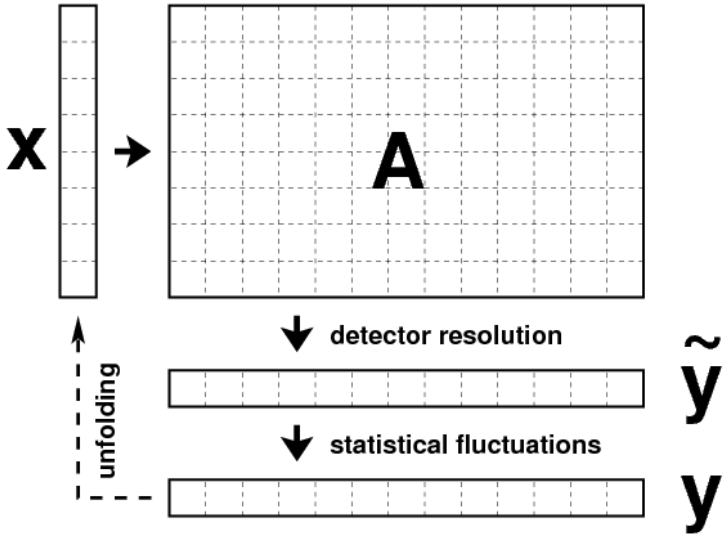

Figure 9.1: Schematic view of the unfolding problem. The plot taken from [119].

are counted in *bins* of a certain variable, or, in other words in some phase space regions which correspond to some restricted window of that variable.

Due to the limited detector resolution, an event, which was originally produced in one bin of a certain variable (p_T, y, E, etc.) may be reconstructed in a different bin of the same variable. This effect is called interbin *migrations*. It is also obvious that if migrations take place, then the measurement in one cross section bin is correlated with the measurements in the other bins. To take into account and correct for the migrations the *unfolding* procedure is applied.

There are various unfolding algorithms used for data analysis in high energy physics experiments. For this analysis the unfolding is performed with the TUnfold [119] tool.

9.2.1 Unfolding with TUnfold

The TUnfold tool aims to solve the problem presented in Fig. 9.1. The main goal of the unfolding is to gain the information about the "true" distribution ($\mathbf{x}$) for a certain variable out of the information which is available after the detector reconstruction represented by the distribution $\mathbf{y}$. The bin-to-bin migration probabilities are described by the migration matrix $\mathbf{A}$. In TUnfold it is recommended to choose the size of the vector $\mathbf{x}$ (or the number of bins of the true distribution) to be smaller than the size of the vector $\mathbf{y}$ (or the number of the bins of the reconstructed distribution). It also works if the number of bins in $\mathbf{x}$ and $\mathbf{y}$ are the same.

In the TUnfold algorithm the distribution $\mathbf{x}$ is extracted as a result of minimizing the following Lagrangian [119]:

$$\mathcal{L} = \mathcal{L}_{min} + \mathcal{L}_{reg} + \mathcal{L}_{area}, \text{ where} \tag{9.2}$$

$$\mathcal{L}_{min} = (\mathbf{y} - \mathbf{A}\mathbf{x})^T \mathbf{V}_{yy}^{-1} (\mathbf{y} - \mathbf{A}\mathbf{x}), \tag{9.3}$$

$$\mathcal{L}_{reg} = \tau^2 (\mathbf{x} - f_b \mathbf{x_0})^T (\mathbf{L}^T \mathbf{L})(\mathbf{x} - f_b \mathbf{x_0}) \text{ and} \tag{9.4}$$

$$\mathcal{L}_{area} = \lambda(\sum_i y_i - \mathbf{e}^T \mathbf{x}). \tag{9.5}$$

Each of the three parts of the Lagrangian $\mathcal{L}$ has its own meaning:

- The first term $\mathcal{L}_{min}$ is a usual χ^2 minimization term. The meaning of the quantities $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{A}$ has been discussed above. The migration matrix $\mathbf{A}$ for this analysis is determined from the MC $t\bar{t}$ signal sample. First, each element A_{ij} of the matrix is filled with the number of events which were generated in bin i, but reconstructed in bin j. Afterwards an additional "zero" row, which contains the numbers of events, which were generated, but not reconstructed, is added to the matrix $\mathbf{A}$ Finally, the matrix $\mathbf{A}$ is normalized such as each element A_{ij} is divided by the sum of the entries in the whole column i, including the number from "zero" row. This normalized matrix $\mathbf{A}$ represents the probability of migrations from bin i to bin j and it enters in this form the unfolding. An example of a normalized migration matrix (probability matrix) is shown in Fig. 9.2.

- The term $\mathcal{L}_{reg}$ is responsible for the *regularization*. In TUnfold the Tikhonov regularization [120] is employed. This procedure aims to suppress unphysical high frequency terms which may appear in the unfolded distribution $\mathbf{x}$. Here, the parameter τ is the *regularization strength*, which governs the relative impact of $\mathcal{L}_{reg}$. The regularization strength is assumed to be constant during the minimization of the Lagrangian $\mathcal{L}$. The regularization strength is therefore determined before the minimization of $\mathcal{L}$ and for this analysis it is determined by minimization of the average global correlation coefficients [119]. This method chooses the regularization strength such that the global correlation coefficient (as defined in eq. 9.6) is minimal.

$$\rho_{glob.} = \sum_i \sqrt{1 - \frac{1}{(\mathbf{V}_{xx}^{-1})_{ii} (\mathbf{V}_{xx})_{ii}}}, \tag{9.6}$$

  where i runs over all components of the vector $\mathbf{x}$.

  The values of the regularization strengths estimated for all the measurements in this thesis are listed in Appendix B. The values of the regularization strengths for this analysis are generally small ($\sim 10^{-4}$), which means that the unfolded data $\mathbf{x}$ is not artificially pushed towards the bias vector $\mathbf{x_0}$ (which is chosen to be the distribution of the quantity $\mathbf{x}$ on the $t\bar{t}$ signal MC generator level).

  The f_b is a normalization factor. For this analysis the normalization factor was chosen to be equal unity.

  The matrix $\mathbf{L}$ represents the regularization conditions. In the TUnfold it is possible to choose unity, first and second derivative regularization conditions. For this analysis the second derivative regularization conditions which aim to smooth high frequency fluctuations were chosen.

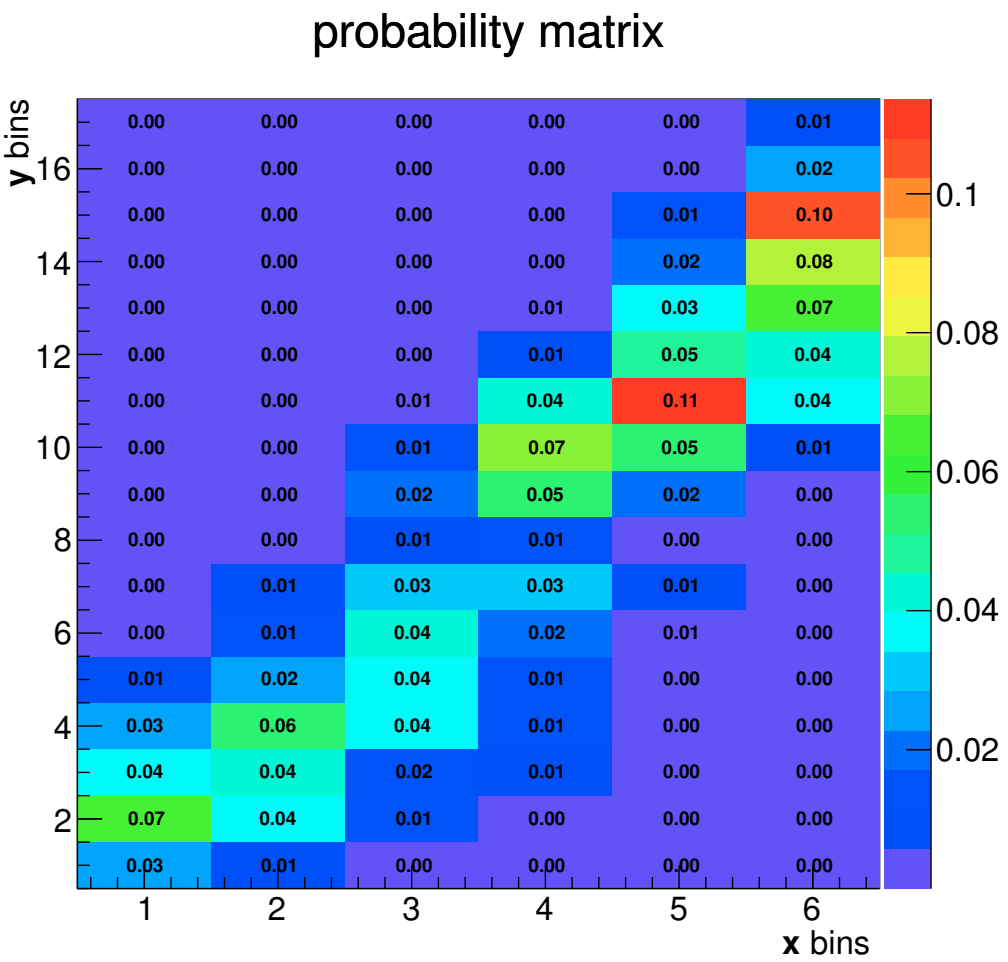

Figure 9.2: Normalized migration matrix $\mathbf{A}$ (probability matrix) in bins of $p_T(t)$. The generator binning corresponds to the $p_T(t)$ bins [0, 65, 125, 200, 290, 400, 2000] GeV and the detector binning corresponds $p_T(t)$ bins ([0, 30, 65, 90, 125, 150, 175, 200, 210, 240, 290, 360, 400, 450, 510, 700, 900, 2000] GeV). The matrix is obtained from the MadGraph + Pythia6 $t\bar{t}$ signal sample. This matrix is produced for illustration and is not used as it is in the analysis.

- The term $\mathcal{L}_{area}$ denotes the area constraint with λ being a fitted parameter. This part of the Lagrangian is optional and can be turned off. If the area constraint is on, then the resulting distribution $\mathbf{x}$ (corrected for the efficiencies $\mathbf{e}$, where $e_j = \sum_i A_{ij}$) is normalized in such a way that is is forced to match the total number of events $\sum_i y_i$. For this analysis the area constraint was not used.

TUnfold allows the unfolding of multidimensional distributions. The distributions are automatically mapped to one-dimensional vectors and the minimization is performed as explained above.

The covariance matrix which is used for the error estimation of the unfolded distribution, is also provided by the TUnfold as a propagation of the covariance matrix of the measured distribution $\mathbf{y}$, $\mathbf{V}_{yy}$:

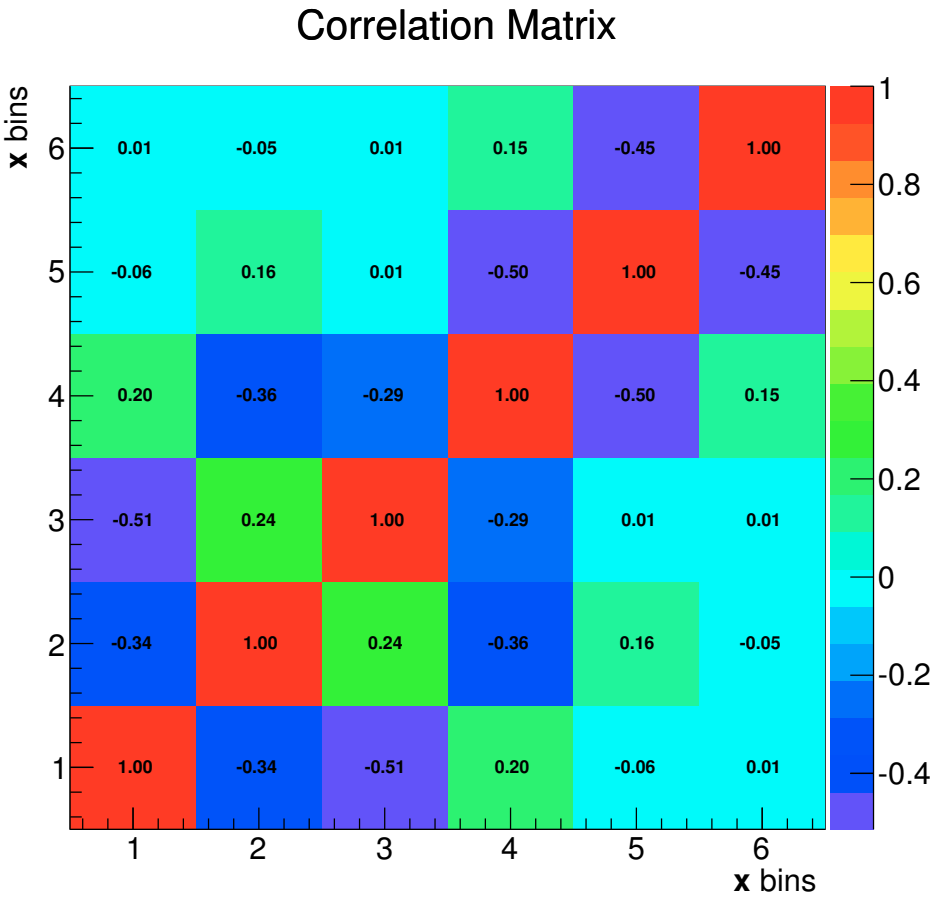

Figure 9.3: Correlation matrix $\mathbf{V}_{xx}$ for the unfolded $\mathbf{x}$ distribution which corresponds to the $p_T(t)$. The binning is the following: [0, 65, 125, 200, 290, 400, 2000] GeV. This matrix is a result of the measurement of the $p_T(t)$ spectrum in this analysis. More details on this measurement are given in Sec. 11.2.

$$\mathbf{V_{xx}} = \mathbf{D}^{xy}\mathbf{V}_{yy}(\mathbf{D}^{xy})^T, \quad \text{where} \tag{9.7}$$

$$(\mathbf{D}^{xy})_{ij} = \frac{\partial x_i}{\partial y_j}. \tag{9.8}$$

An example of the normalized covariance matrix $\mathbf{V}_{xx}$ (also called a *correlation matrix*) is shown in Fig. 9.3.

As an alternative and for comparison, the **S**ingle **V**alue **D**ecomposition (SVD) unfolding [121] (which has been used for the measurement of the single differential $t\bar{t}$ production cross sections at 8 TeV in the non-boosted regeme [35]) was also used for this analysis. The comparison of the performance of the SVD and the TUnfold unfolding tools are presented in Appendix A.

The exact details on how and in which binning each distribution is unfolded in this analysis are presented in Appendix C.

9.3 Cross Sections Determination

The $t\bar{t}$ production cross sections were measured using all events which fulfilled the selection criteria described in Chapter 6 and have a solution of the kinematic reconstruction (see Chapter 7) from the number of unfolded events corrected by the efficiencies and $t\bar{t}$ branching ratios. The integrated and differential cross sections are measured in the boosted region.

The integrated cross sections are measured as a function of the $p_T(t)$ minimal threshold using the following definition:

$$p_T(t) > p_T^{thr.} : \sigma_{p_T^{thr.}} = \frac{N^{signal\ unfold.}_{data,\ p_T^{thr.}}}{\epsilon \cdot BR \cdot L}, \tag{9.9}$$

where $p_T^{thr.}$ is the value of the minimum $p_T(t)$ threshold, $N^{signal\ unfold.}_{data,\ p_T^{thr.}}$ denotes the number of events in data as defined in eq. 9.1 and after the unfolding procedure[1] in the region with $p_T(t) > p_T^{thr.}$, ϵ is the analysis efficiency, BR is the branching ratio of the $t\bar{t}$ dilepton decay mode and L denotes the luminosity recorded by the CMS detector (19.7 fb^{-1}).

The $t\bar{t}$ differential cross sections are defined as follows:

$$\text{For } p_T(t) > 400 \text{ GeV} : \frac{d\sigma}{dx_i} = \frac{N^{signal\ unfold.}_{data,i}}{\Delta x_i \cdot \epsilon_i \cdot BR \cdot L}, \tag{9.10}$$

where x is a variable in bins of which the differential $t\bar{t}$ cross section is calculated, i is the number of the bin in the x distribution and Δx_i is the width of this bin. The number $N^{signal\ unfold.}_{data,i}$ denotes the same as for the eq. 9.9, but only on the i^{th} bin of the variable x.

The cross sections are measured in the *full phase space*. In this case, no restrictions on the lepton, jet and top kinematics are applied on the generator level. This corresponds to the following efficiency ϵ_i determination:

$$\epsilon_i = \frac{N^{signal\ unfold.}_{data,i}}{N^{total}_{gen,i}}, \tag{9.11}$$

where $N^{total}_{gen,i}$ is the total number of generated events in some bin or region i.

This phase space choice relies on the MC model, which extrapolates the measurements to the region of the phase space, which is outside acceptance.

The ratio $\frac{N^{signal\ unfold.}_{data,i}}{\epsilon_i}$ is obtained directly from the unfolding due to the construction of the migration matrix which enters the unfolding.

9.4 Efficiency, Purity and Stability

The effects of interbin migrations can be quantified using two variables – *purity* and *stability*. These quantities are determined from them signal MC sample and are used to check the performance of the reconstruction, unfolding and the binning choice.

[1] $N^{signal\ unfold.}_{data,\ p_T^{thr.}}$ corresponds to the $\mathbf{x}$ distribution, which is the result of the unfolding, if the input $\mathbf{y}$ distribution is filled with events defined in eq. 9.1.

The purity is a measure of the migrations into a certain bin of some variable and is defined as follows:

$$p_i = \frac{N_i^{gen\ and\ reco}}{N_i^{rec}}, \tag{9.12}$$

where $N_i^{gen\ and\ reco}$ is the number of the events, which were generated and reconstructed in the same bin i and N_i^{rec} is the total number of events, which were reconstructed in bin i. The closer the purity approaches unity, the less events have migrated into the bin i from the other bins.

The stability is defined as follows:

$$p_i = \frac{N_i^{gen\ and\ reco}}{N_i^{gen\ and\ reco\ anywhere}}. \tag{9.13}$$

Here $N_i^{gen\ and\ reco\ anywhere}$ is the total number of reconstructed events that were also generated in bin i.

The stability is a measure of the events, which migrate out of a certain bin to other bins and the closer it is to unity, the less events migrate out of the bin.

Another quantity, which is mostly used to evaluate the loss of events due to the set of selection and reconstruction procedures applied in the analysis, is called *efficiency*. The efficiency is defined the following way:

$$\epsilon_i = \frac{N_i^{reco}}{N_i^{gen}}. \tag{9.14}$$

The efficiency shows how many events are reconstructed in a certain bin i compared to the number of the generated events in this bins. However, the efficiency defined this way doesn't distinguish between events that were migrating into a bin and events originating from the bin.

For each distribution, which is used for the cross section measurement, efficiencies, purities and stabilities were checked. The respective plots are shown in the Appendix D. The purities for all the distributions are not dropping lower than 20% for all the measurements, which are performed in this analysis, which means that the binning is choice is reasonable. In general, the binning for all the resulting distributions in this analysis is chosen such that the unfolding performs reasonably well and does not collapse. The most problematic spectra to unfold are those, where the difference in statistics between two neighboring bins of the distribution is very large. That is why the binning was chosen to avoid extreme drops in statistics.

9.5 Combination of the ee, $e\mu$ and $\mu\mu$ Decay Channels

The equations 9.9 and 9.10 are written as if there is only one decay channel in which the $t\bar{t}$ system is reconstructed. However, in case of this analysis, the $t\bar{t}$ pair is reconstructed in three decay modes: ee, $e\mu$ and $\mu\mu$. Thus, the equations 9.9 and 9.10 need further specifications of how they were employed for the combination of the three dileptonic $t\bar{t}$ final states.

The selection and reconstruction, as described in Chapter 6 and Chapter 7, are done separately for each decay channel. The number of selected and reconstructed events in ee, $e\mu$ and $\mu\mu$ channels are added together for the unfolding:

$$\mathbf{y} = \mathbf{y}^{ee} + \mathbf{y}^{e\mu} + \mathbf{y}^{\mu\mu}, \tag{9.15}$$

where $\mathbf{y}$ is the vector of measured distribution, which enters the unfolding minimization (see Sec. 9.2.1).

Thus, the number of unfolded events, which enters the cross section determination (see Eq. 9.9 and Eq. 9.10), includes all three $t\bar{t}$ decay channels, which are considered for this analysis. The branching ratio in the Eq. 9.9 and Eq. 9.10 is taken as a sum of branching ratios for ee, $e\mu$ and $\mu\mu$ decay channels of the $t\bar{t}$ system:

$$BR = BR^{ee} + BR^{e\mu \text{ and } \mu e} + BR^{\mu\mu}. \tag{9.16}$$

10 | Systematic Uncertainties

An important source of systematic uncertainties is the limited knowledge about detector resolutions and efficiencies. Additionally, the assumptions, which are used in the MC models may influence the final results of the analysis and have to be taken into account in the systematic uncertainty assessment.

This chapter gives a brief overview of the systematic variations considered in this work. It closely follows the procedures described in [104]. The systematic uncertainties are performed by appropriate variations of the MC simulation samples. Each time a certain assumption is varied, the whole analysis is repeated for the variation. The difference between the nominal result and the result with the varied assumption is counted as a systematic uncertainty.

For the experimental uncertainties usually a variation of the corresponding scale factors is done. The model systematics are usually determined by generating separate MC samples with different assumptions and using them in the analysis flow instead of the nominal signal MC sample.

In this chapter it is also discussed how the separate sources of the systematics are combined ino one total uncertainty.

10.1 Experimental Uncertainties

The following variations of the scale factors and correction factors were considered in this work:

- **Trigger efficiency**: the scale factors related to the trigger efficiencies are described in Sec. 6.2. These scale factors are varied up and down within their estimated statistical uncertainties.

- **Pileup correction**: the vertex multiplicity, to which the pileup correction is related (see Sec. 6.4), is determined from the instantaneous luminosity multiplied by the total pp inelastic cross section. The latter was varied for the systematic uncertainties determination by $\pm 5\%$ and the reweighting of the MC primary vertex distribution was repeated for each variation.

- **Jet energy scale**: the jet energy was calibrated as described in Sec. 5.4.3. The uncertainties related to these correction are estimated in [101]. The correction factors were varied in the MC within the estimated uncertainties.

- **Jet energy resolution**: the jet energy resolutions (JER) in MC are corrected to match the ones in data by $|\eta|$ dependent factors [104] and are varied up and down for the systematics determination by the values of their uncertainty.

- *b*-**tagging**: the scale factors related to the *b*-tagging efficiencies are estimated by a range of variations within the estimated uncertainties of the scale factors [122]. The variations are done in many dimensions, depending on the p_T, η and flavour of the jets [104].

  The median point is defined in $|\eta|$ ($|\eta|^{median} = 0.75$) and in p_T ($p_T^{median} = 65$ GeV for *b*- and *c*-jets and $p_T^{median} = 45$ GeV for ligh jets[1]) of the jets. Then the separate variations of the *b*-tagging scale factors are done depending on $|\eta|$ and p_T. The variation up was done by increasing the *b*-tagging scale factors by their uncertainty value if the p_T (or $|\eta|$) are larger than the median value and by decreasing the scale factors by their uncertainties if the p_T (or $|\eta|$) are smaller than the median. The variation down is done with the inverse logics - if the p_T (or $|\eta|$) are larger than the median, the scale factors are decreased and vise versa. The results of the variations depending on p_T and $|\eta|$ of the jets are added in quadrature. The variation for *b*- and *c*-jets is done simultaneously and the variation for light jets is done separately. The resulting cross section variations are also added in quadrature.

- **Kinematic reconstruction**: the uniform scale factors related to the efficiency of the $t\bar{t}$ kinematic reconstruction (See Sec. 7.4) are varied up and down by ($\pm 1\%$), which is of the order of the statistical uncertainties of those scale factors.

The varied scale factors were applied to background and to signal MC samples simultaneously.

10.2 Model Uncertainties

The model uncertainties considered in this analysis are related to the modeling of the $t\bar{t}$ signal process. In the nominal analysis flow, the MadGraph + Pythia generator combination with CTEQ6 PDF set was used for the signal process simulation with LO accuracy. The tested assumptions of the simulation are described in the following.

10.2.1 PDF Variations

The $t\bar{t}$ signal sample is reweighed according to the errors of the 44 eigenstates of the CTEQ66 PDF sets [79]. Each time when the measured cross section after variation is larger than the nominal one, it is assumed to be a variation with a "+" sign. If the measured cross section after variation has a lower value than nominal it is assumed to have a "-" sign. All the variation with the same sign are added in quadrature. Finally the systematic uncertainty related to the PDF is presented as two numbers, one with a "+" sign and the other with a "-" sign.

[1] A *light jet* refers to the jets originating from an *u*, *d* or *s* quark or from a gluon.

10.2.2 Hard Process Model

The hard process of the $t\bar{t}$ signal production is alternatively modeled exploiting the POWHEG matrix element generator with NLO accuracy. POWHEG is also interfaced with PYTHIA for the showering and hadronization modeling, thus the difference compared to the nominal signal model is only in the hard process simulation. The difference in the results compared to the nominal MADGRAPH sample is assumed as the hard process model systematic uncertainty.

10.2.3 Hadronization Model

Unlike for the previous case, to see the effect caused by the hadronization model one needs to keep the same matrix element generator, but apply different hadronization models. This is done by comparing the results of two samples, which are generated with POWHEG + PYTHIA and POWHEG + HERWIG. The difference in the results obtained with using these two samples is assumed to be the systematic uncertainty from the hadronization model.

10.2.4 Matching Scale Variation

The scale at which the hard scattering process is matched to the showering is assumed to be 20 GeV for the nominal analysis flow. This so-called matching scale is varied up and down by a factor two.

10.2.5 Hard Scale Variation

The assumed model parameter called "hard scale", which corresponds to the renormalization scale of the QCD, is varied by factors 2 and 1/2. The changes of the results are the corresponding systematic uncertainties.

10.2.6 Top Mass Variation

For the nominal simulated sample the mass of the top quark is assumed to be 172.5 GeV. The variation of ± 1 GeV should account for the systematic uncertainties related to the mass assumption.

In general, the deviations of the cross section results when using a $\bar{t}$ MC with different op mass grow linearly with the increase of the deviation of the mass (see Fig. 10.1). To reduce the statistical effects due to the limited statistics of the simulated signal samples that were generated with different masses, the systematics was measured for the two samples with the top mass varied by ± 6 GeV and the difference in the results compared to the nominal ones is divided by 6. This is assumed to be the systematic variation due to the top mass assumption.

10.3 Additional Sources of Systematics

There are also some sources of systematics, which are accounted for in this work, but they can't be referred neither to experimental uncertainties in a sense of variation of the scale

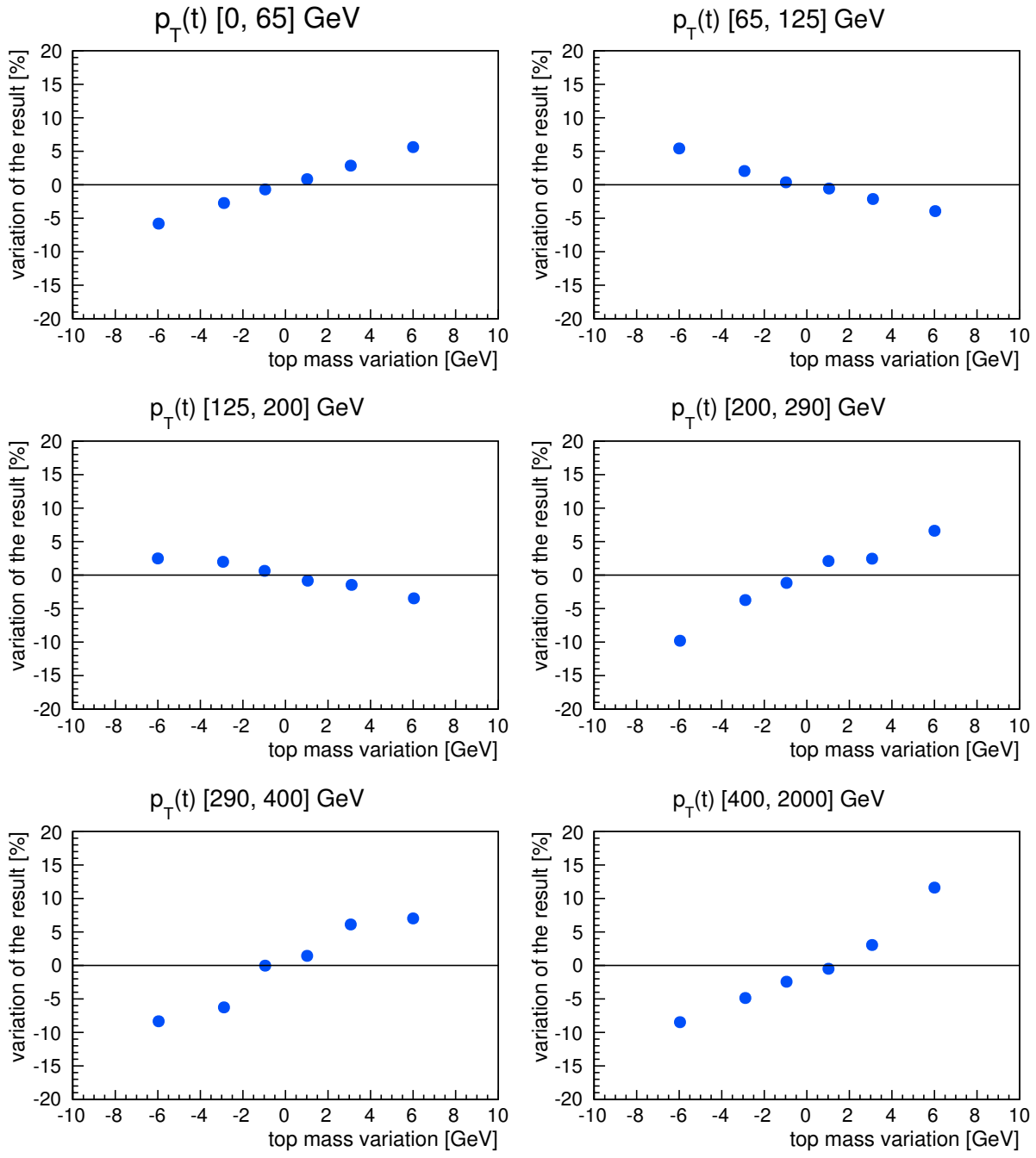

Figure 10.1: Value of corresponding systematic variation on the cross section as a function of the top mass variation for different bins of $p_T(t)$.

factors, nor to the model uncertainties related to the modeling of the $t\bar{t}$ signal process. These are the following uncertainties:

- **Luminosity**: the luminosity value used for the cross section determination is varied within its uncertainty of 2.6%. The uncertainty of the measured luminosity is estimated using the pixel counting algorithm [123].
- **Background rates variation**: all the yields of the background processes are normalized to the data luminosity. The normalization factors of all the backgrounds, except for Drell-Yan, are varied simultaneously by ±30%. The normalization of the Drell-Yan yields is varied separately, also by ±30%. These conservative variations should account for both absolute cross section and shape uncertainties.

10.4 Total Systematic Uncertainty

The total systematic uncertainty is determined from the different separate uncertainties described above in a way, which was already discussed for the PDF variations (see Sec. 10.2.1). All the variations from different sources with "+" sign and with "-" sign are separately added in quadrature. The total systematic uncertainty is presented with the two numbers: δ^{+}_{syst} and δ^{-}_{syst}. The result should then be interpreted as following:

$$\text{Result} = \text{Result}_{nominal} \; {}^{+\delta^{+}_{syst}}_{-\delta^{-}_{syst}}. \tag{10.1}$$

11 | Results

Various measurements and checks were performed using the selection and reconstruction optimized for the analysis of the $t\bar{t}$ production in the boosted regime (see Chapter 6 and Chapter 7).

This chapter presents the measurements of the integrated cross sections as a function of the minimum threshold on $p_T(t)$ and the differential $t\bar{t}$ production cross sections in the boosted region ($p_T(t) > 400$ GeV) as functions of six different variables describing the t and $t\bar{t}$ dynamics.

Additionally, measurements in the high $M(t\bar{t})$ region were performed to investigate some discrepancies between data and models observed in the double differential $t\bar{t}$ production cross section measurements [9]. In this analysis it was possible to make a statistically reasonable measurements in mass regions higher than what was presented in [9].

All the distributions presented in this chapter are counting only the top quark and not the antitop to avoid the problem of statistical correlation between both (double counting of event).

The measured cross sections can be visually compared to the different models: MadGraph+Pythia, Powheg + Herwig, Powheg + Pythia, MC@NLO + Herwig.

All the detector level control distributions, which correspond to the variables, in which the cross sections were measured, are shown in Appendix G.

11.1 Consistency Checks

The single differential cross section for the $t\bar{t}$ production have been measured previously for the 8 TeV data sample collected with the CMS detector [35, 104]. The measurements were performed in the non-boosted regime and the selection was optimized for it: the lepton isolation requirements were harder and no event cleaning was applied.

The single differential cross sections in bins of $p_T(t)$ and $M(t\bar{t})$ measured with the framework developed for this analysis can be compared with the previously published results. This is a test of the whole analysis chain and no significant discrepancies are expected to occur.

Fig. 11.1 presents the ratio of the normalized $t\bar{t}$ production cross sections[1] in bins of $p_T(t)$ measured in this analysis over the ones published before [35, 104]. The ratio is in agreement with unity, which confirms the consistency with the previous cross section measurements.

The cross sections in bins of $p_T(t)$ were previously measured up to 400 GeV [35, 104]. This measurement is shown in Fig. 11.2 compared to the measurement obtained in this

[1] The *normalized differential cross sections* are normalized to the inclusive $t\bar{t}$ production cross section.

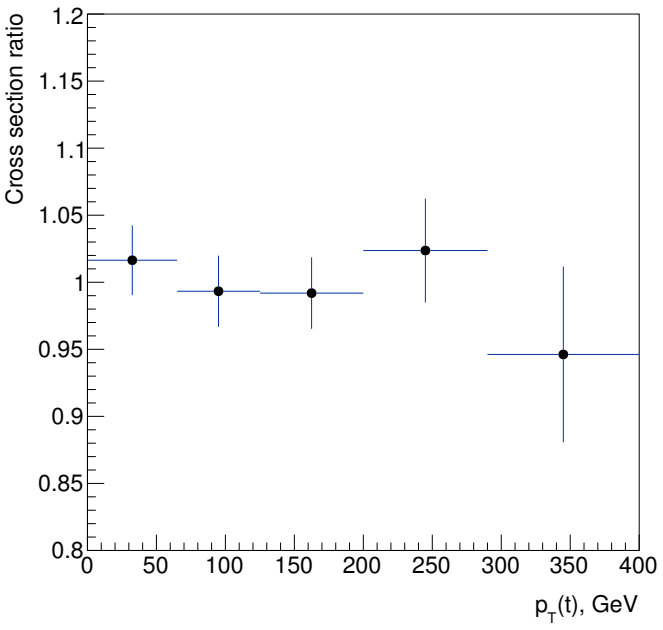

Figure 11.1: The ratio of the $t\bar{t}$ production cross sections in bins of $p_T(t)$ measured in this analysis over the ones measured in [35, 104]. The error bars represent the statistical uncertainties of the cross sections measured in this analysis.

analysis. With the new measurement one is able to make a reasonable measurement for the $p_T(t)$ above 400 GeV. The tendency that the different theories predict a harder $p_T(t)$ spectrum than the measured one is confirmed above 400 GeV. The worst description of the data is provided by LO predictions from MadGraph + Pythia.

The total integrated cross section measured in this analysis in the boosted region ($p_T(t) > 400$ GeV) was compared to analogous measurement based on the semileptonic $t\bar{t}$ decay channel [118]. In the latter measurement only the top quarks decaying hadronically were reconstructed in the boosted regime. These top quarks were reconstructed as single top "fat jets". The measurement [118] resulted in the following:

$$\sigma^{\text{semileptonic}}_{p_T>400\text{ GeV}}(t\bar{t}) = 1.44\pm0.10\text{ (Stat.}\oplus\text{Syst.)}\pm0.13\text{ (PDF)}\pm0.15\text{ (Hard Scale)}\pm0.04\text{ (lumi) pb}, \tag{11.1}$$

where the uncertainties are subdivided to the PDF and hard scale variations, luminosity and statistical and other experimental systematic uncertainties.

In this analysis the integrated $t\bar{t}$ production cross section in the region with high transverse momenta of the top quarks ($p_T(t) > 400$ GeV) resulted in the following:

$$\sigma^{\text{dileptonic}}_{p_T>400\text{ GeV}}(t\bar{t}) = 1.12 \pm 0.09\text{ (Stat.)}\,^{+0.08}_{-0.10}\text{ (Syst.) pb}, \tag{11.2}$$

where the total systematic uncertainties include the contributions from all the sources described in Chapter 10. These contributions are summarized in the first $p_T(t)$-threshold column $p_T(t) > 400$ GeV in Table 11.1. Compared to the integrated $t\bar{t}$ production cross section measured in the semileptonic decay channel, the contribution of the PDF variations

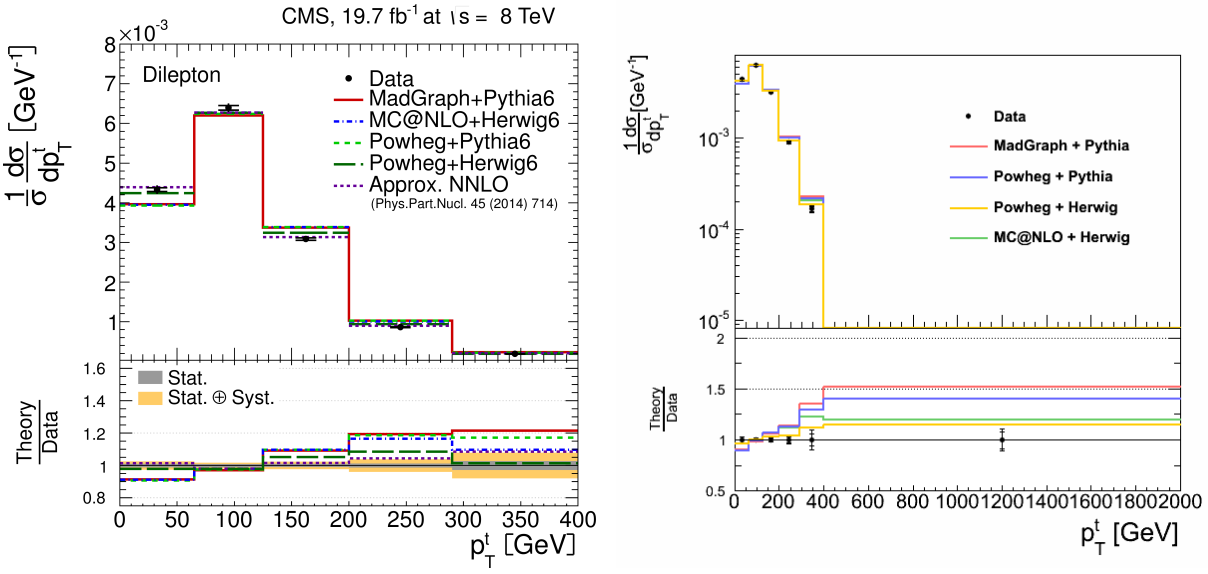

Figure 11.2: Normalized bin averaged differential $t\bar{t}$ production cross sections as a function of $p_T(t)$. The left plot represents the results from [35, 104]. Here, the inner (outer) error bars indicate the statistical (combined statistical and systematic) uncertainties. The measurements are compared to predictions from MadGraph+Pythia, Powheg+Pythia, Powheg + Herwig and MC@NLO + Herwig. The lower part of the plot shows the ratio of the predictions to data. On the right plot the cross sections obtained in this analysis are presented. The inner error bars correspond to the statistical only uncertainties and the outer error bars are the combined statistical and systematic uncertainties. The following predictions are compared to the data: MadGraph + Pythia (red line), Powheg + Pythia (blue line), Powheg + Herwig (orange line) and MC@NLO + Herwig (green line). The ratio of the predictions over the measured cross sections are shown in the bottom panels with the error bands (left) and bars (right) corresponding to the measurement uncertainties.

is very small. The reason for it is likely related to the procedure of estimation of the uncertainties related to the PDF variations in the measurement with the semi-leptonic channel: there one relies on MC predictions for the estimation of the PDF uncertainty by comparing the integrated cross sections derived in the nominal MC sample and in the samples with varied PDFs. However, the PDFs are not precisely estimated for the region with high momenta of the top quarks (and high values of x, respectively) and thus their variations are large.

In general, the integrated $t\bar{t}$ production cross section in the boosted region measured utilizing the dileptonic final state is somewhat lower than for the semileptonic case, however both cross sections are in agreement within their uncertainties.

11.2 Cross Sections as Functions of $p_T(t)$ Thresholds

The Fig. 11.3 shows the total cross section as a function of the minimum $p_T(t)$ threshold. This measurement is done performing a simultaneous unfolding of the $p_T(t)$ and

$M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ spectra (for the details of the binning used for the unfolding see Appendix C). The unfolding in $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ is performed only in one bin on the generator level and four bins on detector level, which means that is is not an actual unfolding, but rather a weighted averaging of the cross section measurement from the four different $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ regions (four bins on the detector level). A comparison of the results with and without application of the cleaning conditions and with doing the simultaneous unfolding of the $p_T(t)$ and $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ spectra is presented in Appendix F. The numerical values for the cross sections, presented in Fig. 11.3, are summarized in the table 11.1, in which also all the systematic uncertainties as well as the cross sections predicted by the various models are shown.

The cross sections in the predictions are falling off with increasing $p_T(t)$ threshold less steeply than the data. This trend was also observed before in the single- [35, 104] and double differential [9] cross sections. The worst agreement is observed between the measurements and MadGraph+Pythia predictions and the best description is provided by Powheg+Herwig. The cross sections rapidly drop to zero with increasing threshold on the top quark transverse momentum, showing a small number of events in the highly boosted region.

The measurement was also performed separately for every $t\bar{t}$ decay channel (see Fig. 11.4). The measurements of each cross section point in the three different channel were tested to be consistent with a constant value with a χ^2 goodness of fit test. For almost all the cross sections the corresponding χ^2 probability is above 5%, from which one can conclude that the results in the ee, $e\mu$ and $\mu\mu$ channels are consistent. For example, for the cross section above 400 GeV the test results in a χ^2 of 3.9 for two degrees of freedom, which corresponds to a χ^2 probability of $\sim 14\%$; for the cross section above 550 GeV the χ^2 probability is $\sim 28\%$. Only for the cross section above 500 GeV the χ^2 probability is $\sim 3\%$.

The measurement is compared to the previous measurements of $t\bar{t}$ production in boosted regime from CMS [118] and ATLAS [124] experiments using the semi-leptonic channel. The CMS measurement in the semi-leptonic channel is discussed in the previous chapter and also in the ATLAS measurement one reconstructs in the boosted region only the hadronically decaying top quark. Both measurements (from CMS and from ATLAS) provide differential cross sections as a function of $p_T(t)$ in the boosted region. To make a comparison to the present measurement, the differential cross sections from [118] and [124] are integrated above a certain bin, assuming the errors in all the bins which are integrated as fully uncorrelated. A comparison is presented in Fig. 11.5. The measurements presented in this analysis are well in agreement with the ATLAS results in the semi-leptonic channel. The cross sections measured with the CMS semi-leptonic channel tend to be somewhat higher. The overall precision of the measurement, which is presented in this analysis, is compatible to the previous measurements for the first three $p_T(t)$ thresholds ($p_T(t) > 400$ GeV, $p_T(t) > 450$ GeV, $p_T(t) > 500$ GeV). This indicates that despite the lower branching ratio of the dileptonic channel, and thus lower statistics, the results have a high precision in the boosted region due to the specifically developed reconstruction strategy - the reconstructed event sample is clean (has a small background fraction).

$p_T(t)$ [GeV]	> 400	> 450	> 500	> 550	> 600	> 650
σ [pb]	1.115	0.539	0.254	0.087	0.036	0.008
MADGRAPH [pb]	1.645	0.868	0.47	0.258	0.15	0.085
MCATNLO [pb]	1.41	0.733	0.393	0.211	0.112	0.063
POWHEG [pb]	1.557	0.826	0.442	0.245	0.143	0.082
POWHEGHERWIG [pb]	1.278	0.657	0.345	0.193	0.111	0.059
Source	Uncertainty [%]					
Statistical	8.3	11.6	17.6	42.5	74.7	272.9
Total Sys. Pos.	7.3	10	12.3	28	54.9	274.7
Total Sys. Neg.	8.9	10.1	9.2	16.5	43.4	206.2
Lumi Up	-2.8	-2.8	-2.9	-3.1	-3.4	-5.1
Down	2.9	3	3.1	3.3	3.6	5.4
b-tagging Pos.	1.4	1.4	1.4	0.9	0.9	0
Neg.	1.4	1.7	2.2	2.7	2.1	15.6
Kin Up	-0.4	-0.4	-0.4	-0.5	-0.5	-0.9
Down	0.4	0.4	0.4	0.5	0.5	0.9
JER Up	1.3	1.4	2.1	7.8	5	19.5
Down	-2.7	-3.1	-5.4	-11.1	-16.7	-79
JES Up	-1.9	-1.1	-0.5	6.7	9.5	31.2
Down	2.1	2.6	4.1	10.9	17.4	58.1
PU Up	0	-0.1	-0.2	-0.4	-1.3	0
Down	-0	0	0.1	0.6	1.3	-1.1
Trig Up	-1.2	-1.2	-1.2	-1.3	-1.5	-2.1
Down	1.2	1.2	1.2	1.4	1.5	2.1
Lept Up	-2.2	-2.3	-2.3	-2.6	-2.9	-4.4
Down	2.3	2.3	2.4	2.7	2.9	4.5
BG Up	1.6	1.3	0.5	-1	-4.3	-21.6
Down	-5.1	-4.9	-4.4	-3.5	-1.3	10.7
DY Up	0	0	0	-0.1	-0.1	-0.4
Down	-0	-0	-0	0.1	0.1	0.4
Mass Up	1.1	1.6	3.1	10	13.6	47.7
Down	-0.6	-0.4	-0.2	-2.8	-1.4	-10.3
Match Up	0.6	0.7	4	6.3	2.7	21.5
Down	-0.8	-2.8	-1	5.1	13.7	149.9
Scale Up	-3.1	-3.8	-2.4	18.4	22.5	106.6
Down	3.3	6.5	8.8	15.7	27.1	102.7
PDF Pos.	0.2	0.3	0.7	1.2	2.1	2.9
Neg.	0.2	0.3	0.7	1.2	2.2	3.2
Hadronization	-3.4	-3.8	1.7	5.2	11.4	22
Hard Scat.	2	3.5	2.2	8.6	37.7	186.8

Table 11.1: The $t\bar{t}$ production cross sections as a function of the $p_T(t)$ minimum threshold. The measured and predicted cross sections are presented. The statistical and systematic uncertainties for the measured cross sections are also given.

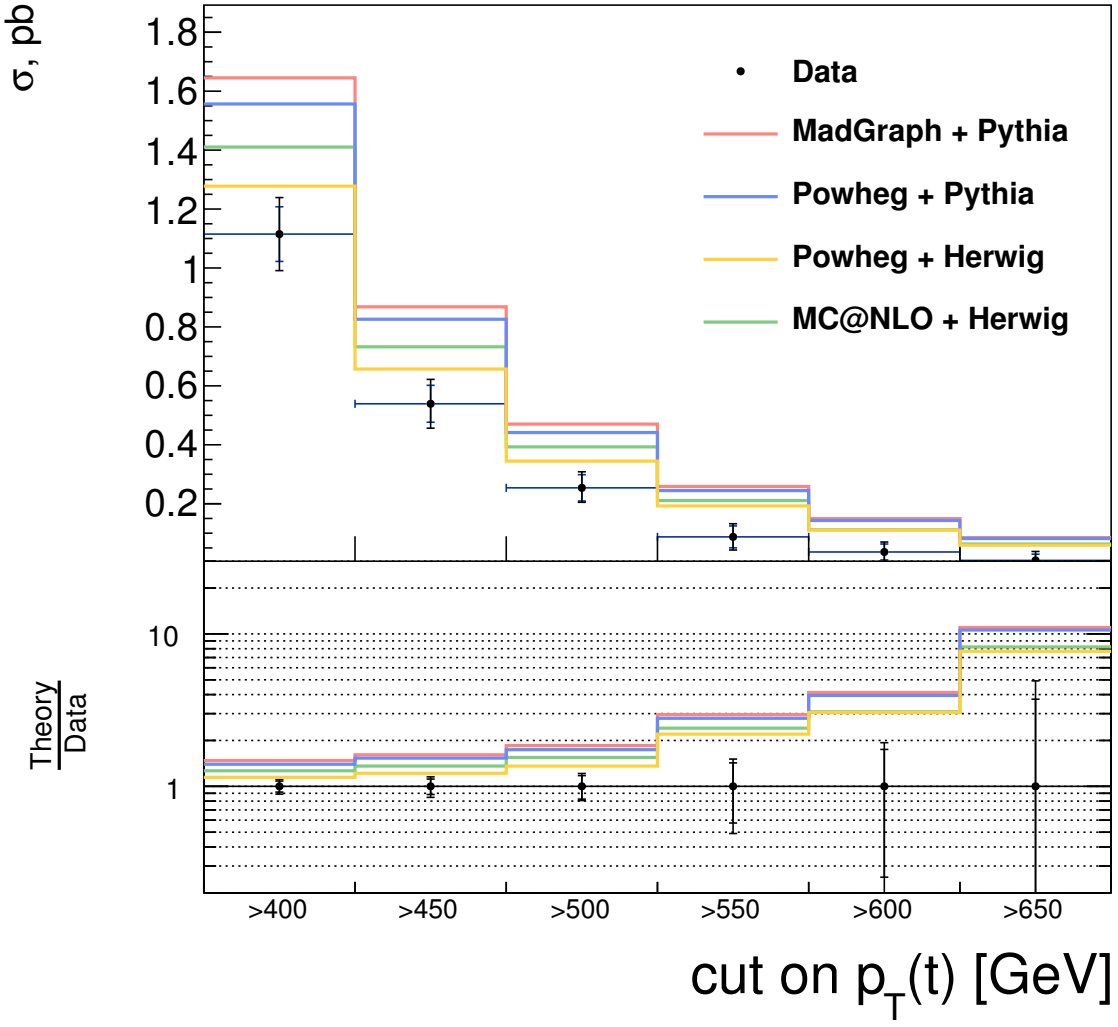

Figure 11.3: The $t\bar{t}$ production cross section as a function of the $p_T(t)$ minimum threshold. The black dots represent the data. The inner error bars correspond to the statistical only uncertainties and the outer error bars are the combined statistical and systematic uncertainties. The predicted cross sections from MadGraph + Pythia is represented with a red line, the cross sections from Powheg + Pythia are drawn with a blue line, Powheg + Herwig predictions are marked with an orange line and MC@NLO + Herwig is represented with a green line. The ratio of the predictions over measured cross sections are shown in the bottom panels with the error bars around the constant line at unity corresponding to the measurement uncertainties.

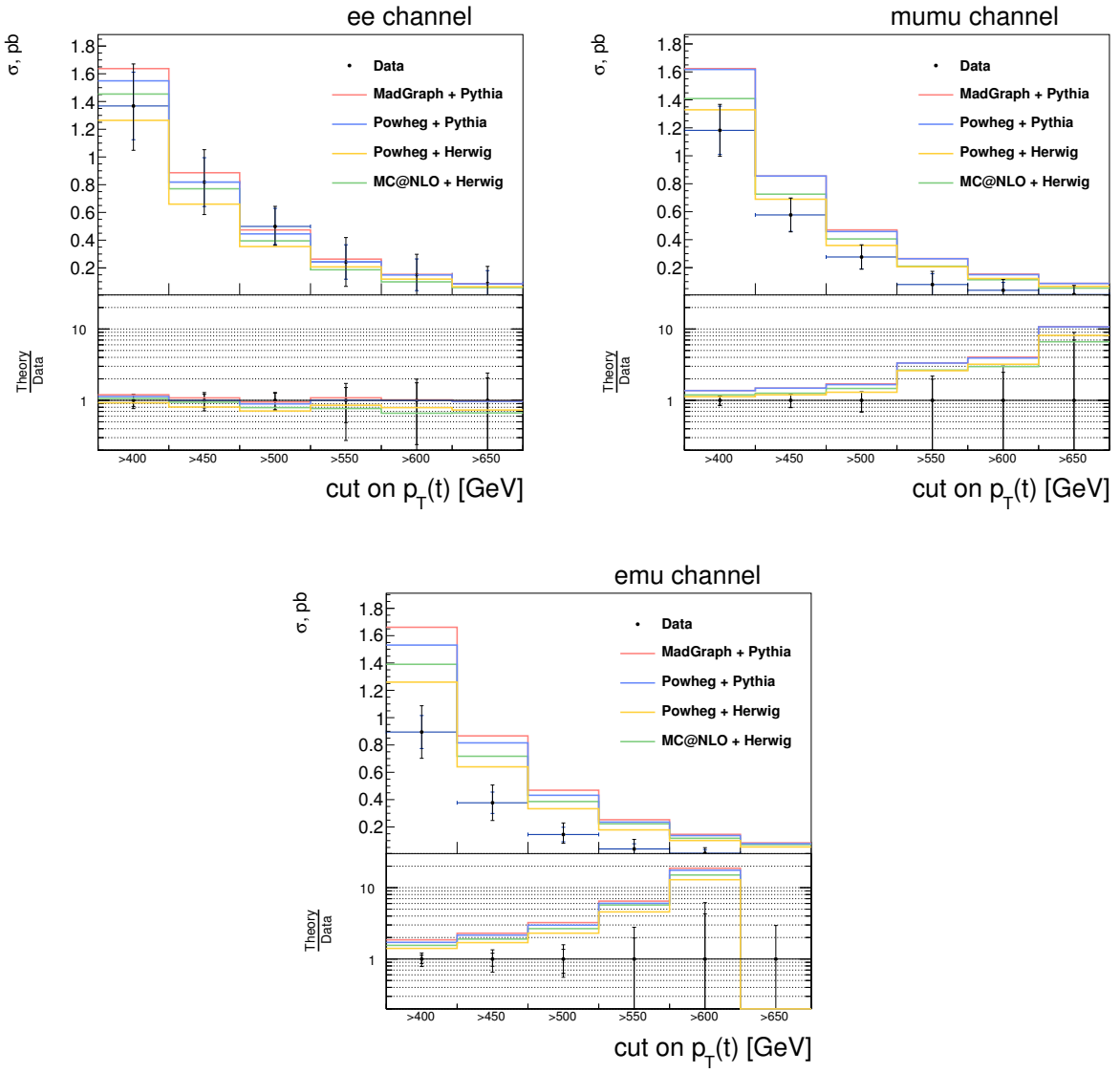

Figure 11.4: The $t\bar{t}$ production cross section as a function of the $p_T(t)$ minimum threshold in the ee (top left), $\mu\mu$ (top right) and $e\mu$ (bottom) $t\bar{t}$ decay channels. Other details as in Fig. 11.3.

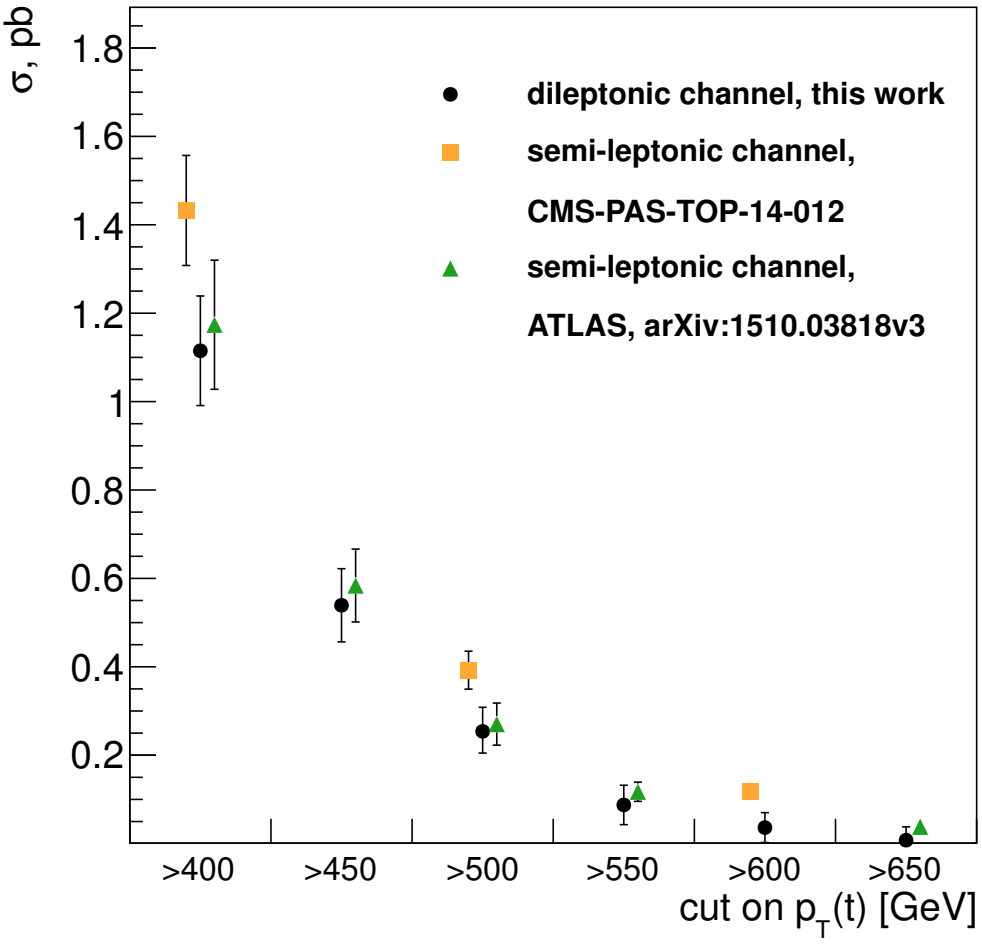

Figure 11.5: The $t\bar{t}$ production cross section as a function of the $p_T(t)$ minimum threshold. The black dots represent the results of the measurement described in this analysis, orange squares are the results of the integrating of the differential cross sections measured in [118] and the green triangles denote the results of integrating the differential cross sections from [124]. The error bars are the combined statistical and systematic uncertainties.

11.3 Differential Cross Sections in the Boosted Regime

Differential cross sections in bins of three different variables are measured in the boosted region ($p_T(t) > 400$ GeV). For these cross sections a simultaneous unfolding of the variables under study and the $p_T(t)$ is performed (see details of the unfolding in Appendix C). The simultaneous unfolding in $p_T(t)$ allows to obtain results in the region $p_T(t) > 400$ GeV, correctly taking to account the migrations from and to the regions with lower transverse momenta of the top quark.

The differential measurements in the boosted region provide a test of the Standard Model in a not well explored kinematic space. In particular, the boosted top quarks are produced with high-x gluons in the proton ($x \gtrsim 0.1$) and the PDFs are not precisely known at very high x values ($x > 0.2$). Thus, the measurement of the top quarks in the boosted regime can provide additional information which can be used for the PDF fits for the high x values.

In the remainder of this section, the cross section plots are presented. The cross sections in the lower p_T bins for the corresponding variables as well as the tables with the numerical cross section values including statistical and systematic uncertainties are presented in Appendix H. The corresponding correlation matrices are listed in Appendix J.

11.3.1 Cross Sections in Bins of $|y(t)|$

The cross sections in bins of the absolute rapidity of the top quark in the boosted region is presented in Fig. 11.6. This variable is sensitive to the PDFs and to the QCD dynamics of the hard process of the $t\bar{t}$ production. The worst disagreement is observed between the data and the LO MadGraph + Pythia predictions, while Powheg + Herwig gives the best description of the data. The overall tendencies and distribution shapes are similar in data and in simulation. However, the spectrum predicted by MadGraph + Pythia falls slightly steeper with increasing $|y(t)|$ than the other predictions.

11.3.2 Cross Sections in Bins of $|y(t\bar{t})|$

Fig. 11.7 shows the absolute rapidity of the $t\bar{t}$ system in the boosted region. The variable $y(t\bar{t})$ is highly sensitive to the PDFs (in the leading order can be expressed as $y(t\bar{t}) \sim \ln(\frac{x_1}{x_2})$).

In general the theory models predict more events with high $|y(t\bar{t})|$. The best agreement is provided by the Powheg + Herwig predictions and the worst description is observed between the data and MadGraph + Pythia.

11.3.3 Cross Sections in Bins of $\Delta\eta(t\bar{t})$

The variable $\Delta\eta(t\bar{t}) = |\eta(t) - \eta(\bar{t})|$ denotes the absolute pseudorapidity separation between the top and the antitop quark. This variable is largely independent of the PDFs and is sensitive to the QCD dynamics of the production process. The cross sections in bins of $\Delta\eta(t\bar{t}) = |\eta(t) - \eta(\bar{t})|$ for the high $p_T(t)$ are presented in Fig. 11.8. Several tendencies are observed. All the NLO predictions tend to show a larger separation in pseudorapidity

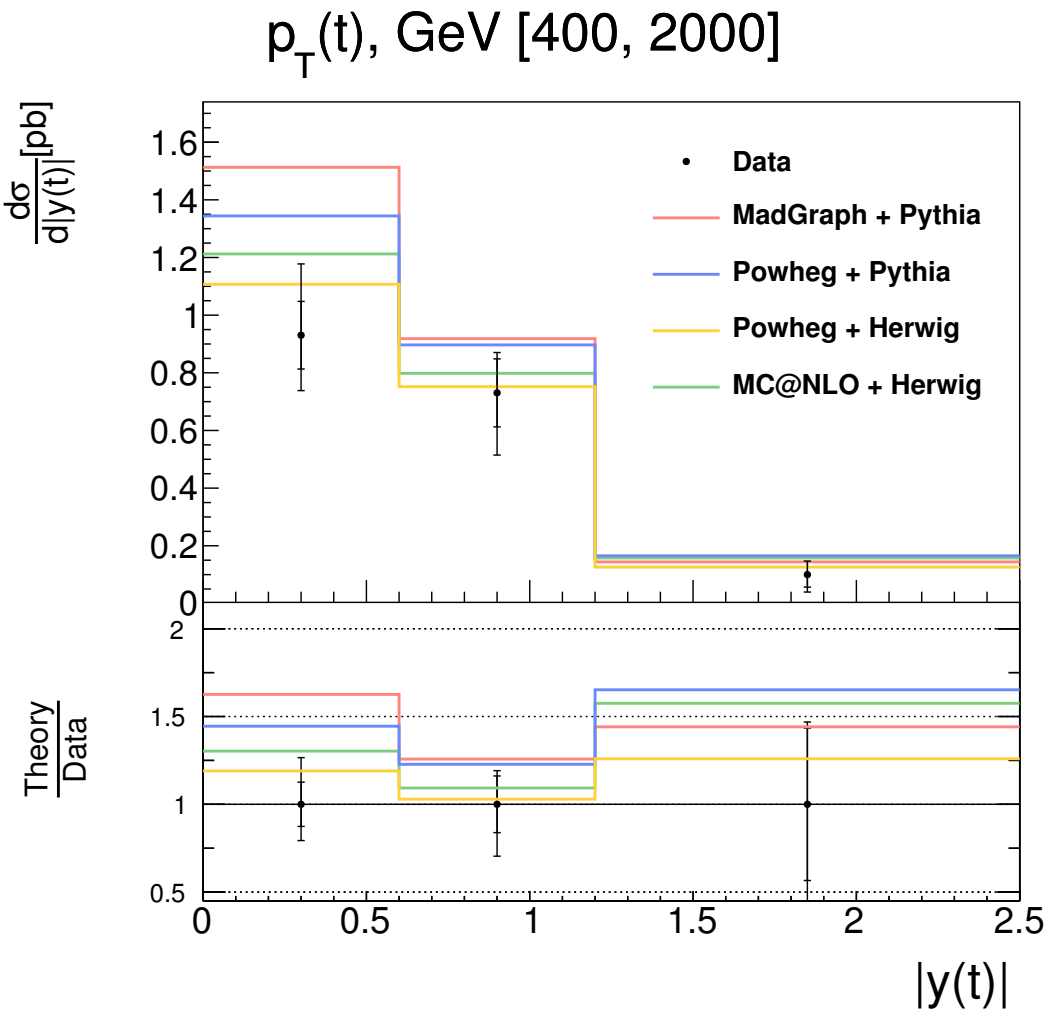

Figure 11.6: The $t\bar{t}$ production cross sections in bins of $|y(t)|$ in the high $p_T(t)$ region. For other details see Fig. 11.3.

between top and antitop compared to the data results. This effect is the strongest for POWHEG+PYTHIA. The LO MADGRAPH+PYTHIA predictions show the opposite effect.

11.4 Measurements as Functions of the $M(t\bar{t})$

The application of the 2D isolation (see Sec. 6.6.1) and employing the variable $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ for the reduction of migration effects either as a cleaning condition (see Sec. 8.2.1) or as an additional distribution for unfolding (see Appendix F) also provide an improvement of the measurements as a function of $M(t\bar{t})$ (see Appendix E). That is why a set of measurements depending on $M(t\bar{t})$ has been performed in this analysis.

The cross sections in bins of the invariant mass of the $t\bar{t}$ system with simultaneous unfolding of the $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ spectrum (for the details of the binning chosen for the unfolding see Appendix C) is presented in Fig. 11.9. The invariant mass of the $t\bar{t}$ system is highly sensitive to the PDFs (in the LO approximation $M(t\bar{t}) \sim \sqrt{x_1 x_2}$).

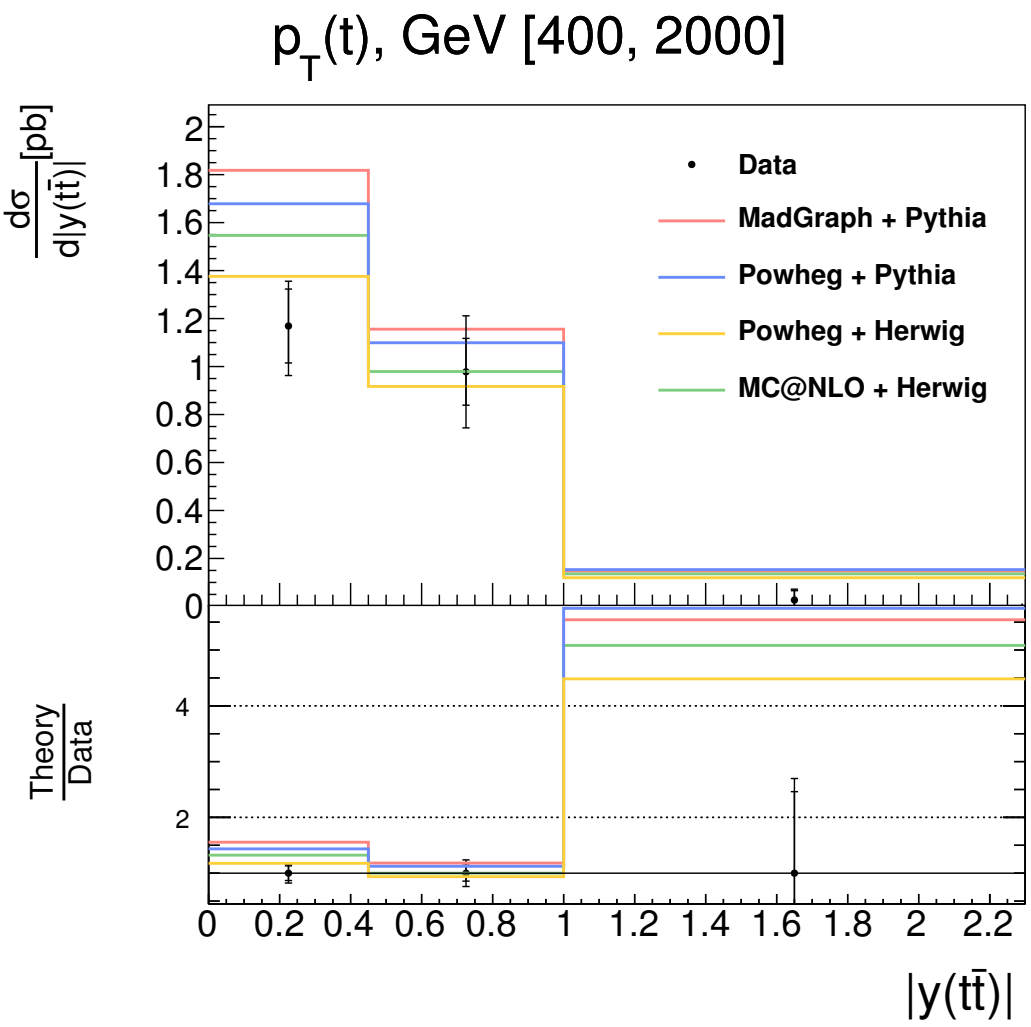

Figure 11.7: The $t\bar{t}$ production cross sections in bins of $|y(t\bar{t})|$ in the high $p_T(t)$ region. For other details see Fig. 11.3.

An overall good agreement between the measured results and the predictions is observed. The best agreement is provided by MadGraph + Pythia. An overall tendency is that the NLO models predict a slightly harder spectrum in $M(t\bar{t})$ compared to what is being observed in the data. The worst agreement between the data and the prediction points are observed for MC@NLO + Herwig. The corresponding numbers of the cross sections in data and predictions and values of systematic and statistical uncertainties of the measurements are presented in Table 11.2.

11.5 Double Differential Cross Sections in Bins of $M(t\bar{t})$

A number of double differential $t\bar{t}$ production cross section measurements in the $e\mu$ channel only has been performed previously [9] in bins of the invariant mass of the $t\bar{t}$ system. There were some trends observed for the highest bin of the $M(t\bar{t})$ (see Fig. 11.10):

- The models predict slightly more central production in $|y(t)|$. The worst agreement

$M(t\bar{t})$ [GeV]	340 to 380	380 to 470	470 to 620	620 to 820	820 to 1100	1100 to 2200
$\frac{d\sigma}{dM(t\bar{t})}$ $[\frac{pb}{GeV}]$	0.94	1.09	0.466	0.122	0.024	0.0012
Source	Uncertainty [%]					
Statistical	2.7	1	1.1	2.5	5.8	12.1
Total Syst. Pos.	10	7.7	8.7	9.3	20.7	21.1
Total Syst. Neg.	10.8	8.5	9.7	10.5	19.9	23.9
Lumi Up	-2.9	-2.6	-2.7	-2.7	-2.6	-2.9
Down	3.1	2.7	2.9	2.8	2.7	3
b-tagging Pos.	1.1	0.8	0.8	0.8	1	1.8
Neg.	1.5	0.7	0.7	1.1	1	1.4
Kin Up	-0.36	-0.32	-0.35	-0.36	-0.35	-0.39
Down	0.36	0.33	0.35	0.36	0.35	0.39
JER Up	0.7	0.2	0.1	0.5	-1.7	3.2
Down	-1.2	-0.9	-0.8	-0.9	-1.5	-1.5
JES Up	0.6	-3.1	-3.3	-1.4	-2.3	-2.5
Down	-0	3.4	3.1	1.5	0.2	4.1
PU Up	0.9	0.3	0.2	0.2	0.1	-0
Down	-0.9	-0.3	-0.2	-0.3	0.1	-0.2
Trig Up	-1.3	-1.2	-1.3	-1.3	-1.2	-1.4
Down	1.4	1.2	1.3	1.3	1.3	1.4
Lept Up	-2.3	-2.1	-2.2	-2.2	-2.1	-2.3
Down	2.4	2.1	2.3	2.2	2.1	2.4
BG Up	2	3.3	1.2	2.2	3.9	0.5
Down	-5.7	-5.6	-3.7	-5.2	-6.4	-4.2
DY Up	-1.3	-0.3	-0.1	-0.2	0.1	0
Down	1.5	0.3	0.1	0.2	-0.1	0
Mass Up	-4.3	-0.7	0	1	1.4	1.4
Down	3.5	0.9	-0.1	-1.8	-0.4	-3.4
Match Up	-0.2	0.4	0.1	0.5	7.9	-7
Down	4	-1.6	-1.4	0	-1.9	-5.4
Scale Up	-0.3	3.4	2	0.8	-0.8	-11.6
Down	-0	-2.2	-2.7	-0.5	2.3	7.8
PDF Pos.	0.4	0.3	0.3	0.3	0.2	0.7
Neg.	0.4	0.3	0.2	0.2	0.2	0.7
Hadronization	-3.6	-2.8	-4.5	-3.6	-6.9	4.5
Hard Scat.	5.7	1.2	5	6.9	16.7	17.7

Table 11.2: Cross sections as a function of $M(t\bar{t})$. The table also presents the summary of all the statistical and systematic uncertainties.

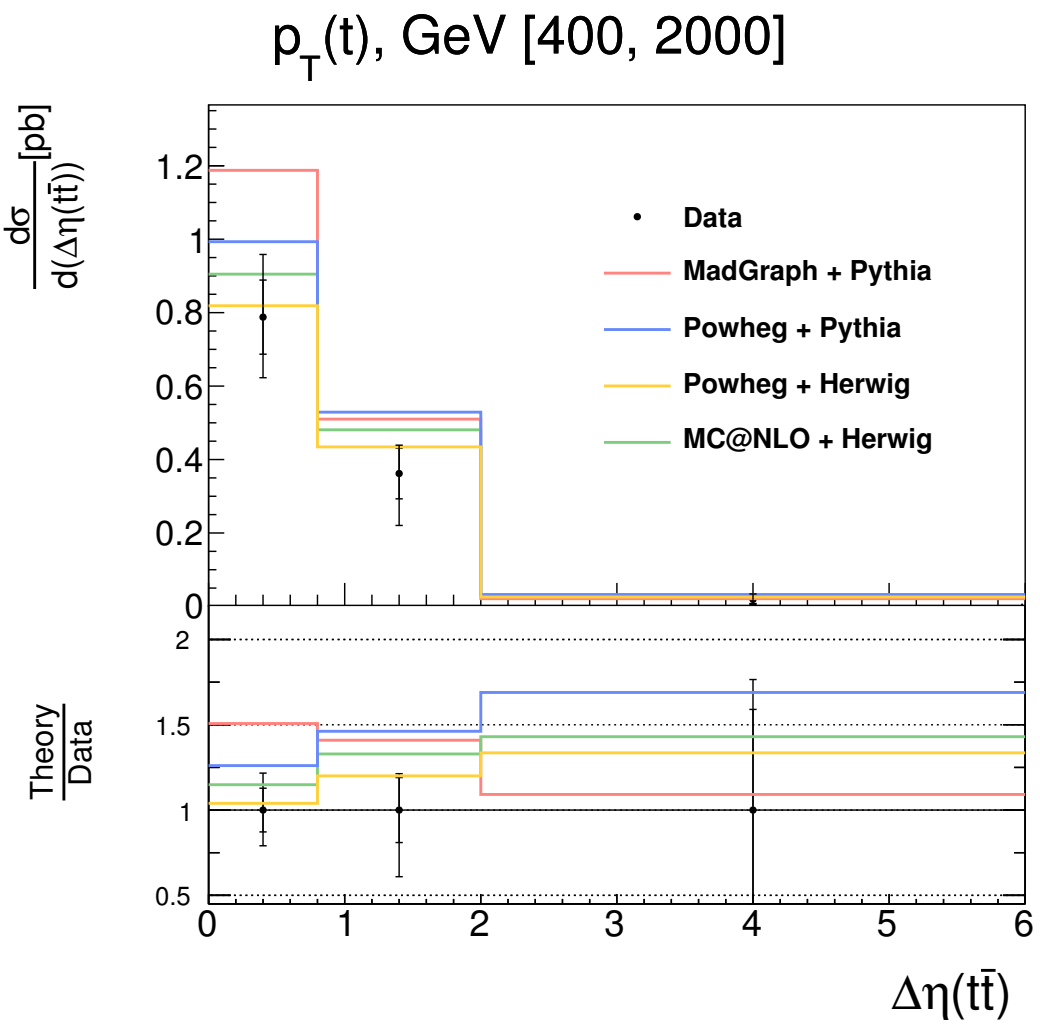

Figure 11.8: The $t\bar{t}$ production cross sections in bins of $|\Delta\eta(t\bar{t})|$ in the high $p_T(t)$ region. For other details see Fig. 11.3.

is observed between data and MadGraph + Pythia prediction.

- The models tend to predict less separation in pseudorapidity between the top and the antitop quark from the produced $t\bar{t}$ system for high $M(t\bar{t})$. This trend is more pronounced for the MadGraph + Pythia predictions, while the best description of the data is given by Powheg + Herwig.

- The models predict less central production of the $t\bar{t}$ pairs in rapidity $|y(t\bar{t})|$ in the high mass region. Here MadGraph + Pythia provides the best description of the data.

The overall conclusion of these observations was that there might be several effects in the MC simulation which could cause the wrong description of the data. It was shown that it is less likely that the effects originate from the wrong simulation of the number of hard radiations in the event, but rather because of the mixing of matrix element hard radiation and parton showering.

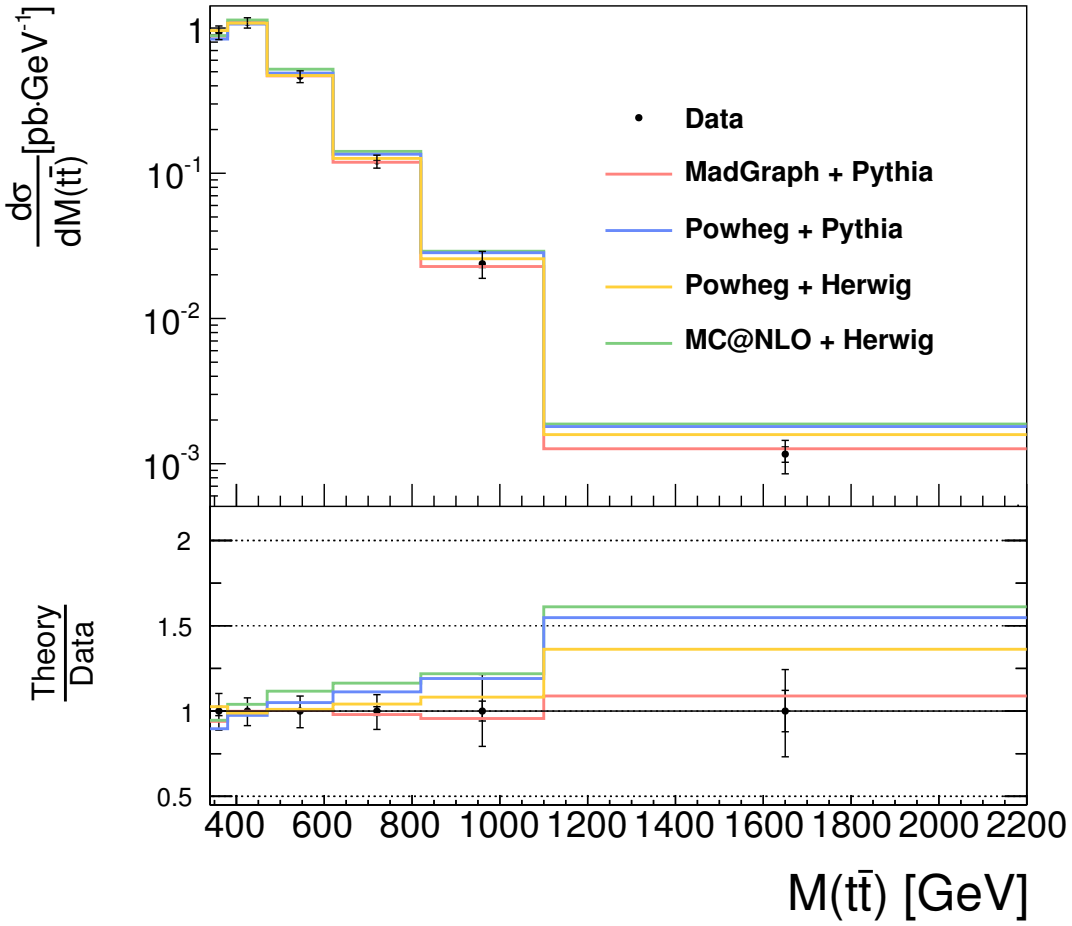

Figure 11.9: The $t\bar{t}$ production cross sections in bins of $M(t\bar{t})$. For other details see Fig. 11.3.

In the current analysis one is able to reproduce the measurements in the high mass region with the selection specially tuned for the boosted top reconstruction and to go higher in the mass of the $t\bar{t}$ system to cross check the tendencies observed in [9]. In general the measurements presented in this work show slightly better precision compared to the previous measurements [9] as in the current measurement all the three channels of the $t\bar{t}$ dileptonic decay (ee, $\mu\mu$ and $e\mu$) are combined, while the previous measurement [9] was done employing the same data but using only the $e\mu$ channel. The event cleaning against the migrations (see Sec. 8.2.1), which is performed in this analysis, is also a source of improvements.

The normalized and unnormalized double differential $t\bar{t}$ production cross sections in bins of $M(t\bar{t})$, which are measured in this analysis, are presented in the following. The normalized double differential cross sections are presented only for the comparison to the previous measurement [9]. The values of the cross sections discussed in the following are given in the tables in Appendix I including statistical and systematic uncertainties. The corresponding correlation matrices are shown in Appendix J. The details of how each of

the following distributions was unfolded are given in Appendix C.

11.5.1 Cross Sections in Bins of $|y(t)|$

The unnormalized and normalized cross sections in bins of $|y(t)|$ for different $M(t\bar{t})$ are presented in Fig. 11.11 and Fig. 11.12 respectively. Moving from smaller to higher values of $M(t\bar{t})$ the overall trend is: for the lower and moderate masses the agreement between data and different MC predictions is good. However, for the larger masses ($M(t\bar{t}) \in (600, 1000)$ GeV) a slight trend is observed towards the prediction of a more central top quark production in different models (mainly MadGraph + Pythia). This trend is in agreement with previous observations (see Fig. 11.10). The measurement in the higher mass region ($M(t\bar{t}) \in (1000, 2000)$ GeV) shows that this effect is getting stronger for all tested MC predictions and most pronounced for MadGraph + Pythia. The best agreement between measurements and predictions is observed for Powheg + Herwig.

One should take note that the difference between MadGraph+Pythia and the data is even more pronounced on the control distributions plotted on the detector level (see Appendix G). The same is true for all other discrepancies discussed below in Sec. 11.5.2 - 11.5.3.

11.5.2 Cross Sections in Bins of $\Delta\eta(t\bar{t})$

In Fig. 11.13 the absolute cross sections in bins of $\Delta\eta(t\bar{t})$ in different ranges of the invariant mass of the $t\bar{t}$ system are presented. The corresponding normalized cross sections are shown in Fig. 11.14. A good agreement between measurements and predictions is observed for the smallest $M(t\bar{t})$ bin ($M(t\bar{t}) \in (350, 390)$ GeV). However, starting from the moderate $M(t\bar{t})$ bins, all the models start to predict a smaller pseudorapidity separation than observed in the data. This tendency is getting more pronounced the higher the invariant mass of the $t\bar{t}$ system is, especially for MadGraph + Pythia, as was also observed in [9]. For the highest $M(t\bar{t})$ region measured in this analysis, the tendency is also visible for all the predictions except for Powheg+Herwig, which also provides the best description of the measurements. In general, the shape of the $\Delta\eta(t\bar{t})$ distribution in data and in NLO predictions in the highest $M(t\bar{t})$ bin (where the cross section is raising with increasing $\Delta\eta(t\bar{t})$) changed compared to all the other bins of $M(t\bar{t})$. However, the shape did not change for the LO MadGraph + Pythia predictions.

11.5.3 Cross Sections in Bins of $|y(t\bar{t})|$

Fig. 11.15 and Fig. 11.16 present the unnormalized and normalized differential $t\bar{t}$ production cross sections as a function of $|y(t\bar{t})|$ for different $M(t\bar{t})$ regions. In general, the agreement between data and predictions is good. The same slight trend is observed as for the previous measurements [9]: the models predict less central $t\bar{t}$ production than the data in the bin $M(t\bar{t}) \in [600, 1000]$ GeV and MadGraph + Pythia also provides the best description of the data. This effect is getting more pronounced in the highest bin of the invariant mass of the $t\bar{t}$ system – $M(t\bar{t}) \in [100, 2000]$ GeV.

11.6 Possible Extension for the Z' Resonance Search

The measurement presented in this thesis provides a unique possibility for a search of heavy resonances decaying into $t\bar{t}$, such as $Z' \to t\bar{t}$ [125]. The previous searches for heavy resonances using the dileptonic $t\bar{t}$ decay channel [126, 127] investigated the visible $t\bar{t}$ mass $M(\bar{l}bl\bar{b}MET)$ only, while with the full kinematic reconstruction applied in this analysis one has access to the total mass $M(t\bar{t})$.

In Fig. 11.17 the possible positions of the Z' resonance in the $M(t\bar{t})$ and visible mass $M(\bar{l}bl\bar{b}MET)$ spectra are presented. To do this, the $t\bar{t}$ signal MC sample was used. Windows of the hypothetical Z' mass of 750 GeV, 1000 GeV and 1500 GeV (with t width of $\pm 1\%$ of the mass) have been selected on the generator level and the reconstructed $M(t\bar{t})$ and $M(\bar{l}bl\bar{b}MET)$ are plotted in this window with a certain magnification factor.

From Fig. 11.17 one can observe that using the reconstructed invariant mass of the $t\bar{t}$ pair an almost unbiased estimate of the resonance mass is obtained. The observed relative RMS[2] for the Z' mass spectra in $M(t\bar{t})$ are 0.216, 0.219 and 0.306 for $M_{Z'} = 750$ GeV, $M_{Z'} = 1000$ GeV and $M_{Z'} = 1500$ GeV, respectively. The corresponding values of the relative RMS in the visible mass spectrum are 0.152, 0.163 and 0.175. This indicates that the full kinematic reconstruction does not bring an improvement in the resolution of the full mass $M(t\bar{t})$ compared to that of the visible mass in particular at high invariant masses. One can conclude from this that for such searches the visible mass provides an observable which is favorable to use. However, further detailed studies including a more realistic simulation of Z' production and decay would be needed to substantiate the conclusions.

[2] A *root mean square*, or RMS of a certain distribution is the the square root of the mean squared deviations from the mean.

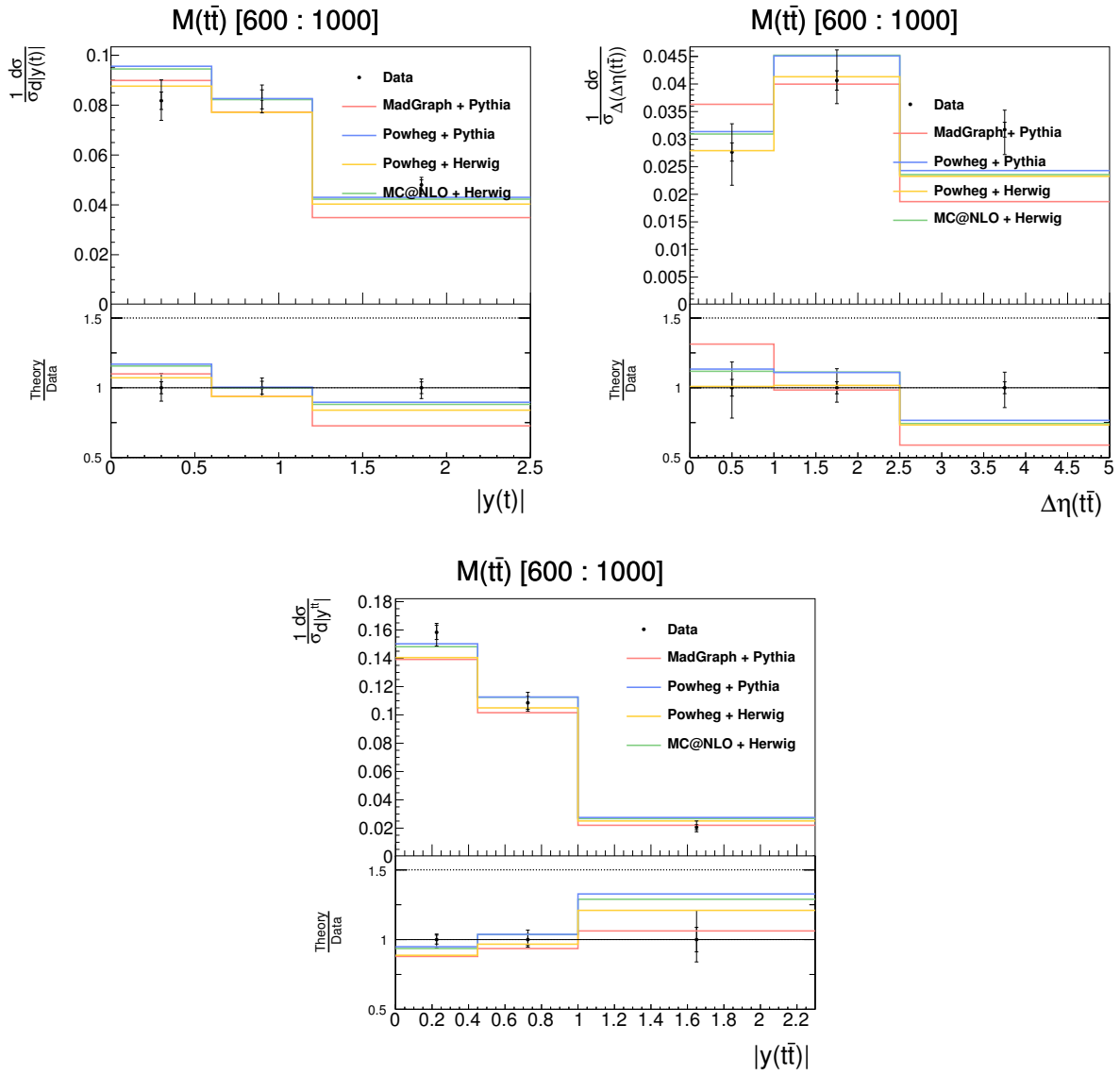

Figure 11.10: Normalized differential cross sections in bins of $|y(t)|$ (top left), $\Delta\eta(t\bar{t})$ (top right) and $|y(t\bar{t})|$ (bottom) in the $M(t\bar{t}) \in (600,\ 1000)$ GeV bin, as measured in I. Korol's PhD thesis [9]. The inner error bars show the statistical uncertainties from the data. The outer error bars are the combined statistical and systematical uncertainties on the data. The predicted cross sections from four different models are also presented: MadGraph + Pythia (red line), Powheg + Pythia (blue line), Powheg + Herwig (orange line) and MC@NLO + Herwig (green line). The ratio of the predictions over the measured cross sections are shown in the bottom panels with the error bars corresponding to the measurement uncertainties.

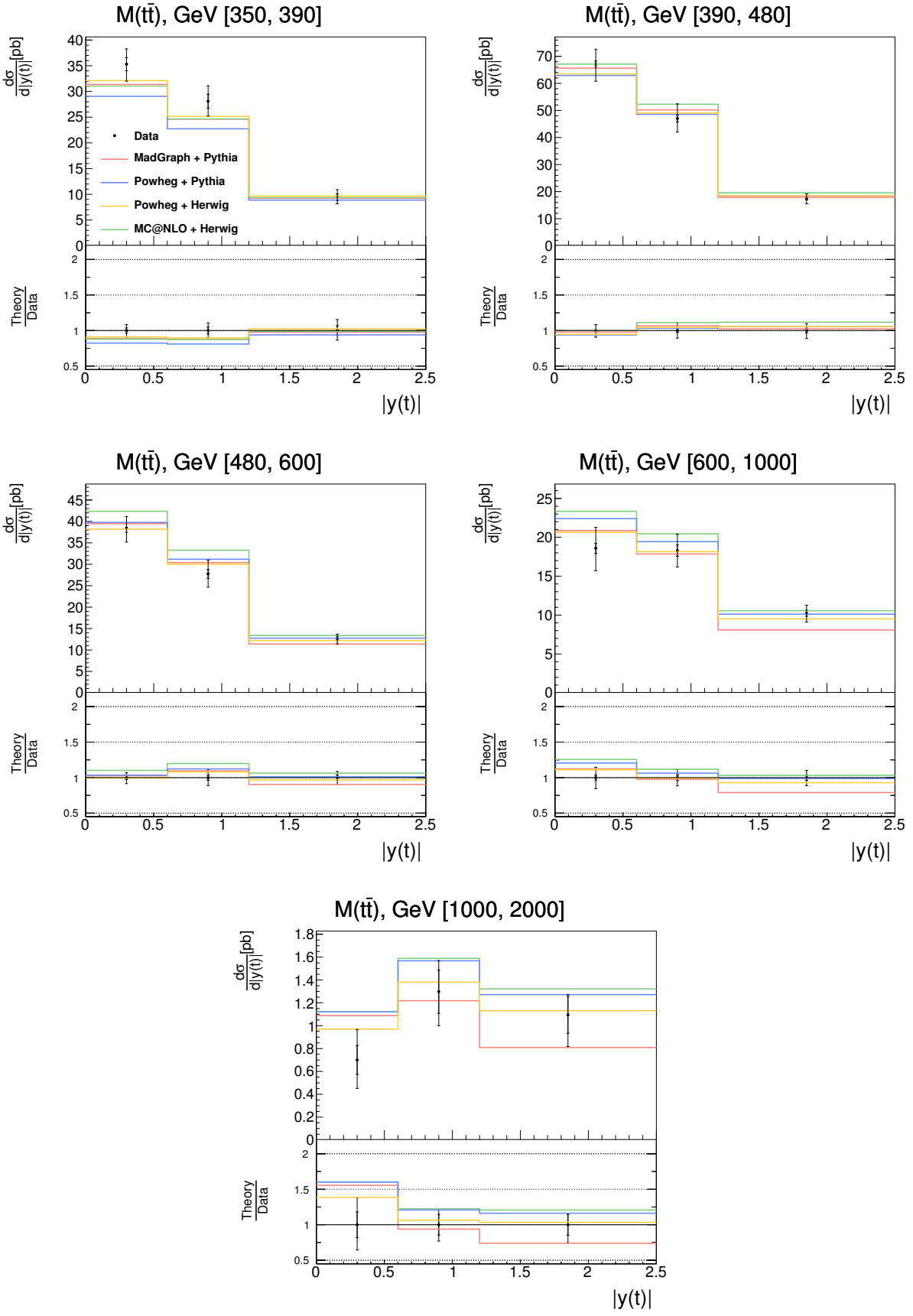

Figure 11.11: The $t\bar{t}$ production cross sections in bins of $|y(t)|$ for different $M(t\bar{t})$ region. For other details see Fig. 11.3.

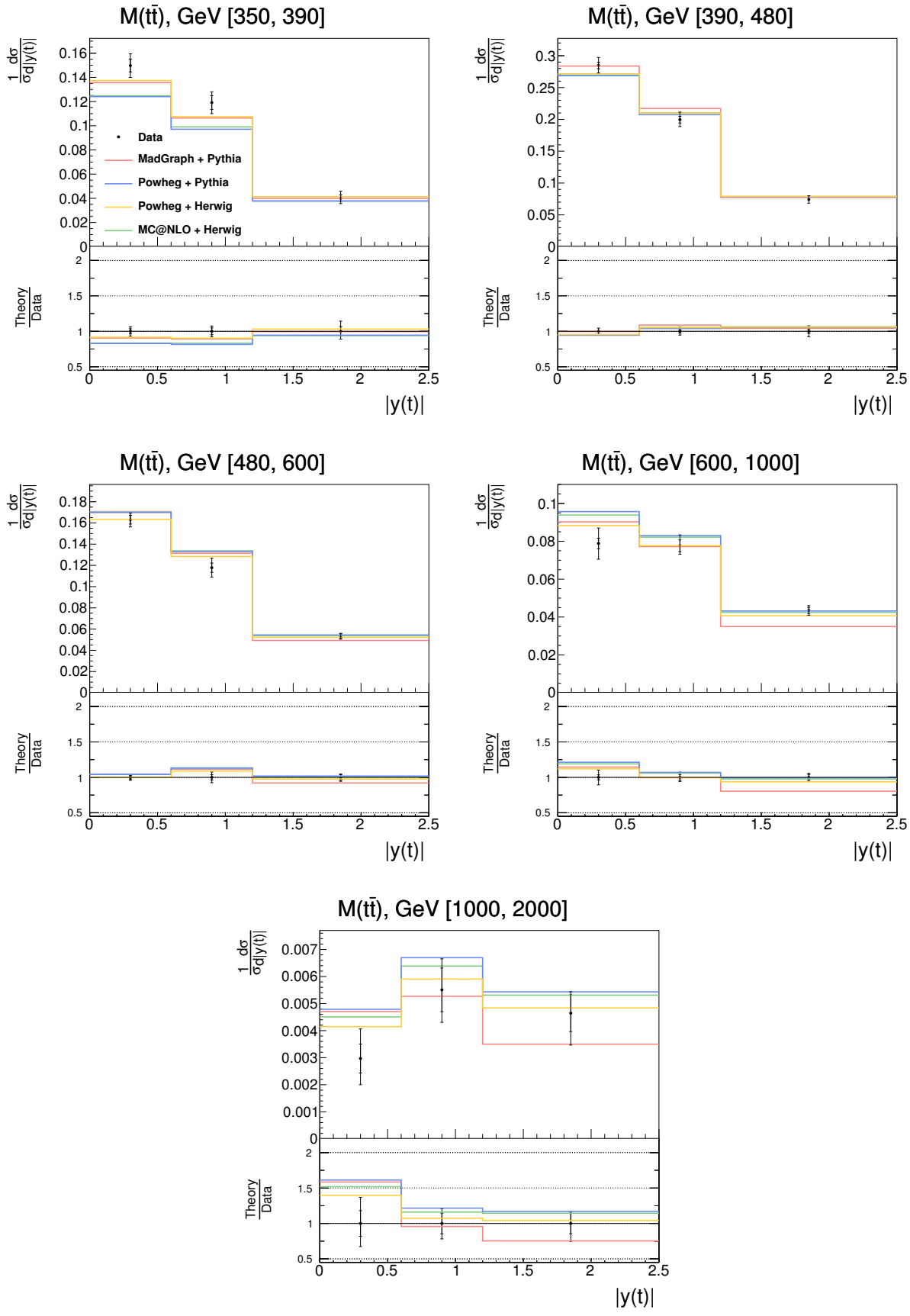

Figure 11.12: Normalized $t\bar{t}$ production cross sections in bins of $|y(t)|$ for different $M(t\bar{t})$ region. For other details see Fig. 11.3.

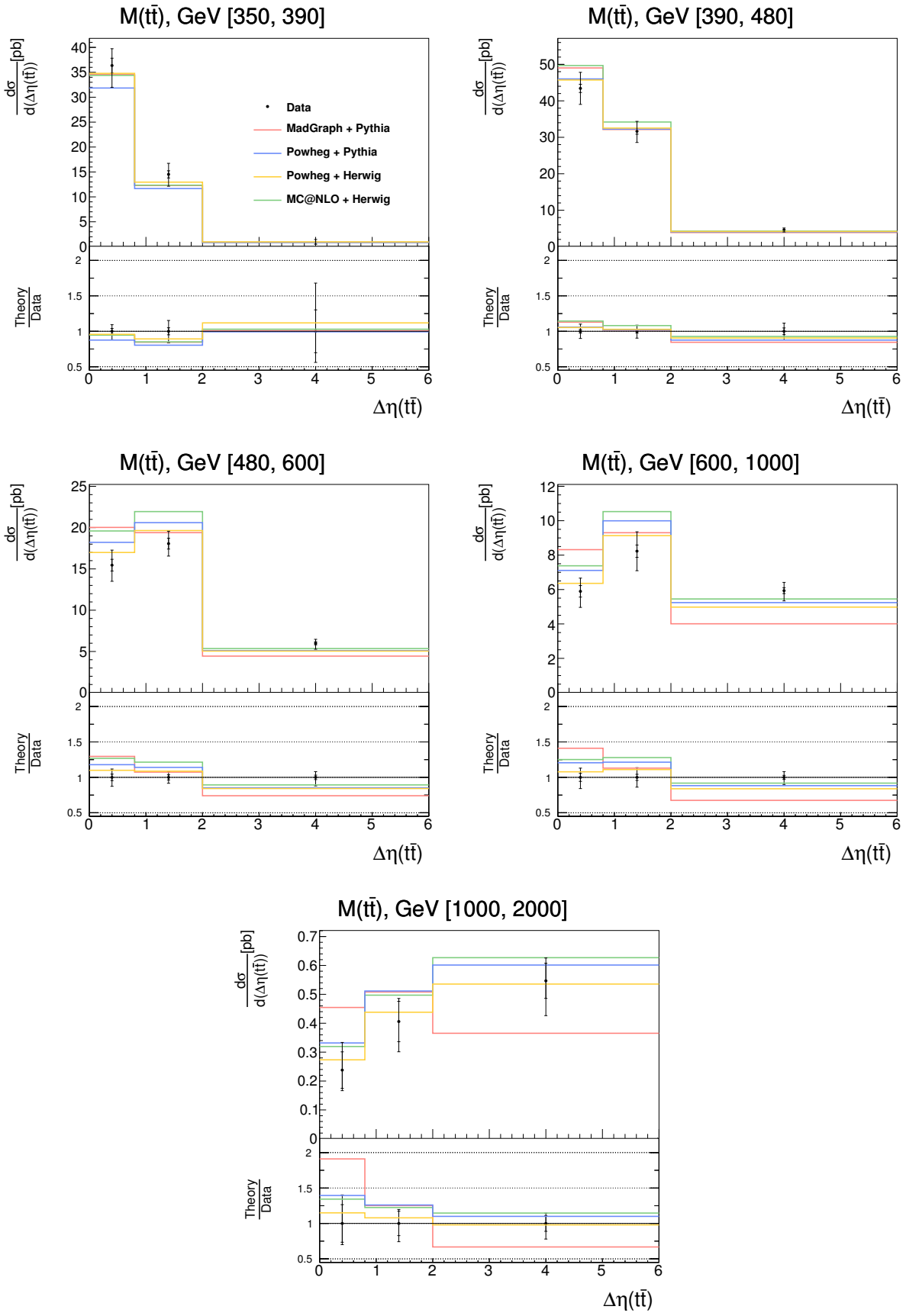

Figure 11.13: The $t\bar{t}$ production cross sections in bins of $\Delta\eta(t\bar{t})$ for different $M(t\bar{t})$ region. For other details see Fig. 11.3.

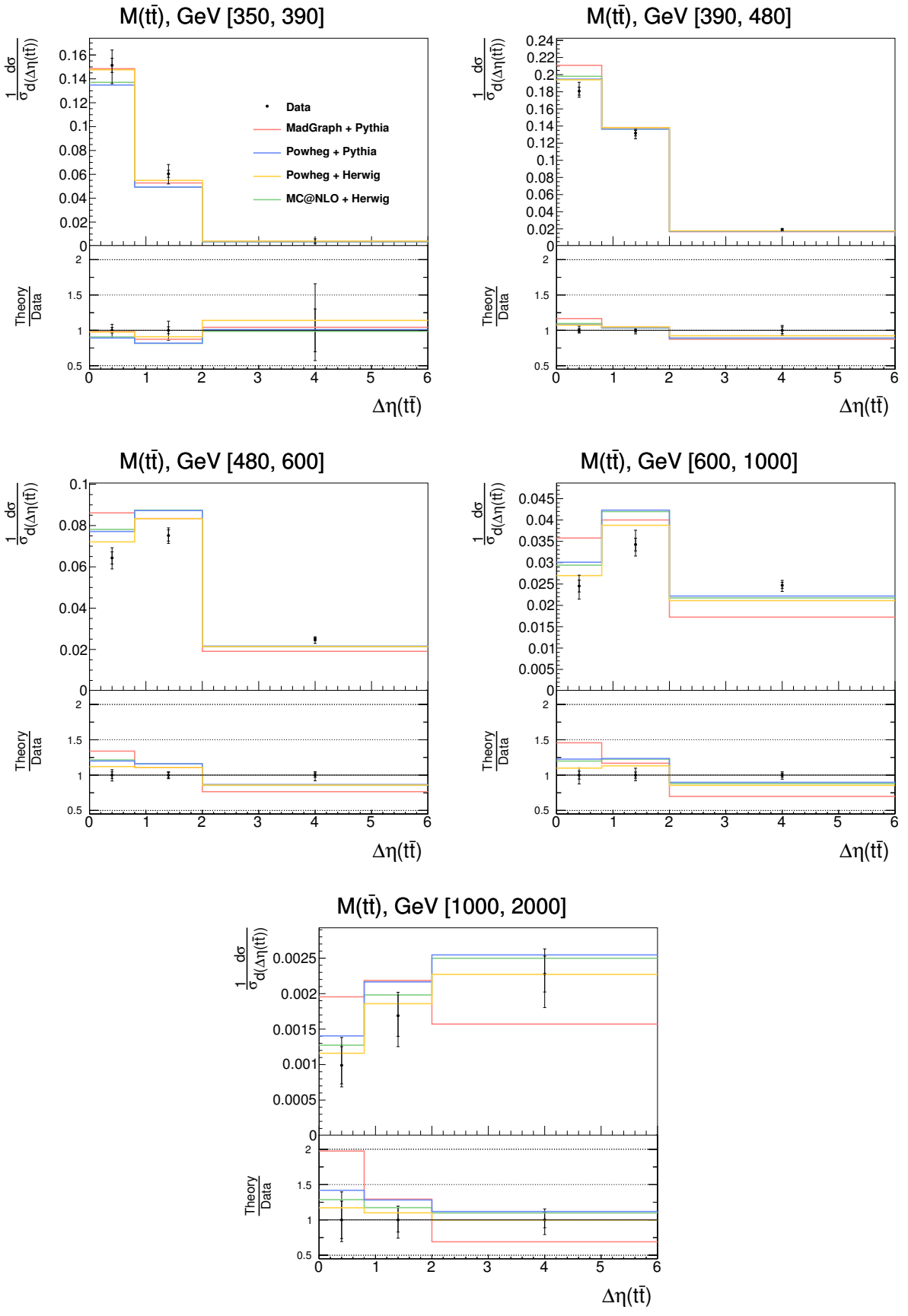

Figure 11.14: Normalized $t\bar{t}$ production cross sections in bins of $\Delta\eta(t\bar{t})$ for different $M(t\bar{t})$ region. For other details see Fig. 11.3.

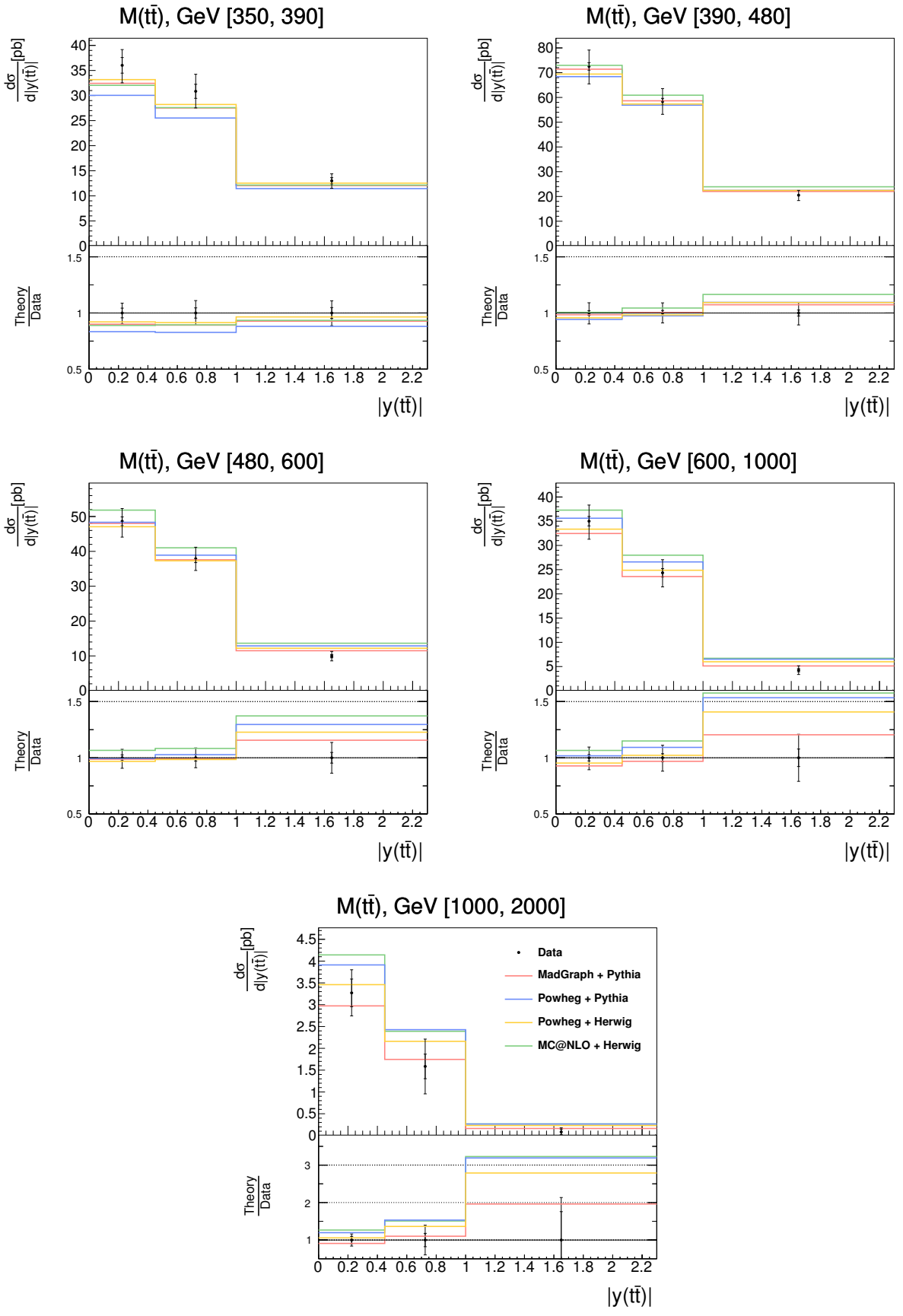

Figure 11.15: The $t\bar{t}$ production cross sections in bins of $|y(t\bar{t})|$ for different $M(t\bar{t})$ region. For other details see Fig. 11.3.

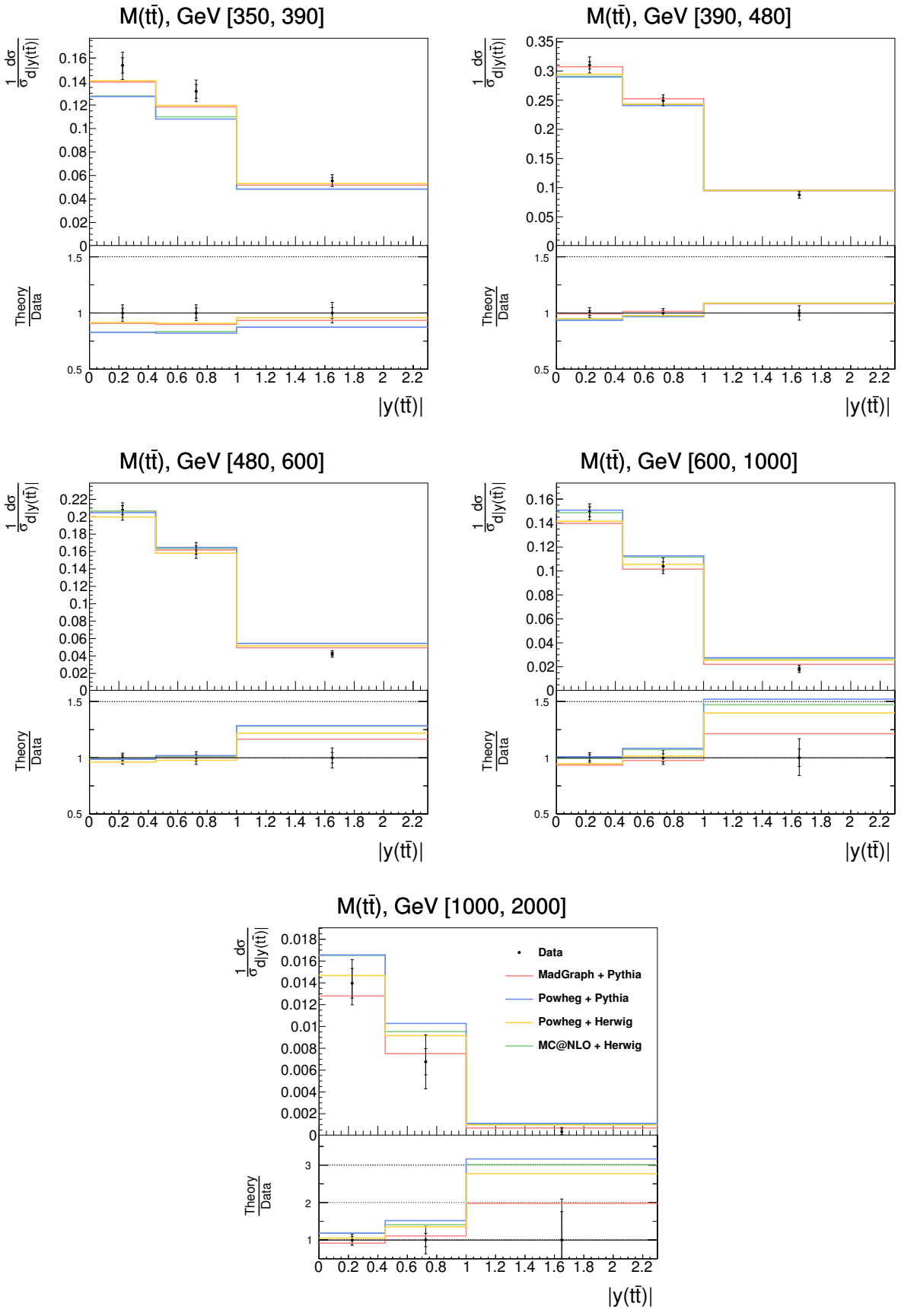

Figure 11.16: Normalized $t\bar{t}$ production cross sections in bins of $|y(t\bar{t})|$ for different $M(t\bar{t})$ region. For other details see Fig. 11.3.

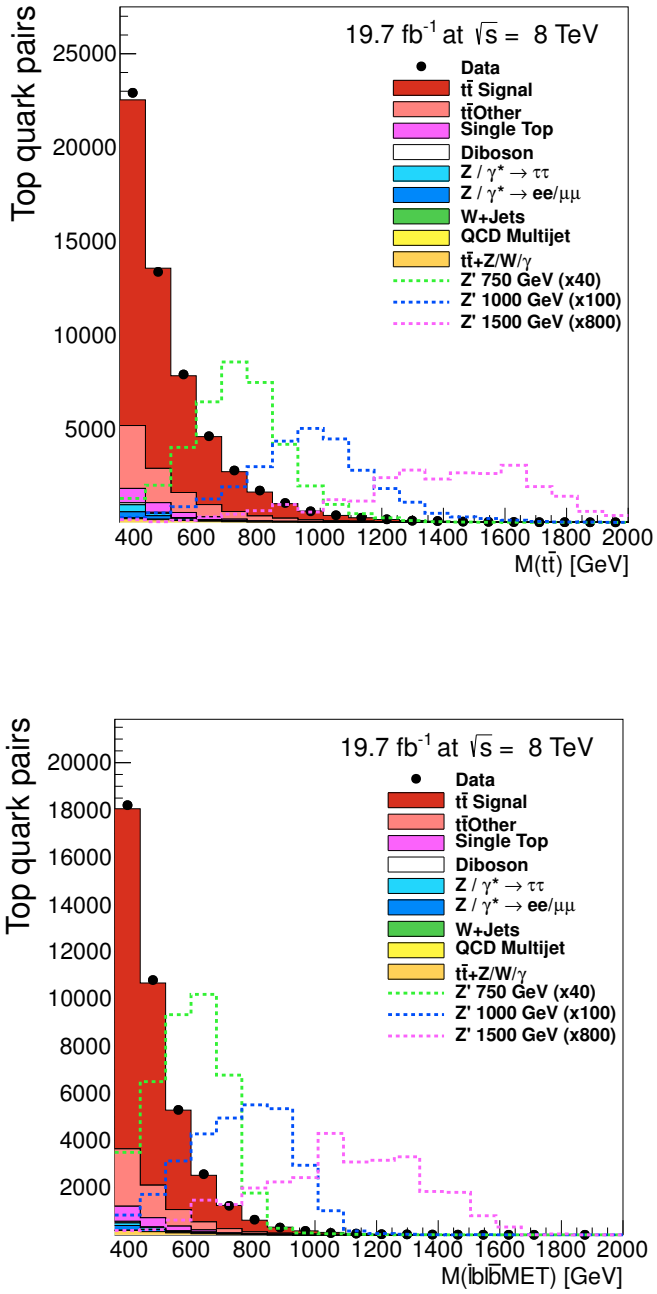

Figure 11.17: Control distributions of the reconstructed invariant mass of the $t\bar{t}$ (top) and visible mass of $M(\bar{l}b\bar{l}\bar{b}MET)$. Black dots show the data points and filled histograms represent the MC. Histograms with different colors represent contributions from different background processes. The error bars represent the statistical uncertainties of the data. The dashed green line represents a hypothetical Z' peak with the mass of 750 GeV magnified by the factor of 40, the dashed blue line shows the hypothetical Z' peak with the mass of 1000 GeV magnified by a factor 100 and the dashed magenta line shows the hypothetical Z' peak with the mass of 1500 GeV magnified by a factor of 800.

12 | Summary

The results of the analysis presented in this thesis are the first measurements of the production cross sections of $t\bar{t}$ pairs in pp collisions which have been optimized for the boosted topologies in the dileptonic decay channel. These results were obtained using the data collected by the CMS detector in 2012, when the LHC was operating with a center-of-mass pp collision energy of 8 TeV.

The analysis of the dilepton channel allows to separate for the boosted top quarks all decay products of the $t\bar{t}$ system, unlike the cases of semi-leptonic and fully hadronic channels where all the decay products of a boosted top quark with hadronically decaying W are usually reconstructed as a single "fat jet". However, the two neutrinos from the dilepton $t\bar{t}$ decay can not be directly detected. A procedure of full kinematic reconstruction was introduced. This procedure is based on solving the kinematic equations introduced by six kinematic constraints related to transverse momentum conservation and top and W masses. To account for the detector effects, which sometimes lead to insolvable kinematic equations, all the energies and directions of the input objects of the kinematic reconstruction are simultaneously smeared 100 times. The final solution is taken as a weighted average over all the smeared solutions.

The kinematic reconstruction as described above has also been used in the previous measurement of the single- and double differential $t\bar{t}$ production cross sections in the dilepton channel using the CMS data collected at the 8 TeV center-of-mass proton collisions [9, 35].

To account specifically for the features of the top decay in boosted topologies, where the lepton from the W from the top flies in close vicinity of the b-jet from the top, a so-called "2D lepton isolation" was used replacing the standard relative isolation requirement [35]. This isolation is based on rejecting a lepton, which has a small distance in (η, ϕ) to the closest jet and a small projection of the relative transverse momentum with respect to the direction of the closest jet.

The introduction of a cleaning condition based on setting a minimum threshold on a ratio of visible and reconstructed $t\bar{t}$ masses was the other tune to improve the reconstruction of the top quarks in the boosted topologies. The studies of the effectiveness of this condition show a nice improvement on the levels of event reconstruction and cross section measurement.

The final results of this work are consistent with various previous measurements [9, 35, 118, 124]. They also provide extended information about the regions with high transverse momentum of the top quarks ($p_T(t) > 400$) GeV and for high $M(t\bar{t})$ compared to [9, 35]. No deviations from the results presented in previous analyses were observed. One can observe some trends in particular distributions, which indicate that the models are far from being perfect.

The cross sections are also measured as a function of minimum p_T threshold. The measured cross section falls off steeply with increasing threshold. A reasonable measurement from the statistics point of view is possible up to $p_T(t) > 550$ GeV. This measurement was also compared to previous measurements performed in semi-leptonic channel [118, 124], where the hadronically decaying top was reconstructed in a single "fat jet". Results show reasonable agreement and the measurement presented in this work has a compatible precision up to a threshold of $p_T(t) > 500$ GeV.

12.1 Outlook

The measurement presented in this thesis has several perspectives for improvements. Firstly, the number of the events with the $t\bar{t}$ production and further dileptonic decay of the system in the region, where at least one top quark is boosted, is very small. The measurements with the data which are currently collected from the pp collisions in Run II with increased center-of-mass collision energy of $\sqrt{s} = 13 - 14$ TeV are expected to provide much more statistics for the $t\bar{t}$ events with boosted top quark topologies due to the increased production cross sections and higher luminosities.

The other effect, which strongly influences the accuracy of the measurement of the $t\bar{t}$ production with boosted top quarks is the migrations of events with small true $p_T(t)$ to high reconstructed $p_T(t)$. This effect can be treated by further improving the kinematic reconstruction and optimizing the "cleaning conditions", e.g. using information about the jets or about the missing transverse energy. For instance, one could consider to vary the kinematic reconstruction as a function of different jet multiplicities (consider only the combinations with two b-tagged jets for the higher multiplicities to reduce combinatorics). The other option of improvement of the kinematic reconstruction is to use the energy and angle correction factors, which are employed in smearing procedure, as a function of lepton and jet p_T and η. To further increase the efficiency of the reconstruction, one could assign the expected $M(b\bar{b}l\text{MET})/M(t\bar{t})$ spectrum as a weight for different smearings instead of rejecting all the events with $M(b\bar{b}l\text{MET})/M(t\bar{t}) < 0.5$.

In general, a deeper study of the origin of the migrations from the lower $p_T(t)$ to the boosted region should be made. For instance, the migrations might be caused by some specific event topologies, where a small change in the input to kinematic reconstruction causes large change in the solution of the kinematic equations. The detailed studies of the dependence of the migrations on the underlying decays topologies were not performed in this analysis.

Finally, this analysis can be extended to a resonance search, where for the first time a fully reconstructed $t\bar{t}$ invariant mass will be analyzed (in the previous analyses, like [126, 127], the spectrum of visible mass $M(b\bar{b}l\text{MET})$ was analyzed).

A | Comparison between TUnfold and SVD Performance

The performance of the two methods of unfolding (SVD [121] and TUnfold [119]) for this analysis is compared in this section. The cross section as a function of the transverse momentum of the top quark ($p_T(t)$) has been unfolded with the two methods, the results are shown in Fig. A.1. Somewhat larger statistical uncertainties are observed for the TUnfold results in the two bins with higher transverse momentum of the top ($p_T(t) > 400$ GeV) compared to the SVD unfolding.

The reason for that was found to originate from the regularization (see Sec. 9.2.1). The regularization strength (τ) is determined using the minimization of correlation coefficients. However, for the SVD unfolding the minimization results in $\tau \approx 2$, while in the TUnfold it was found to be $\tau \approx 10^{-4}$, which is four orders of magnitude smaller. That means that TUnfold is less dependent on the regularization. To check how the stronger regularization affects the final results, the correlation matrices for both results are compared (see Fig. A.2). The correlations in the results obtained with SVD are very large (up to 99%) and thus indicate overregularization. The correlation matrices in bins of $M(t\bar{t})$ are also compared (see Fig. A.3). The same situation can be also made for this case: the correlations obtained with SVD unfolding in the higher mass bins are of the order of 96-99%. In summary it looks like the results obtained with SVD are oversmoothed due to a very large strength of the regularization. That is why it was decided to use TUnfold tool for this analysis.

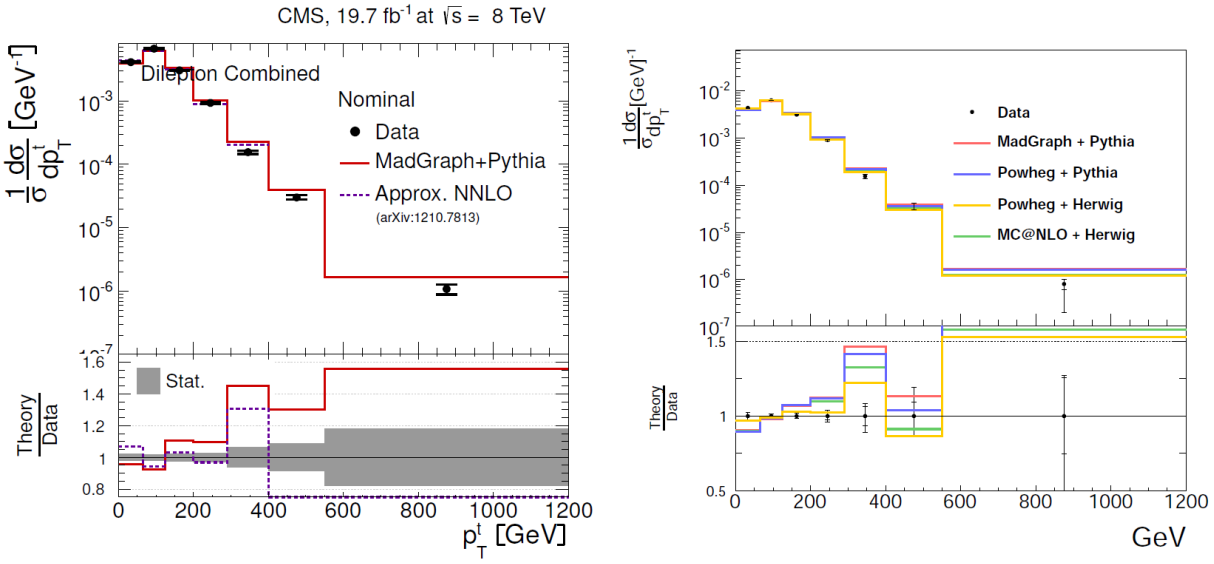

Figure A.1: Normalized differential $t\bar{t}$ production cross section as a function of the $p_T(t)$. The left plot represents the results obtained using the SVD unfolding. Here, the error bars indicate the statistical uncertainties. The measurements are compared to predictions from MadGraph + Pythia and approximate NNLO calculations. The lower part of the plot shows the ratio of the predictions to data. The gray band represents the statistical uncertainties of the measurement. On the right plot the cross sections obtained with TUnfold are presented. The inner error bars correspond to the statistical only uncertainties and the outer error bars are the combined statistical and systematic uncertainties. The predicted cross sections from MadGraph + Pythia is represented with a red line, the cross sections from Powheg + Pythia are drawn with a blue line, Powheg + Herwig predictions are marked with an orange line and MC@NLO + Herwig is represented with a green line. The ratio of the predictions over measured cross sections are shown in the bottom panel. The error bars around the constant line of unity show the uncertainties of the measurement.

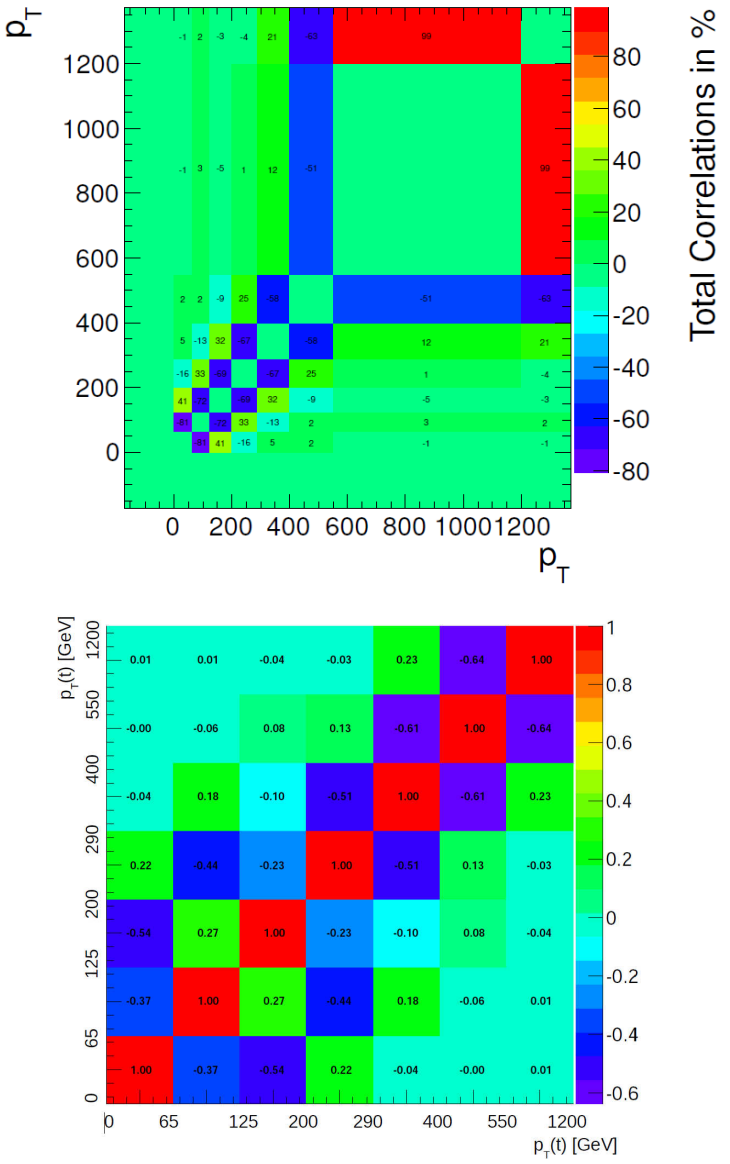

Figure A.2: Correlation matrices for the differential cross sections $\frac{d\sigma}{dp_T(t)}$ obtained with the SVD unfolding (top) and TUnfold (bottom). The diagonal elements of the correlation matrix obtained with SVD unfolding are artificially set to zero for a better visualization.

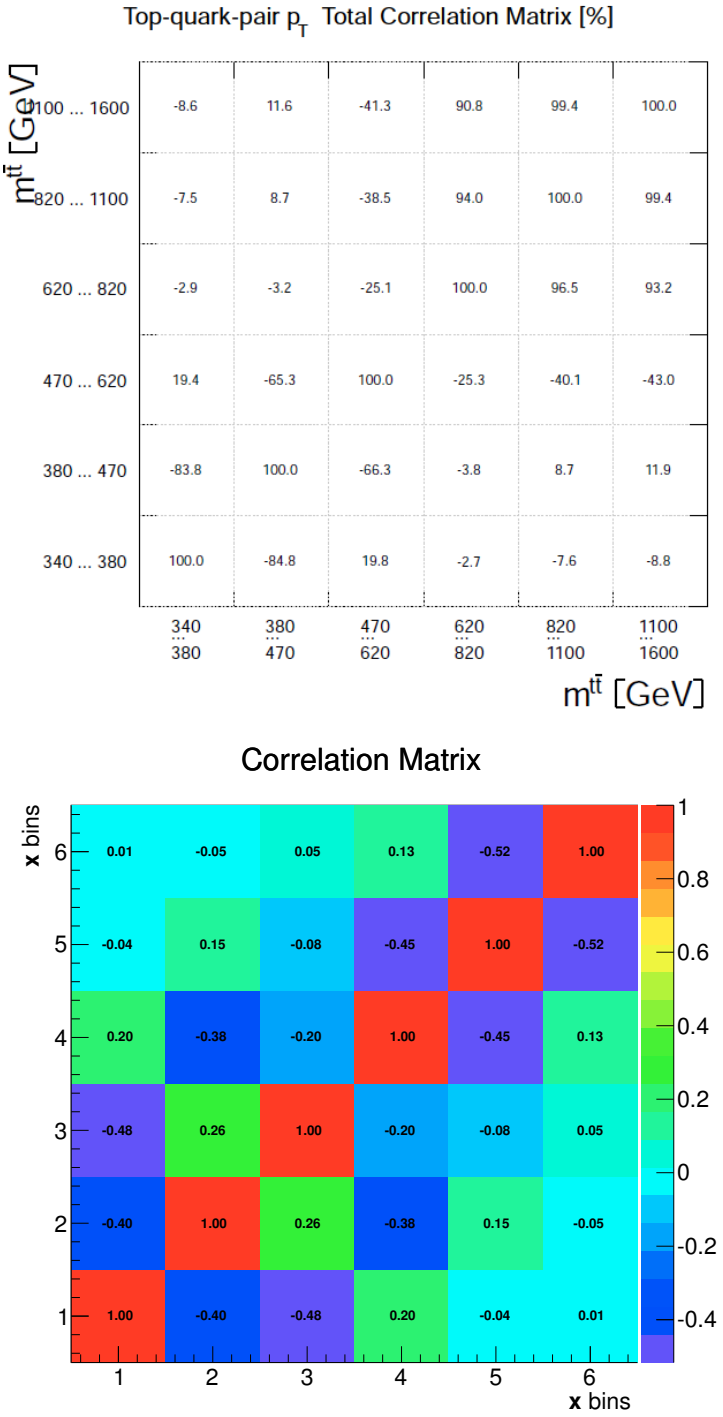

Figure A.3: Correlation matrices for the $M(t\bar{t})$ distributions obtained with the SVD unfolding (top) taken from [104] and TUnfold (bottom). The correlations on the diagonal of the correlation matrix obtained with the SVD unfolding are artificially set to zero for plotting, while the actual values equal to unity.

B | Regularization Strength Parameters

The choice of the regularization strength of every set of unfolded distributions described in Sec. 11.2 and in Sec. 11.4 was based on the minimization of global correlation coefficients (see Sec. 9.2.1). The resulting values of the corresponding parameters τ (which are overall very low) are listed in the following tables:

Variables	τ
$p_T(t) > 400$ GeV	$2.93 \cdot 10^{-4}$
$p_T(t) > 450$ GeV	$2.94 \cdot 10^{-4}$
$p_T(t) > 500$ GeV	$2.94 \cdot 10^{-4}$
$p_T(t) > 550$ GeV	$2.95 \cdot 10^{-4}$
$p_T(t) > 600$ GeV	$2.95 \cdot 10^{-4}$
$p_T(t) > 650$ GeV	$2.95 \cdot 10^{-4}$

Table B.1: Regularization strength parameter τ for the measurement of the $t\bar{t}$ production cross section with different thresholds for the transverse momentum of the top quark. Here, the unfolding is performed simultaneously with the cleaning variable $M(b\bar{l}\bar{b}l\text{MET})/M(t\bar{t})$.

Variables	τ
$\lvert y(t)\rvert$ in bins of $p_T(t)$	$3.31 \cdot 10^{-4}$
$\lvert y(t\bar{t})\rvert$ in bins of $p_T(t)$	$3.35 \cdot 10^{-4}$
$\Delta\eta(t\bar{t})$ in bins of $p_T(t)$	$2.80 \cdot 10^{-4}$

Table B.2: Regularization strength parameter τ for the measurement of the $t\bar{t}$ production cross section for different variables in bins of $p_T(t)$.

Variables	τ
$M(t\bar{t})$	$2.97 \cdot 10^{-4}$
$\lvert y(t)\rvert$ in bins of $M(t\bar{t})$	$3.41 \cdot 10^{-4}$
$\lvert y(t\bar{t})\rvert$ in bins of $M(t\bar{t})$	$3.50 \cdot 10^{-4}$
$\Delta\eta(t\bar{t})$ in bins of $M(t\bar{t})$	$2.73 \cdot 10^{-4}$

Table B.3: Regularization strength parameter τ for the measurement of the $t\bar{t}$ production cross section for different variables in bins of $M(t\bar{t})$. The first row represents the measurement of the cross section in bins of $M(t\bar{t})$ only, which was unfolded simultaneously with thez cleaning variable $M(b\bar{l}\bar{b}l\mathrm{MET})/M(t\bar{t})$.

C | Details of the Unfolding of the Measurements

All the results presented in Chapter 11 were obtained with the Unfolding procedure (see Sec. 9.2.1). For each measurement the binning of each distribution was chosen individually for both detector and generator level (as required by TUnfold tool) to obtain maximum reasonable inter-bin correlations (the correlation matrices are shown in Appendix J).

The following binning and unfolding strategies were chosen for each of the measurements:

- **Cross sections in bins of minimum $p_T(t)$ threshold** (see Fig. 11.3): for this measurement each point on the plot corresponds to the separate measurement of the integral cross section in the highest bin of the corresponding $p_T(t)$ distribution. The following binning was chosen for each of those distributions:
  - for the point $p_T(t) > 400$ GeV: as an input for TUnfold the generator level $p_T(t)$ binning (coarse binning) is chosen to be [0, 65, 125, 200, 290, 400, 2000] GeV (no events above 2000 GeV are observed);
  - for the point $p_T(t) > 450$ GeV: coarse binning is $[0, 65, 125, 200, 290, 450, 2000]$ GeV;
  - for the point $p_T(t) > 500$ GeV: coarse binning is $[0, 65, 125, 200, 290, 500, 2000]$ GeV;
  - for the point $p_T(t) > 550$ GeV: coarse binning is $[0, 65, 125, 200, 290, 550, 2000]$ GeV;
  - for the point $p_T(t) > 600$ GeV: coarse binning is $[0, 65, 125, 200, 290, 600, 2000]$ GeV;
  - for the point $p_T(t) > 650$ GeV: coarse binning is $[0, 65, 125, 200, 290, 650, 2000]$ GeV;

  For all these points the same detector level binning (fine binning) is used – [0, 30, 65, 90, 125, 150, 175, 200, 210, 240, 290, 360, 400, 450, 510, 700, 900, 2000] GeV. The final unfolded distribution is presented in the coarse binning.

  All the distributions, which are listed above, are unfolded simultaneously with the cleaning variable distribution $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ with one bin on the generator level ([0, 1.5]) and four bins on the detector level ([0, 0.3, 0.5, 0.7, 1.5]). The simultaneous unfolding of the two distributions (2D unfolding) is performed in TUnfold.

- **Cross sections in bins of $p_T(t)$** (see Fig. 11.2 left): it is the same measurement as the previously described $p_T(t) > 400$ GeV case, with additional normalization by the bin width.

- **Cross sections in bins of $|y(t)|$ in the boosted regime** (see Fig. 11.6): the distribution of $\frac{d\sigma}{|y(t)|}$ is unfolded simultaneously with the $p_T(t)$ distribution. The binning of the $y(t)$ is the following: [0.0, 0.6, 1.2, 2.5] on the coarse binning and [0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.5, 2.5] for the detector level binning. In $p_T(t)$, the binning is chosen to be [0, 65, 130, 200, 400, 2000] GeV (generator level) and [0, 40, 65, 95, 130, 165, 200, 235, 275, 330, 400, 475, 2000] GeV (detector level). The results in Fig. 11.6 are presented in the last $p_T(t)$ bin only, which corresponds to the boosted regime. The results in the other bins of $p_T(t)$ are presented in Appendix H. For this cross section the cleaning requirement $M(\bar{l}bl\bar{b}MET)/M(t\bar{t}) > 0.5$ is applied.

- **Cross sections in bins of $|y(t\bar{t})|$ in the boosted regime** (see Fig. 11.7): the distribution $\frac{d\sigma}{|y(t\bar{t})|}$ is unfolded simultaneously with the $\frac{d\sigma}{p_T(t)}$. The binning of the $|y(t\bar{t})|$ on the generator level is [0, 0.45, 1.0, 2.3] and on the detector level – [0, 0.2, 0.45, 0.65, 1.0, 1.4, 2.3]. Other details as for the $|y(t)|$ in boosted regime.

- **Cross sections in bins of $\Delta\eta(t\bar{t})$ in boosted regime** (see Fig. 11.8): the strategy for $\Delta\eta(t\bar{t})$ unfolding is the same as for the previous two distributions. The binning of the $\Delta\eta(t\bar{t})$ distribution was chosen the following way: [0.0, 0.8, 2.0, 6.0] for the generator level binning and [0.0, 0.3, 0.8, 1.35, 2.0, 3.0, 6.0] for the detector level.

- **Cross sections in bins of $M(t\bar{t})$** (see Fig. 11.9): the distribution of $\frac{d\sigma}{M(t\bar{t})}$ (generator level binning for $M(t\bar{t})$ is [340, 380, 470, 620, 820, 1100, 2200] GeV and the detector level binning is [340, 360, 380, 425, 470, 550, 620, 720, 820, 950, 1100, 1650, 2200] GeV) is unfolded simultaneously with the $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ (binning is the same as for the measurement as a function of minimum $p_T(t)$ threshold). No other cleaning condition is applied.

- **Cross sections in bins of $|y(t)|$ for different $M(t\bar{t})$ regions** (see Fig. 11.11): the distribution of $\frac{d\sigma}{d|y(t)|}$ is unfolded simultaneously with $M(t\bar{t})$. The binning for $|y(t)|$ is the same as for the measurement in boosted region. The binning for the $M(t\bar{t})$ is [350, 390, 480, 600, 1000, 2000] GeV on the generator level and [350, 365, 390, 410, 420, 431, 445, 460, 480, 500, 530, 563, 600, 650, 740, 850, 1000, 1100, 2000] GeV on the detector level. Additional, the cleaning requirement $M(\bar{l}bl\bar{b}MET)/M(t\bar{t}) > 0.5$ is applied.

- **Cross sections in bins of $|y(t\bar{t})|$ for different $M(t\bar{t})$ regions** (see Fig. 11.15): the same strategy is used as for the previous measurement.

- **Cross sections in bins of $|\Delta\eta(t\bar{t})|$ for different $M(t\bar{t})$ regions** (see Fig. 11.13): the same strategy is used as for the previous measurement.

D | Efficiencies, Purities and Stabilities

Plots, which show efficiencies, purities and stabilities distributions for which the cross sections are measured are presented in Fig. D.5-D.8.

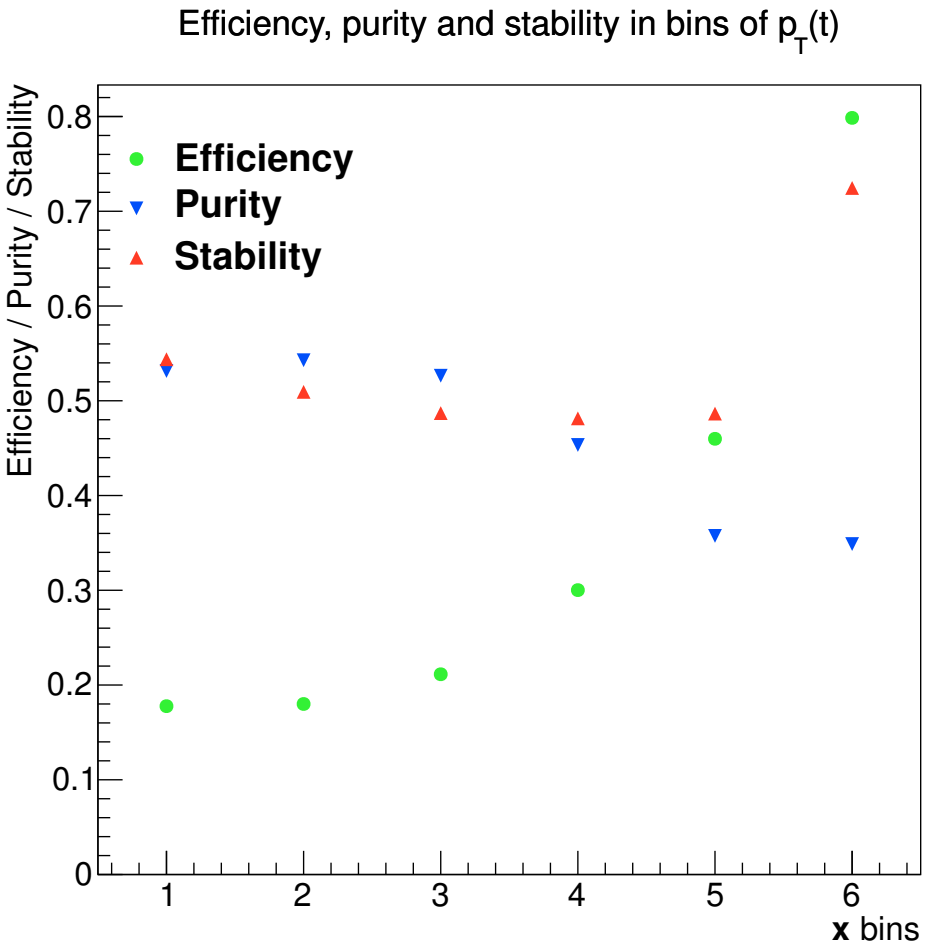

Figure D.1: Efficiencies (green circles), purities (blue invertes triangles) and stabilities (red triangles) for the $p_T(t)$. The binning is the following: [0, 65, 125, 200, 290, 400, 2000] GeV.

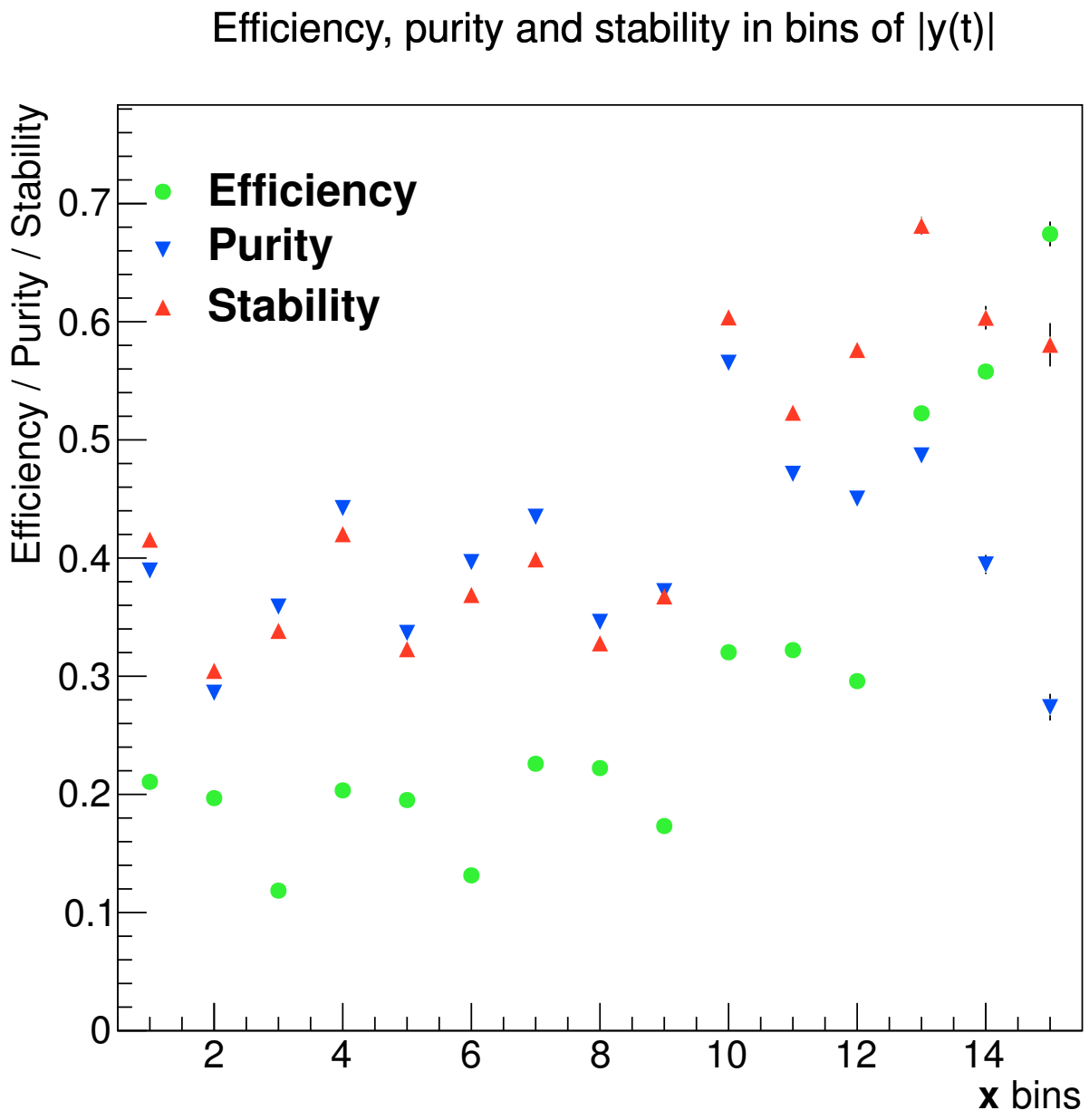

Figure D.2: Efficiencies (green circles), purities (blue invertes triangles) and stabilities (red triangles) for the $|y(t)|$ in bins of $p_T(t)$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $p_T(t)$ – [0, 65, 130, 200, 400, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t)|$ – [0, 0.6, 1.2, 2.5].

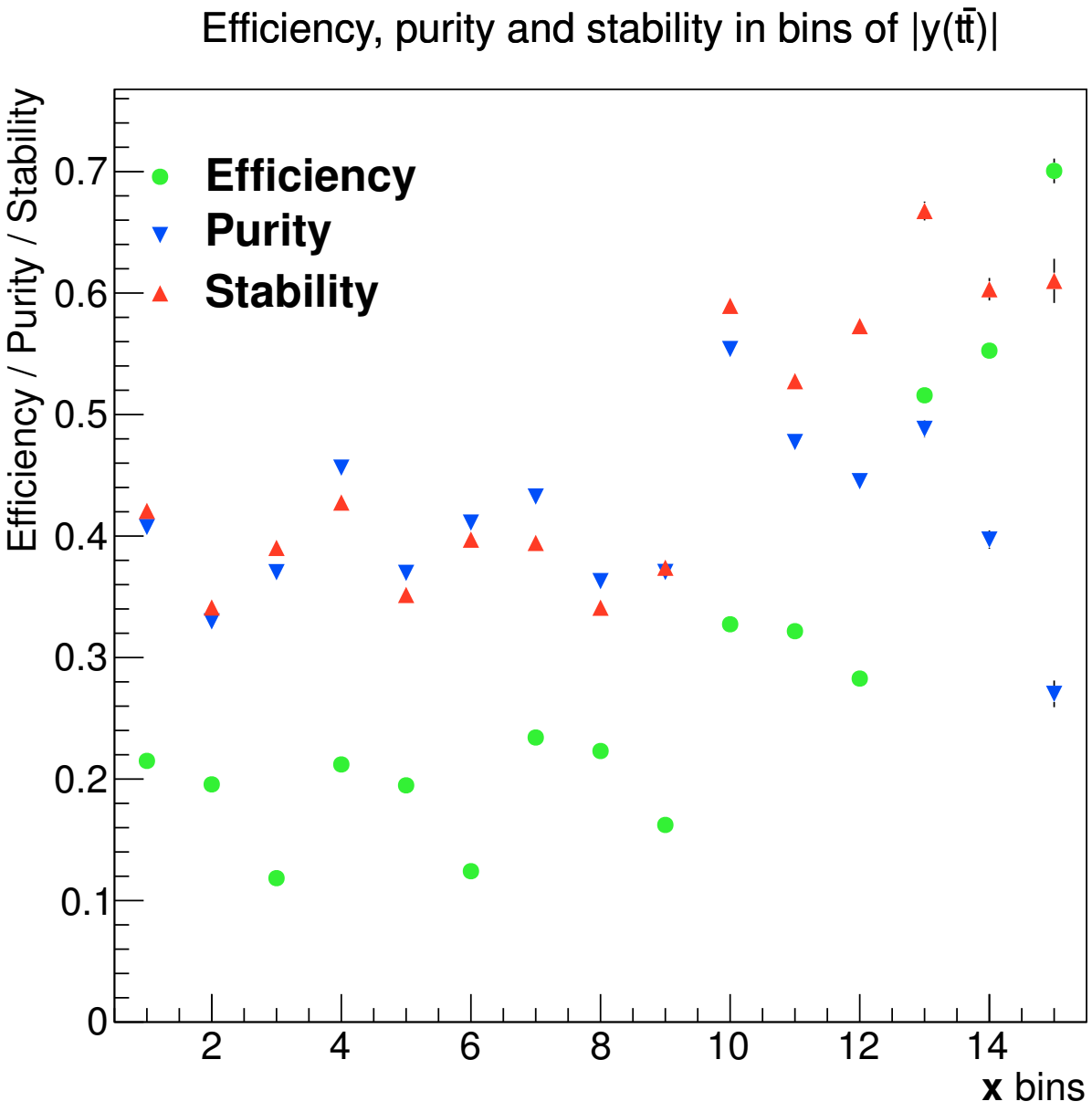

Figure D.3: Efficiencies (green circles), purities (blue invertes triangles) and stabilities (red triangles) for the $|y(t\bar{t})|$ in bins of $p_T(t)$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $p_T(t)$ – [0, 65, 130, 200, 400, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t\bar{t})|$ – [0, 0.45, 1.0, 2.3].

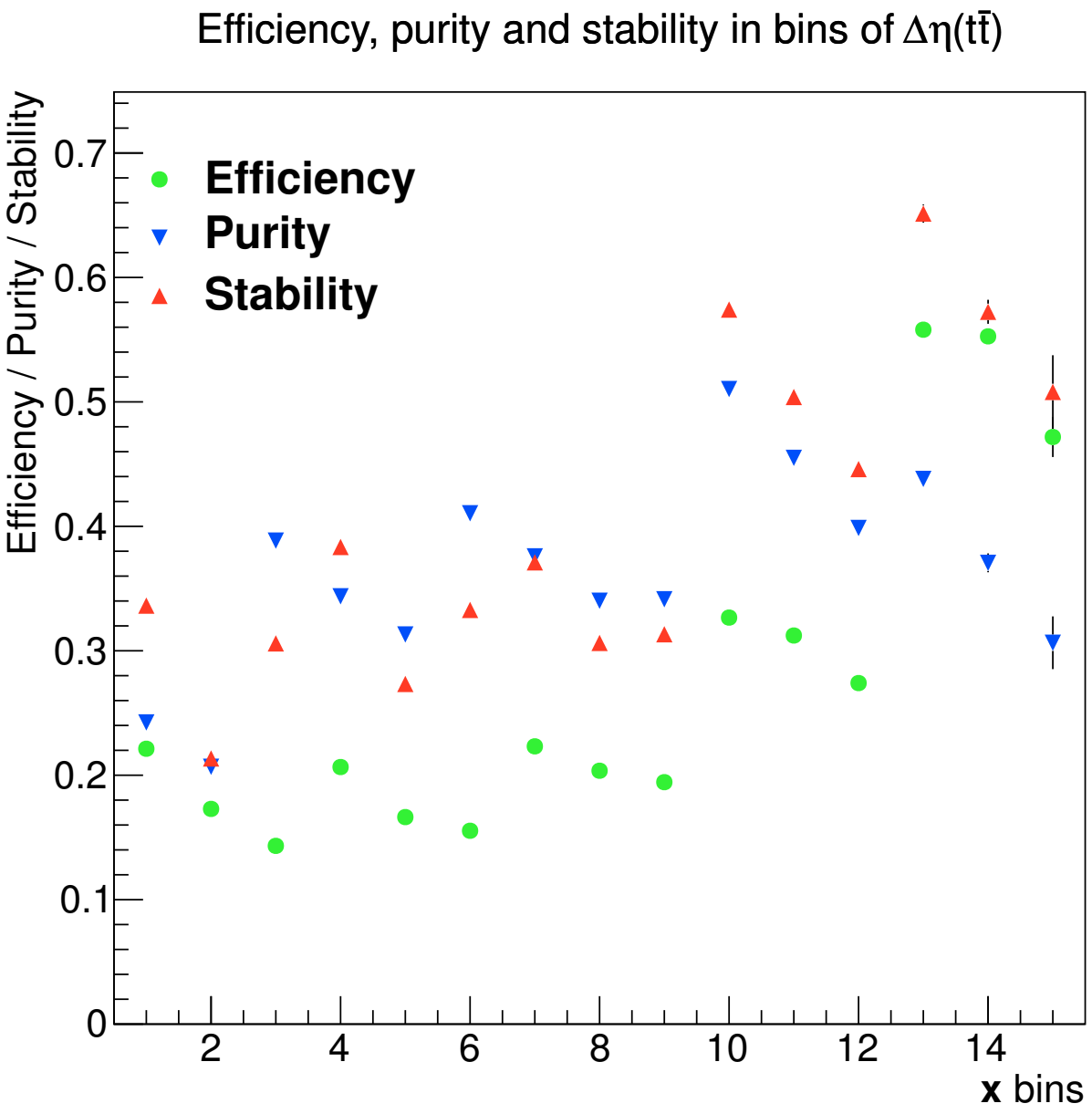

Figure D.4: Efficiencies (green circles), purities (blue invertes triangles) and stabilities (red triangles) for the $\Delta\eta(t\bar{t})$ in bins of $p_T(t)$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $p_T(t)$ – [0, 65, 130, 200, 400, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t\bar{t})|$ – [0, 0.8, 2.0, 6.0].

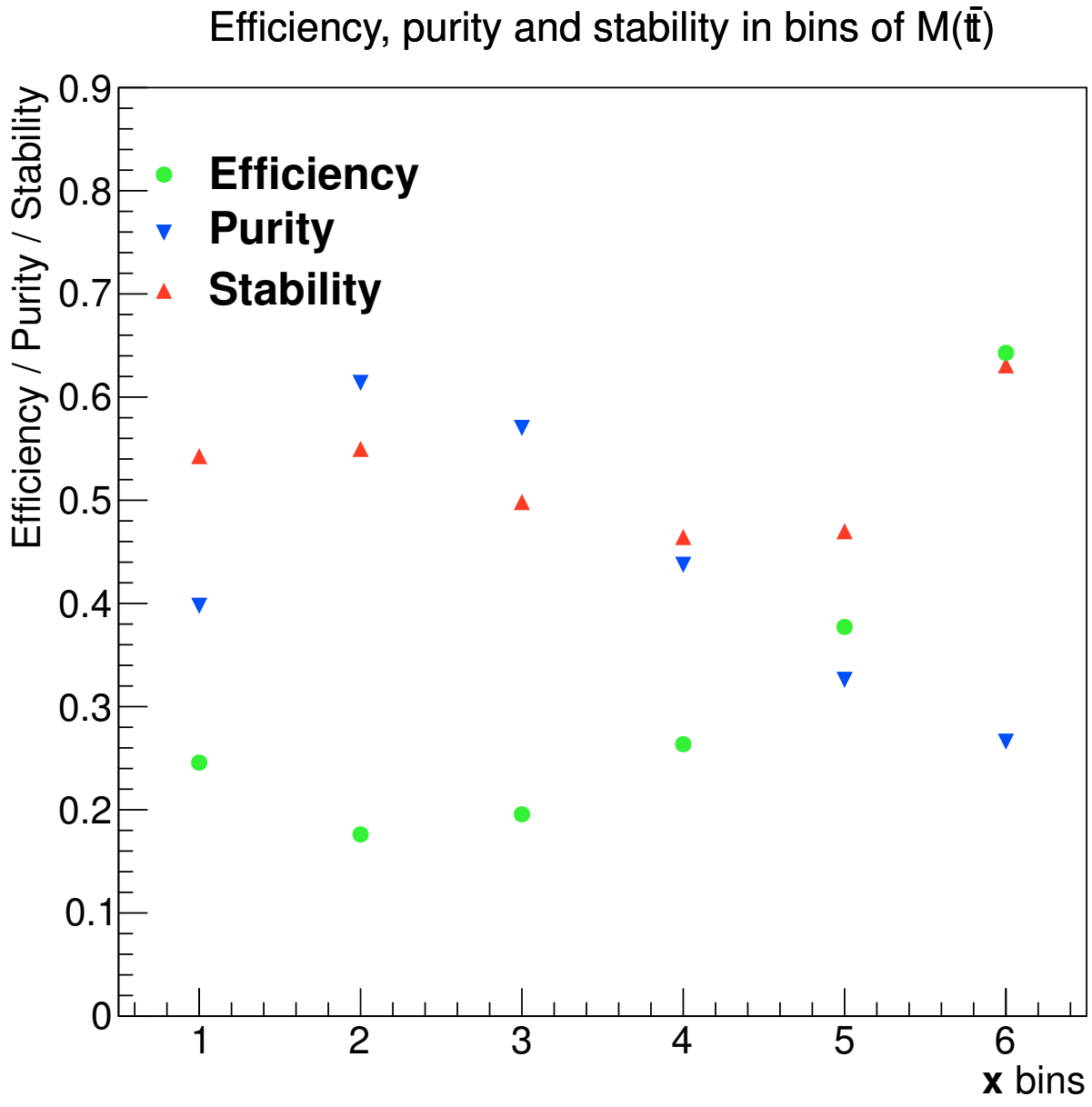

Figure D.5: Efficiencies (green circles), purities (blue invertes triangles) and stabilities (red triangles) for the $p_T(t)$. The binning is the following: [340, 380, 470, 620, 820, 1100, 2200] GeV.

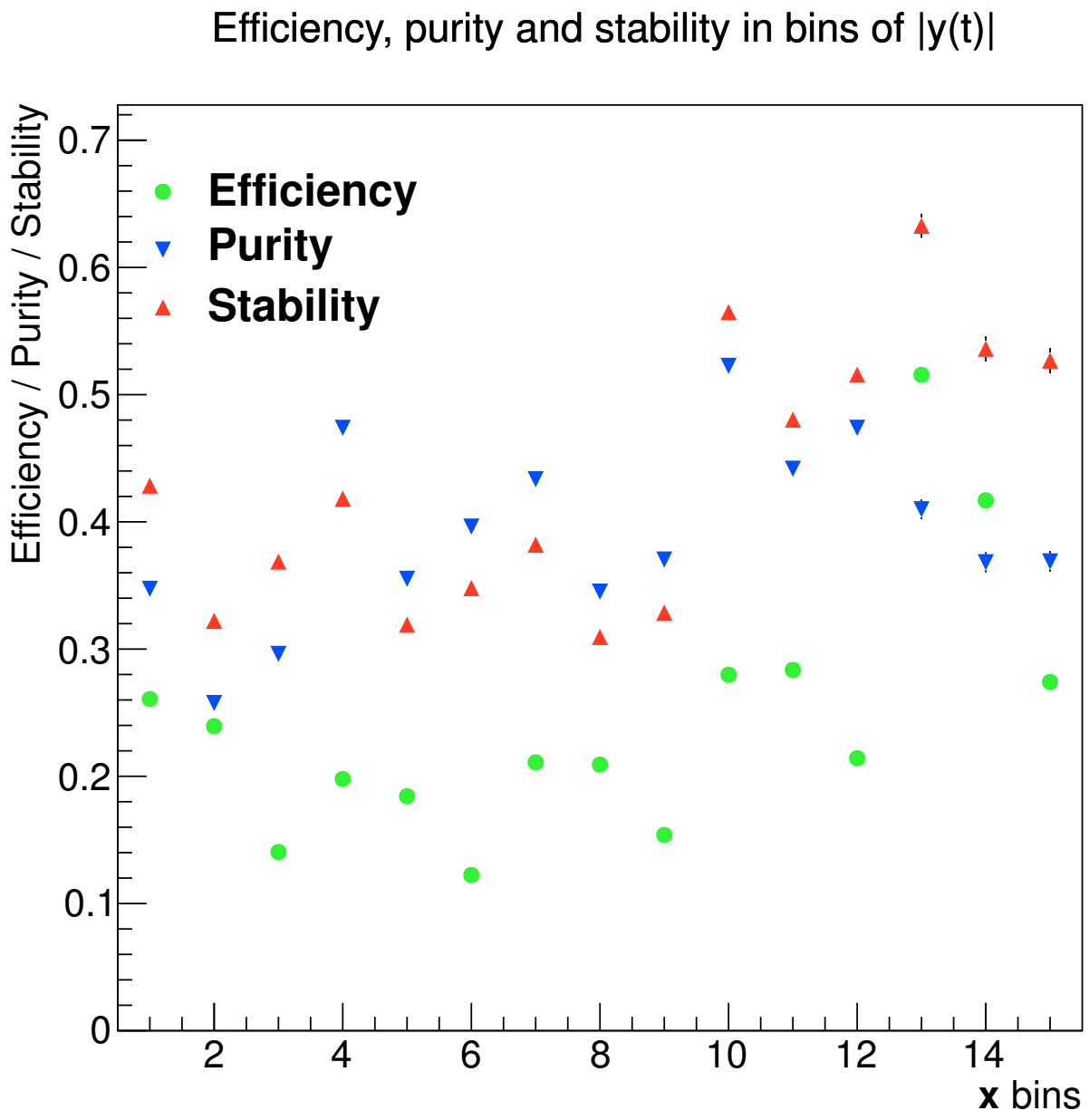

Figure D.6: Efficiencies (green circles), purities (blue invertes triangles) and stabilities (red triangles) for the $|y(t)|$ in bins of $M(t\bar{t})$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $M(t\bar{t})$ – [350, 390, 480, 600, 1000, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t)|$ – [0, 0.6, 1.2, 2.5].

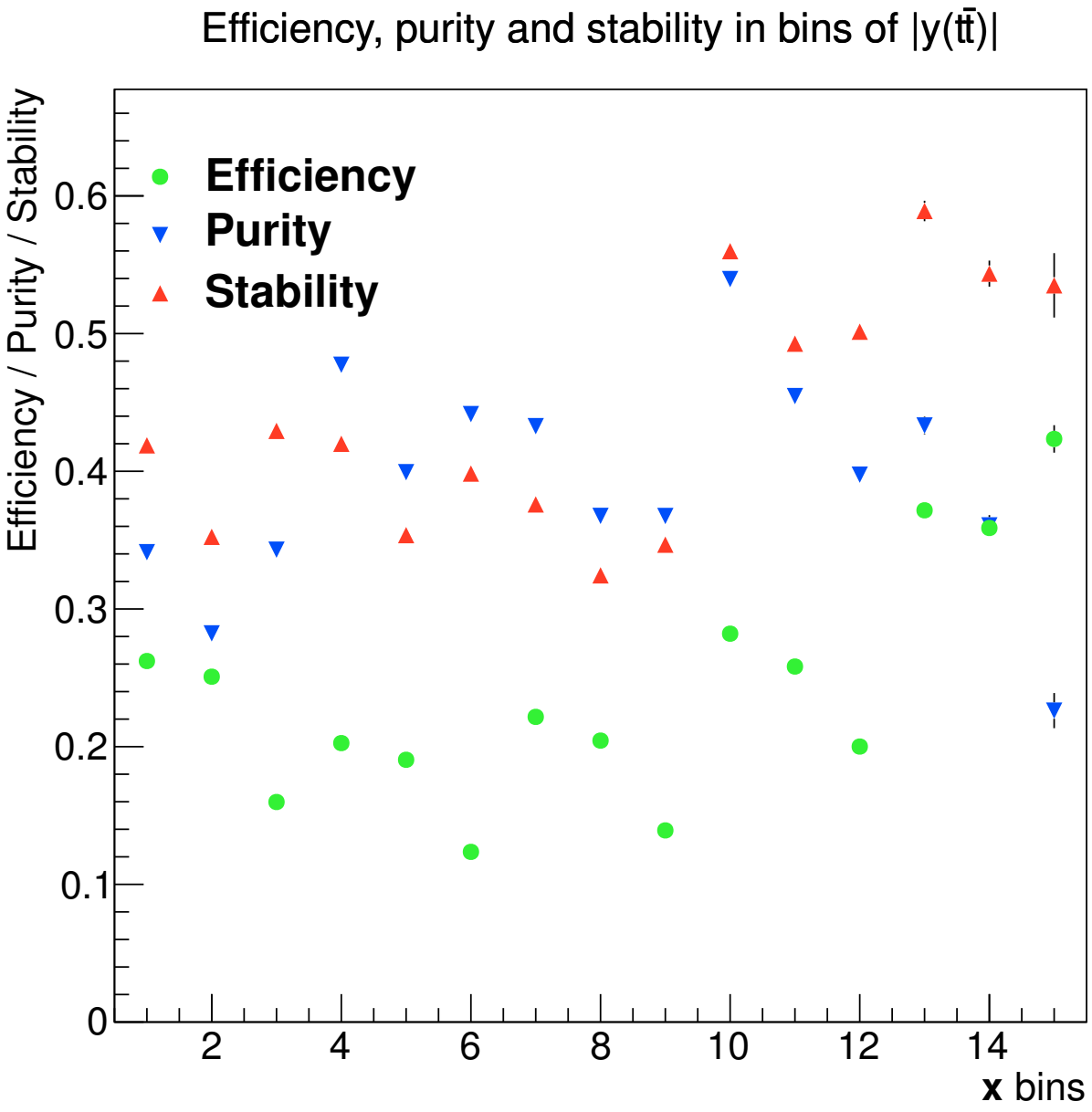

Figure D.7: Efficiencies (green circles), purities (blue invertes triangles) and stabilities (red triangles) for the $|y(t\bar{t})|$ in bins of $M(t\bar{t})$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $M(t\bar{t})$ – [350, 390, 480, 600, 1000, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t\bar{t})|$ – [0, 0.45, 1.0, 2.3].

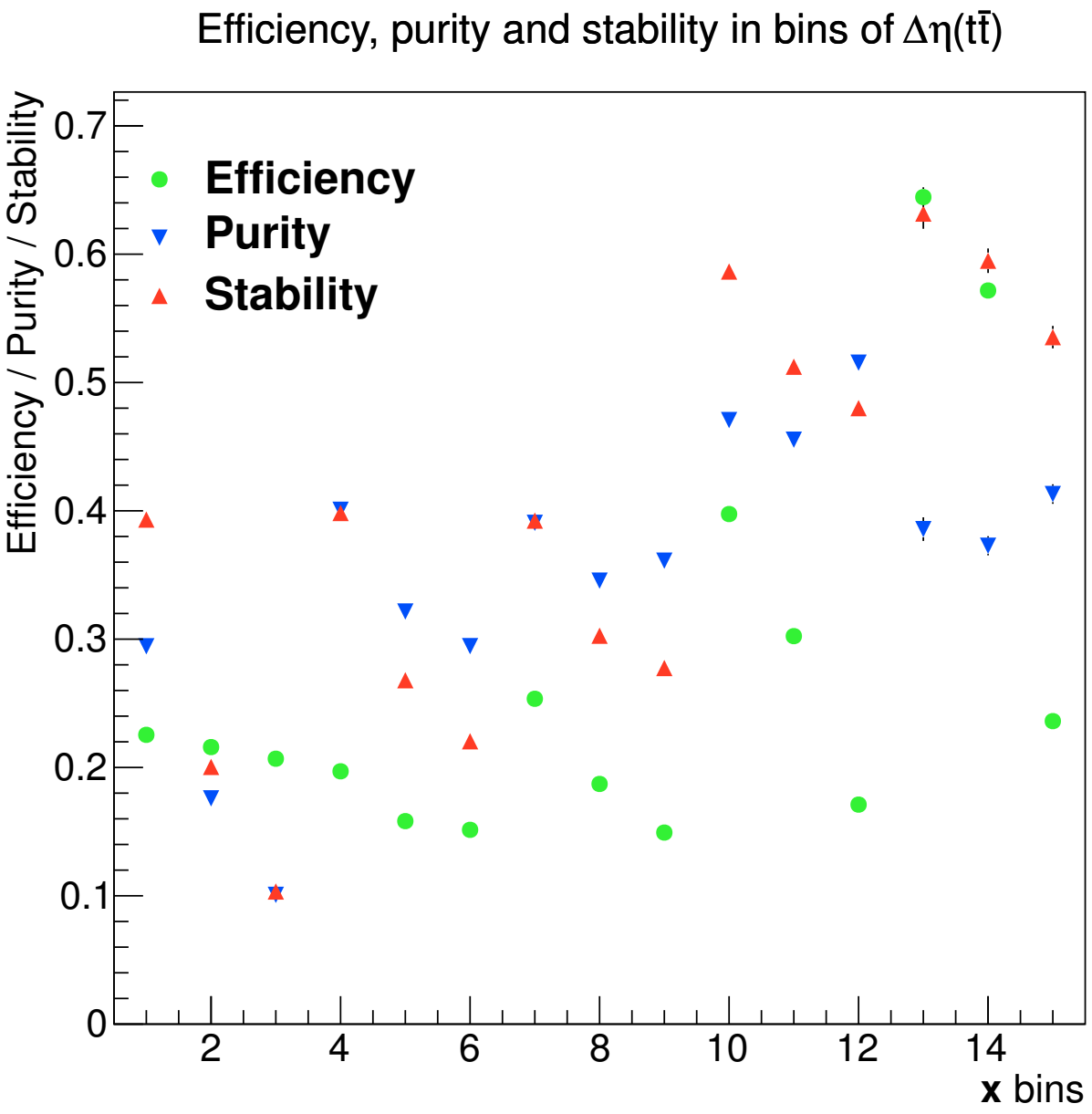

Figure D.8: Efficiencies (green circles), purities (blue invertes triangles) and stabilities (red triangles) for the $\Delta\eta(t\bar{t})$ in bins of $M(t\bar{t})$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $M(t\bar{t})$ – [350, 390, 480, 600, 1000, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t\bar{t})|$ – [0, 0.8, 2.0, 6.0].

E | Cleaning Conditions for the $M(t\bar{t})$ Spectrum

The effectiveness of the cleaning conditions described in Sec. 8.2.1 is tested in detail for the $p_T(t)$ spectrum (see Sec. 8.2.1 and Sec. 8.2). In the following the cleaning condition is examined for the $M(t\bar{t})$ spectrum.

The scatter plot for the invariant mass of the $t\bar{t}$ system without any cleaning requirement is presented in Fig. E.1. A band of the events, which were generated with lower invariant masses of the $t\bar{t}$ system ($M^{Gen}(t\bar{t}) < 1000$ GeV), but were reconstructed with high $M(t\bar{t})$ ($M^{Rec}(t\bar{t}) > 1500$ GeV), is clearly seen on the plot. The application of the cleaning condition ($M(b\bar{b}l$MET$)/M(t\bar{t}) > 0.5$) results in the removement of this band (see Fig. E.2).

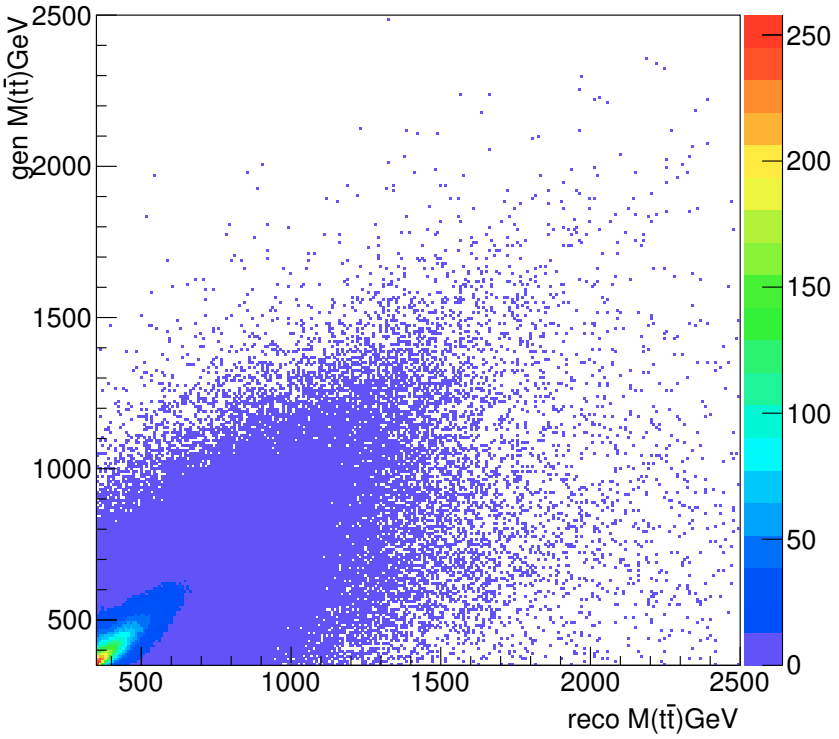

Figure E.1: Scatter plot of the $M(t\bar{t})$ in the combined dilepton channel.

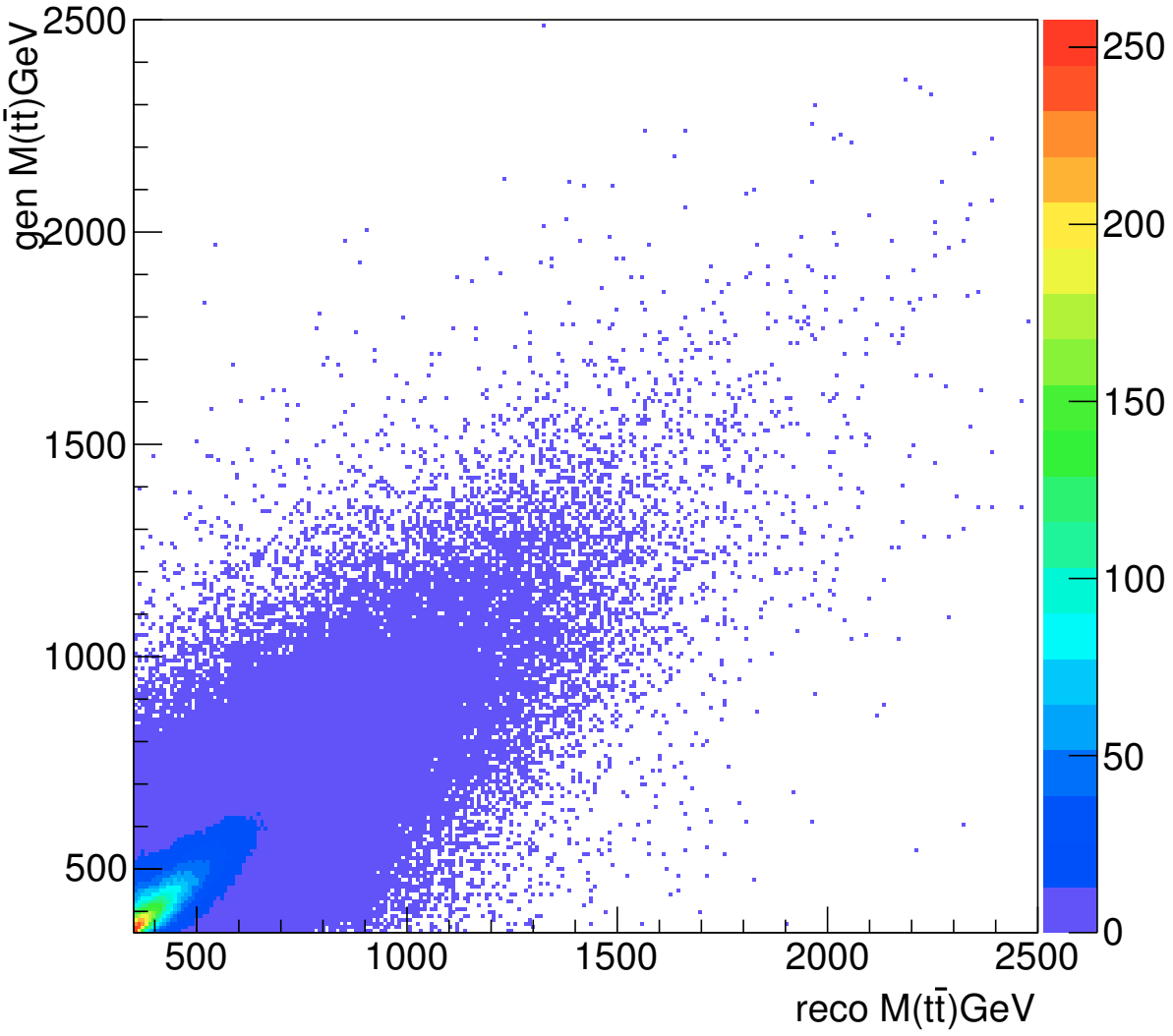

Figure E.2: Scatter plot of the $M(t\bar{t})$ after applying the cleaning condition $M(b\bar{b}l\text{MET})/M(t\bar{t}) > 0.5$ in the combined dilepton channel.

F | Comparison of the Cleaning Condition to the Unfolding

As the statistics of the events with smaller $p_T(t)$ is significantly larger than of those with boosted tops, large migrations are present in the boosted region from the low $p_T(t)$. On one hand one could leave this effect untreated and rely on the unfolding of the $p_T(t)$ spectrum. However, there are also other options, which are implemented in this analysis. First option is to apply the cleaning condition $M(b\bar{l}\bar{b}l\text{MET})/M(t\bar{t}) > 0.5$ (see Sec. 8.2.1), which reduces the migrations to the high $p_T(t)$ bins from the low $p_T(t)$. The other option is to perform simultaneous unfolding of $p_T(t)$ and $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ spectra, where only one bin of $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ is assumed on the generator level with three bins on the detector level distribution. This corresponds to the weighted averaging between the three detector bins of $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$.

The performance of three approaches is compared on the example of the measurement of the integrated cross sections for the region with $p_T(t) > 400$ GeV:

- **No cleaning condition applied**: for the case with no cleaning condition it was discussed that there is a considerable fraction of events which have a large reconstructed transverse momentum of the top, while originally the top quark was produced with lower transverse momentum (see Sec. 8.2). The same migrations are also observed for the invariant masses of the $t\bar{t}$ system (see Appendix E). The integrated cross section obtained for this case results in the following:

$$\sigma^{\text{no cleaning}}_{p_T>400\text{ GeV}}(t\bar{t}) = 1.3\text{ pb} \pm 10.7\%\text{ (Stat.)}\,^{+15.9\%}_{-17.7\%}\text{ (Syst.)}, \tag{F.1}$$

- **Applying cleaning condition**: the cleaning condition rejects a large number of badly reconstructed events (see Sec. 8.2.1). The integrated cross section measured after applying this cut is the following:

$$\sigma^{\text{with cleaning}}_{p_T>400\text{ GeV}}(t\bar{t}) = 1.2\text{ pb} \pm 9.4\%\text{ (Stat.)}\,^{+13.7\%}_{-14.5\%}\text{ (Syst.)}, \tag{F.2}$$

- **Performing simultaneous unfolding with $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ spectrum**: the migrations from lower to higher $p_T(t)$ region are corrected by the double differential unfolding and the measured integrated cross section results in the following:

$$\sigma^{\text{with unfolding}}_{p_T>400\text{ GeV}}(t\bar{t}) = 1.1\text{ pb} \pm 8.3\%\text{ (Stat.)}\,^{+7.3\%}_{-8.9\%}\text{ (Syst.)}, \tag{F.3}$$

Although the unfolding provides overall better results, it was decided not to have an unfolding with a dimension larger than two. That is why the simultaneous unfolding with $M(\bar{l}bl\bar{b}MET)/M(t\bar{t})$ is performed only for the measurement of the cross sections as function of the $p_T(t)$ threshold and as function of $M(t\bar{t})$. The cross section in bins of other variables require additional simultaneous unfolding in $p_T(t)$ or $M(t\bar{t})$, thus for their measurement the cleaning condition is applied.

G | Control Distributions

Control distributions which correspond to all measurements presented in Sec. 11.2, Sec. 11.4 and Appendix H are presented in the following.

The same tendencies, as for the cross sections (see Sec. 11.3 - 11.5), are observed on the detector level control distributions. Moreover, these tendencies are more clearly seen as the number of bins in the control distributions is larger than for the cross sections measurements. Many trends also look less significant after the unfolding procedure applied for the cross section measurements. It is observed that all the trends in description of the data by MC are enhanced for the high $p_T(t)$ (also presented in Appendix H).

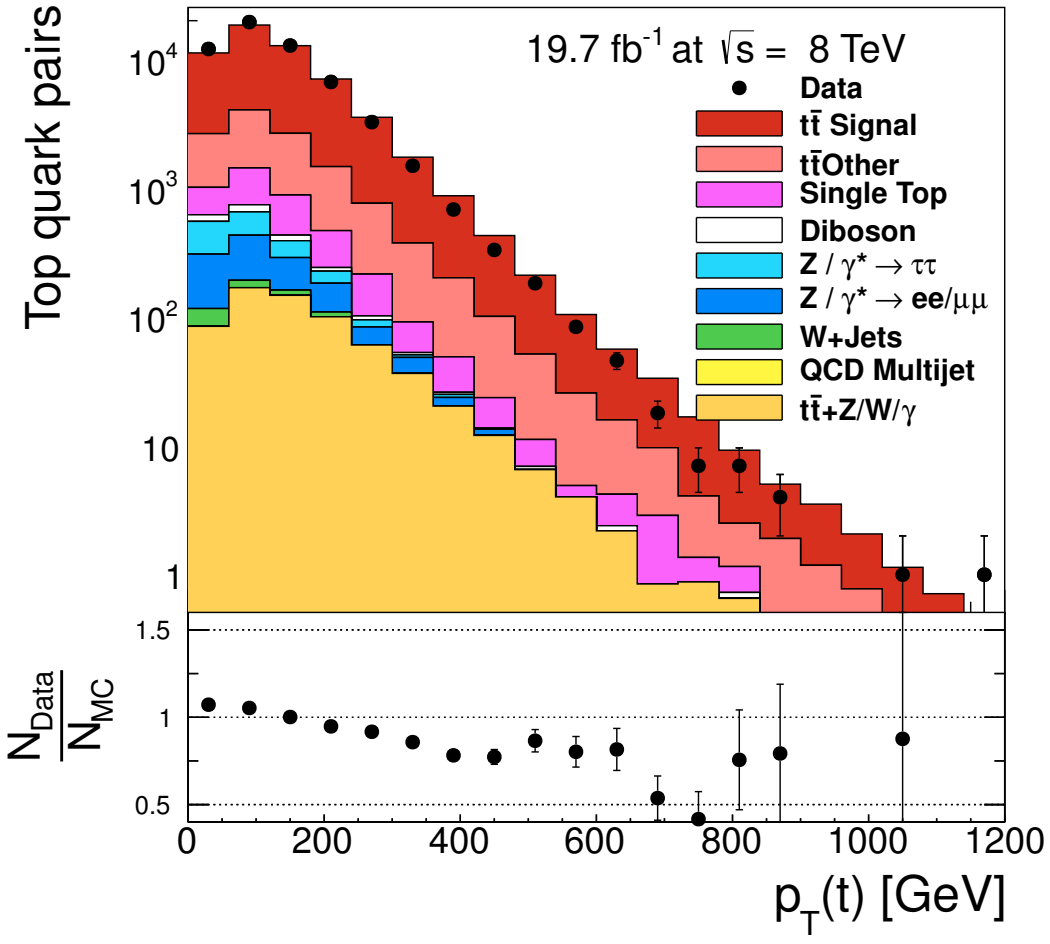

Figure G.1: Control distribution of $p_T(t)$ after applying all selection cuts described in Chapter 6 and the cleaning condition described in Sec. 8.2.1. Black dots show the data points and filled histograms represent the MC. Histograms with different colors represent contributions from different background processes. The ratio of data over simulation is presented in the bottom panel of each distribution. The error bars represent the statistical uncertainties of the data.

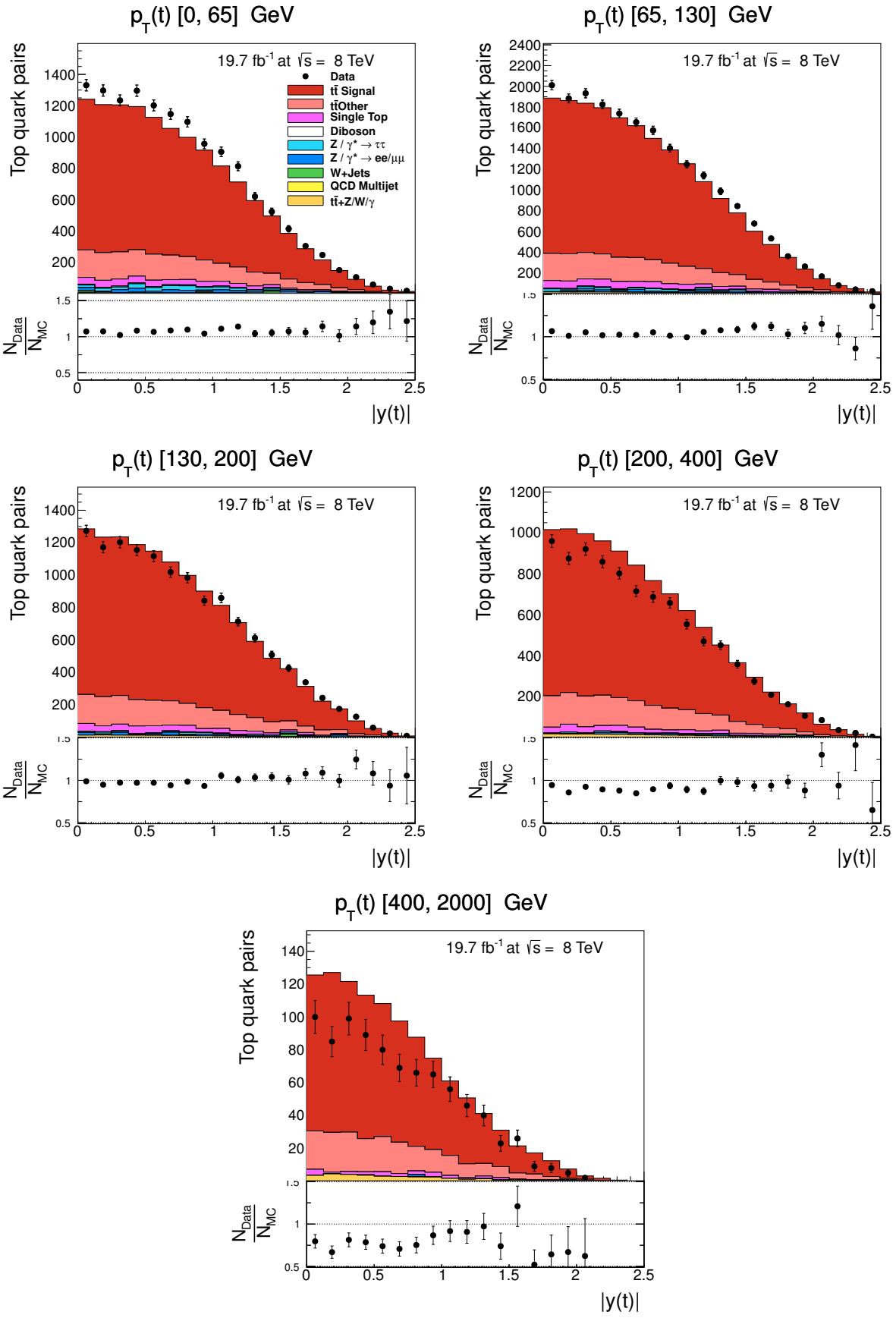

Figure G.2: Control distributions of $|y(t)|$ in different bins of $p_T(t)$ after applying all selection cuts described in Chapter 6. Other details as in Fig. G.1.

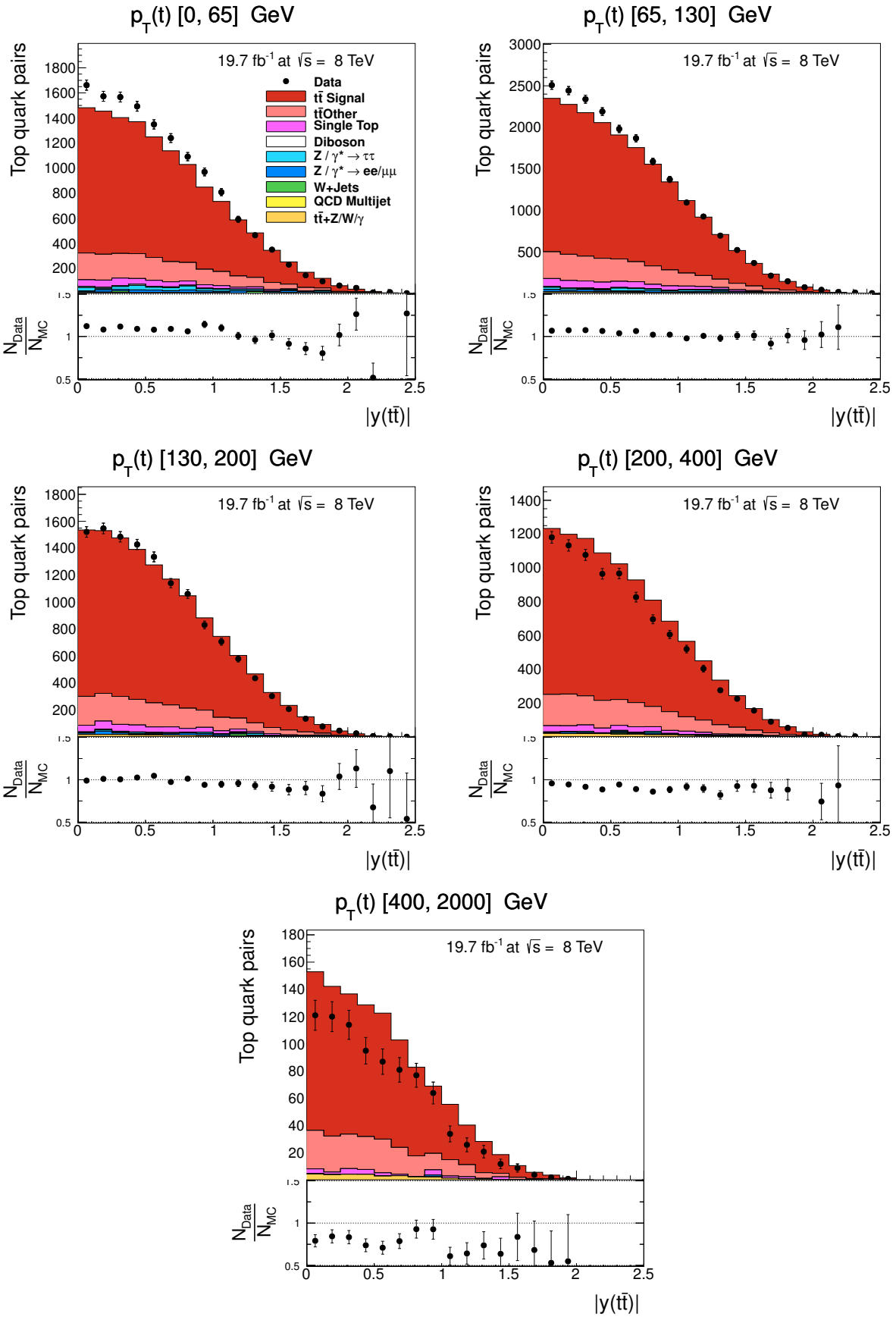

Figure G.3: Control distributions of $|y(t\bar{t})|$ in different bins of $p_T(t)$ after applying all selection cuts described in Chapter 6. Other details as in Fig. G.1.

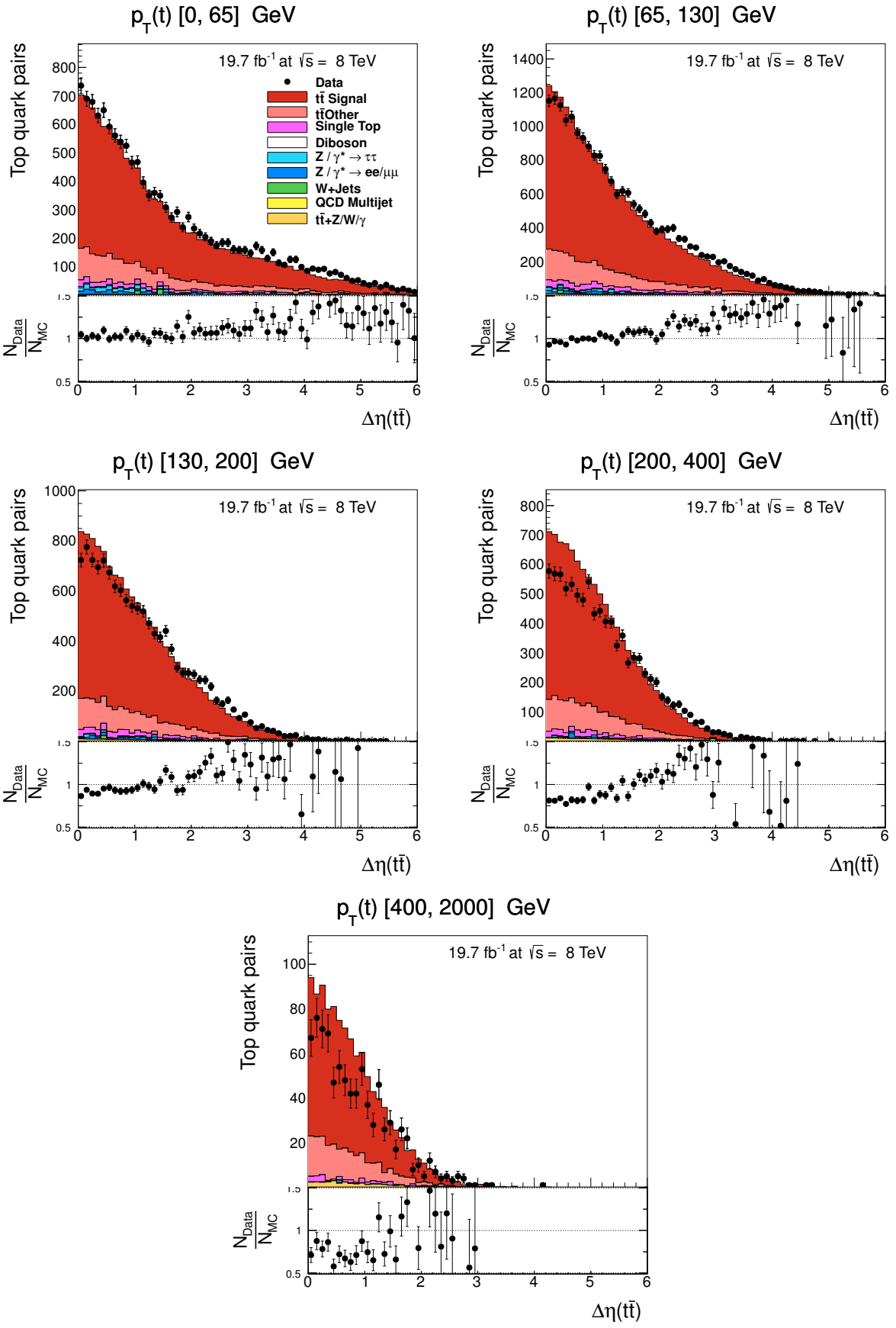

Figure G.4: Control distributions of the $|\Delta\eta(t\bar{t})|$ in different bins of $p_T(t)$. Other details as in Fig. G.1.

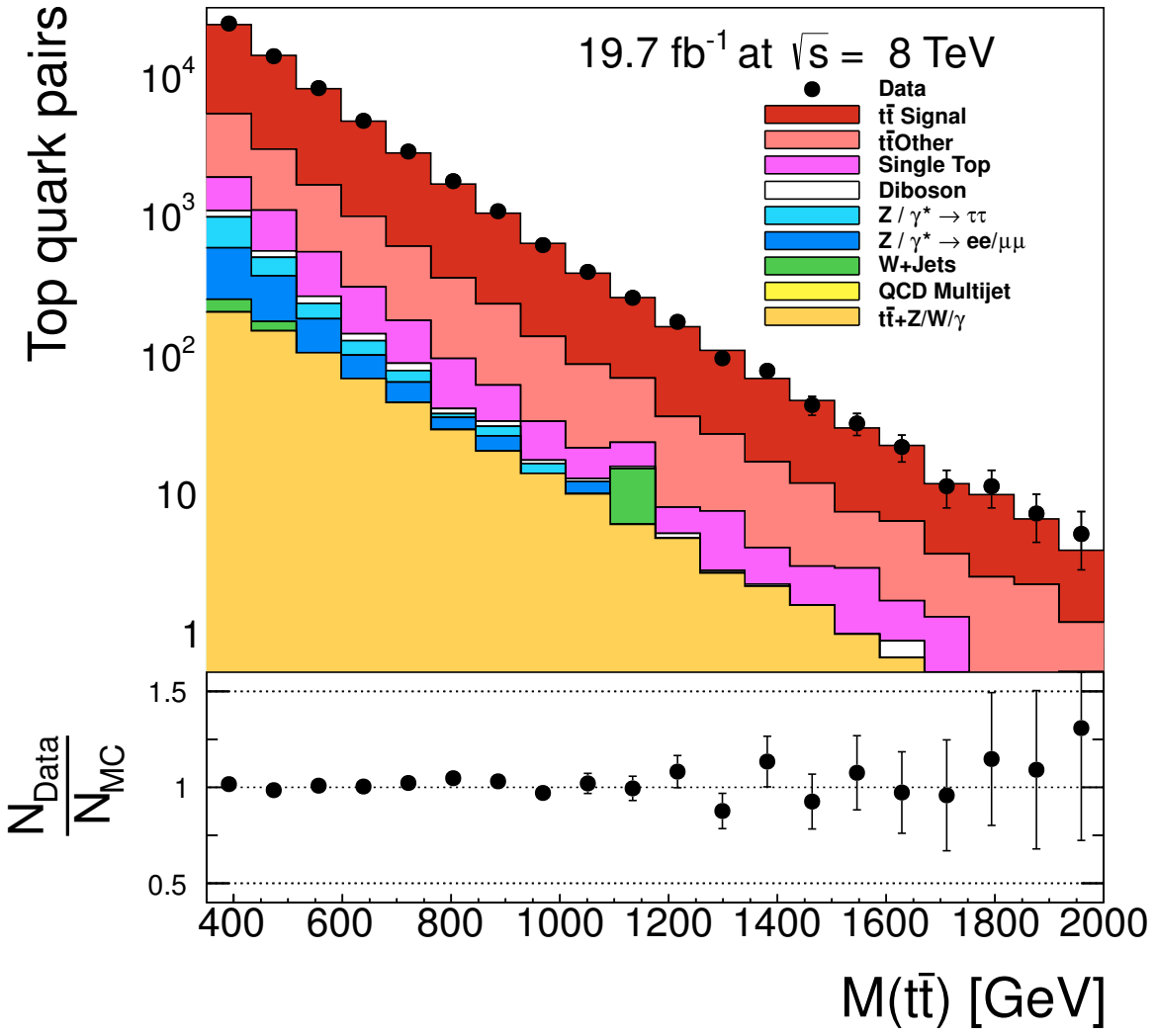

Figure G.5: Control distribution of $M(t\bar{t})$ after applying all selection cuts described in Chapter 6 and the cleaning condition described in Sec. 8.2.1. For other details see Fig. G.1.

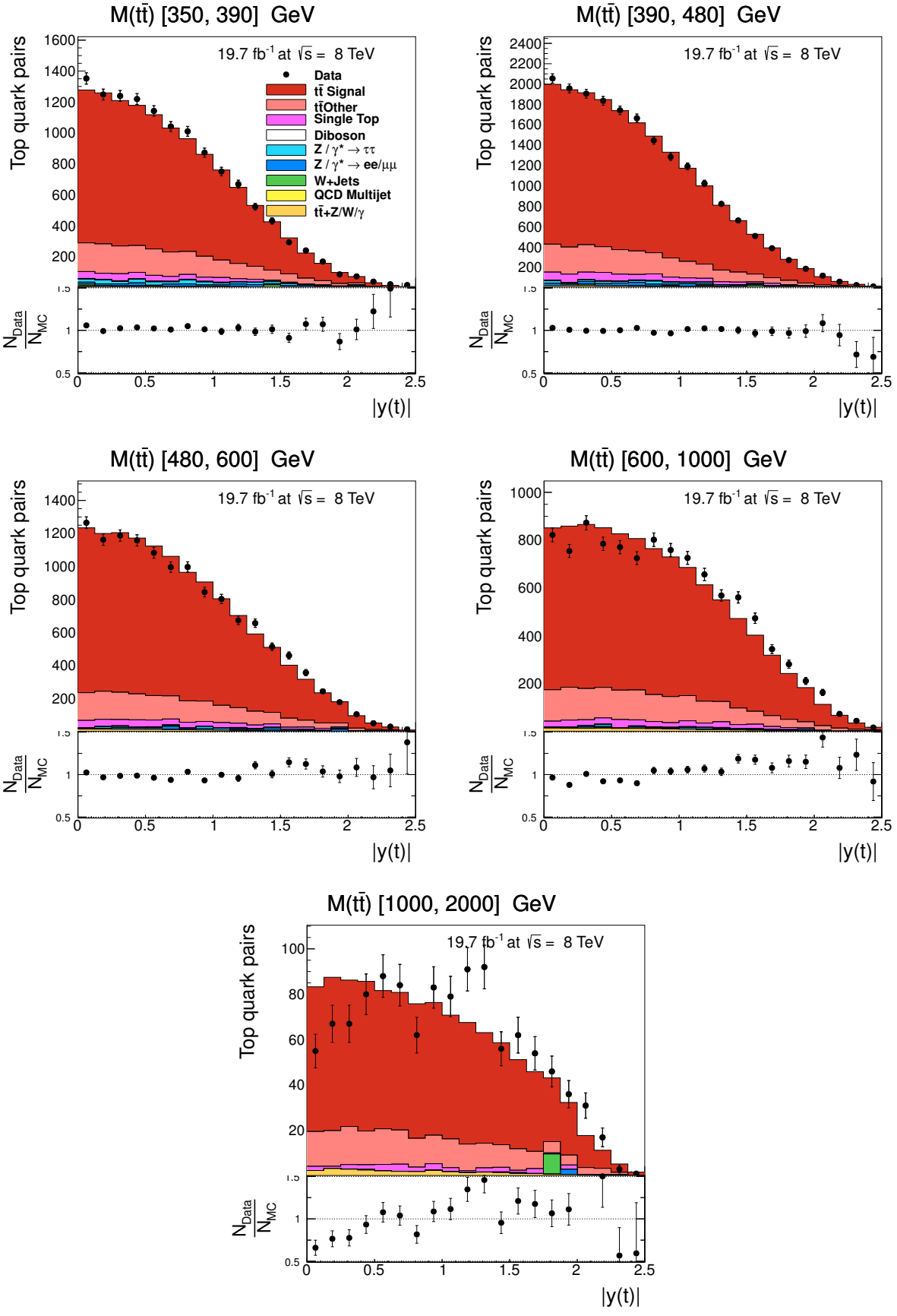

Figure G.6: Control distributions of $|y(t)|$ in different bins of $M(t\bar{t})$ after applying all selection cuts described in Chapter 6. Other details as in Fig. G.1.

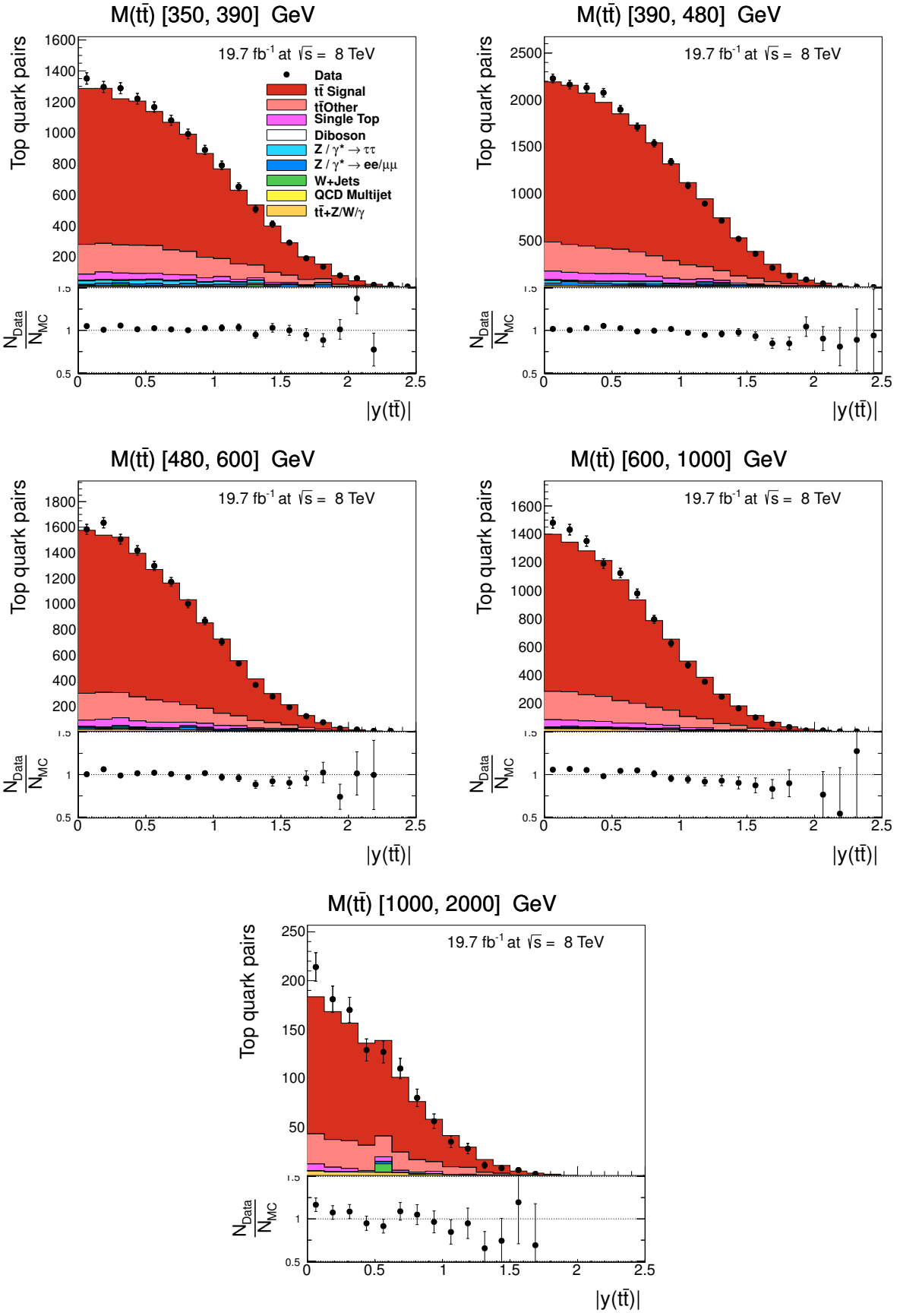

Figure G.7: Control distributions of $|y(t\bar{t})|$ in different bins of $M(t\bar{t})$ after applying all selection cuts described in Chapter 6. Other details as in Fig. G.1.

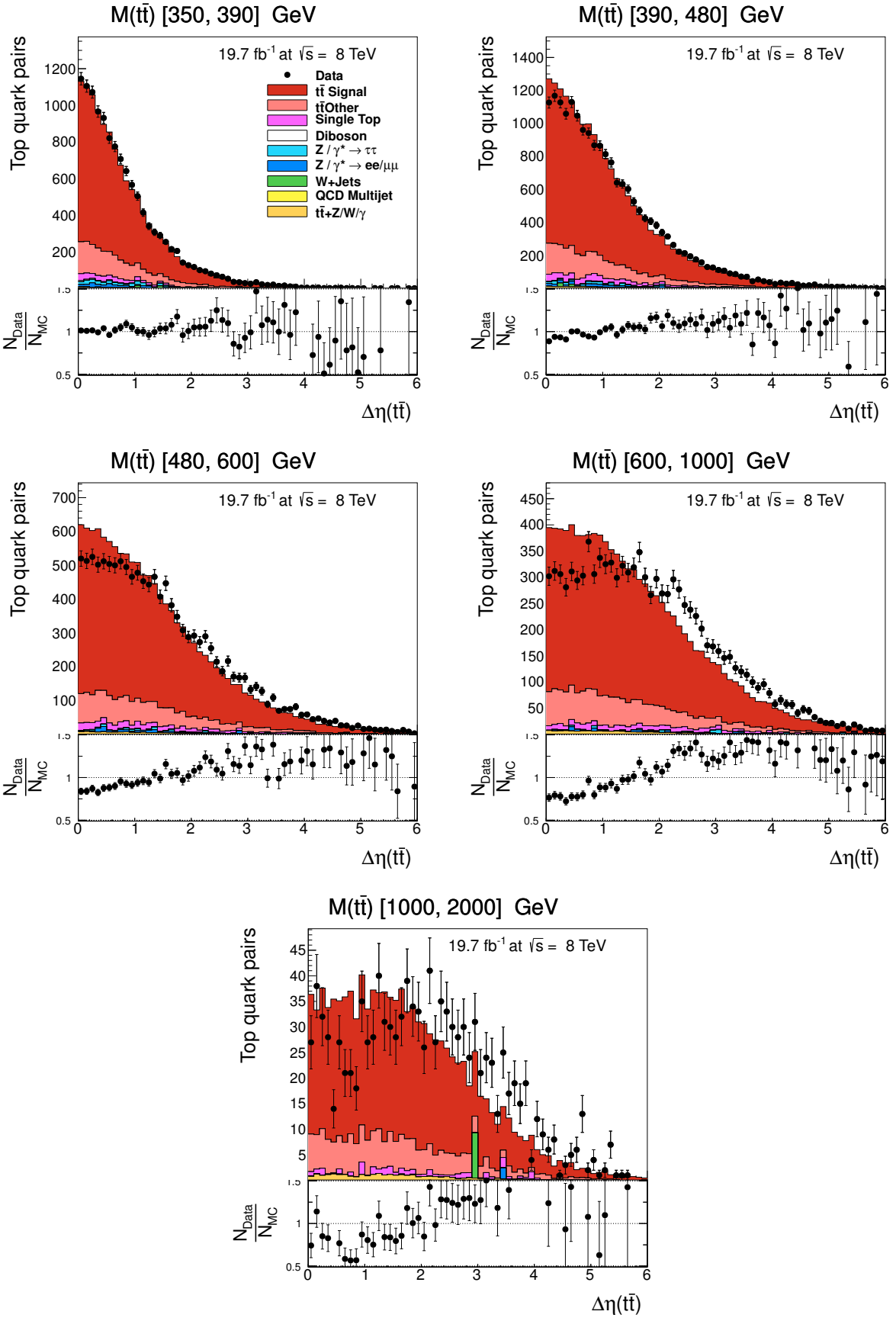

Figure G.8: Control distributions of $|\Delta\eta(t\bar{t})|$ in different bins of $M(t\bar{t})$ after applying all selection cuts described in Chapter 6. Other details as in Fig. G.1.

H | Cross Sections for Different Bins of $p_T(t)$

H.1 Cross Sections: Plots

The cross sections in the boosted region ($p_T > 400$ GeV) are presented in Sec. 11.3.1-11.3.3. To obtain each of those cross sections a double differential unfolding simultaneously with the $p_T(t)$ distribution has been performed. Therefore, there are also measurements for the regions, where $p_T(t)$ is smaller than 400 GeV. The cross sections in bins of $|y(t)|$, $|y(t\bar{t})|$ and $\Delta\eta(t\bar{t})$ for all the five unfolded bins of $p_T(t)$ are presented in Fig. H.1-H.3.

H.2 Tables with Cross Section Numbers

In the following the tables with the numbers for cross sections shown in Sec. 11.2 and Sec. H.1 are presented. The corresponding uncertainties are also listed in the tables.

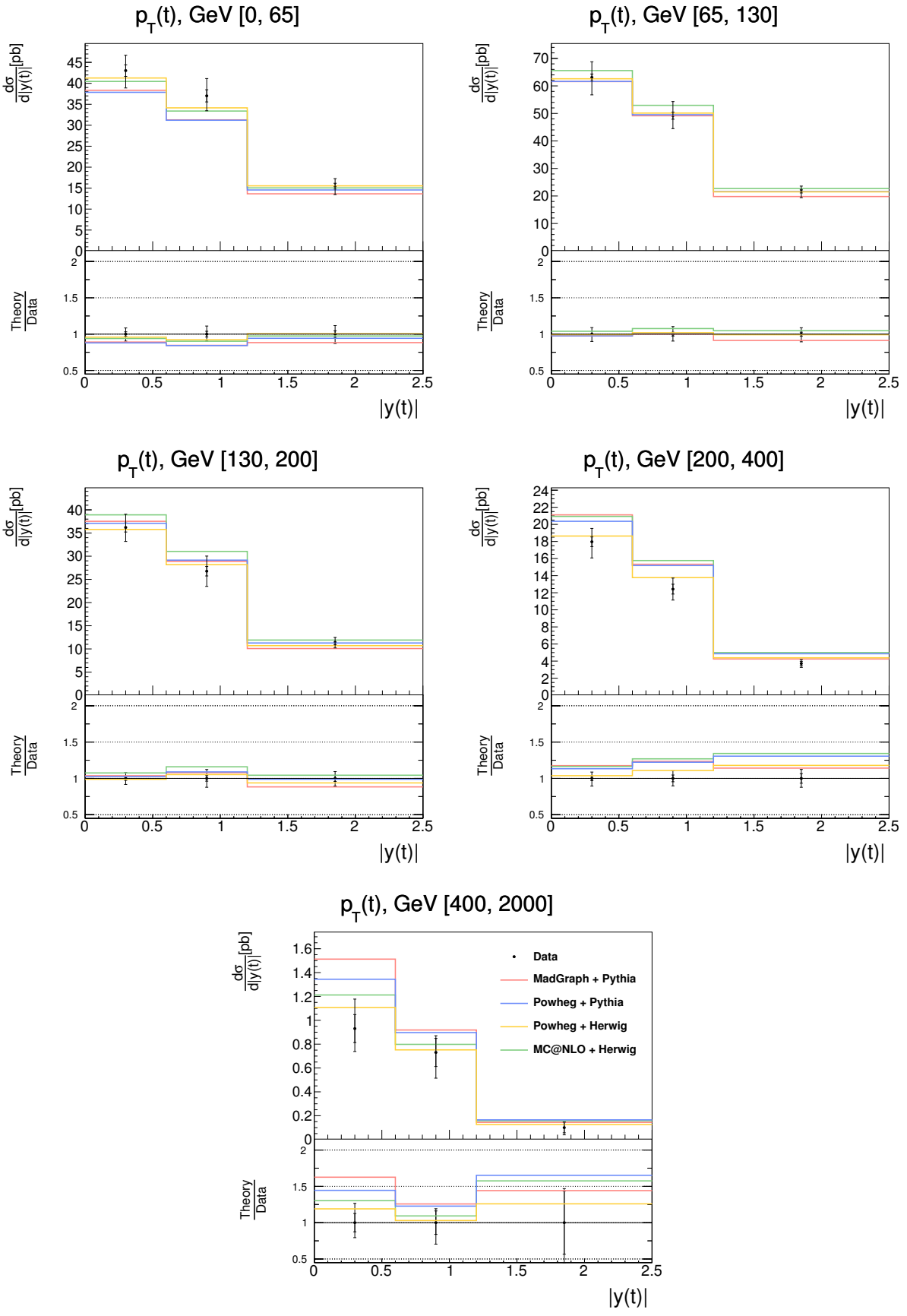

Figure H.1: The $t\bar{t}$ production cross sections in bins of $|y(t)|$ for different $p_T(t)$ regions. The other details as in Fig. 11.3.

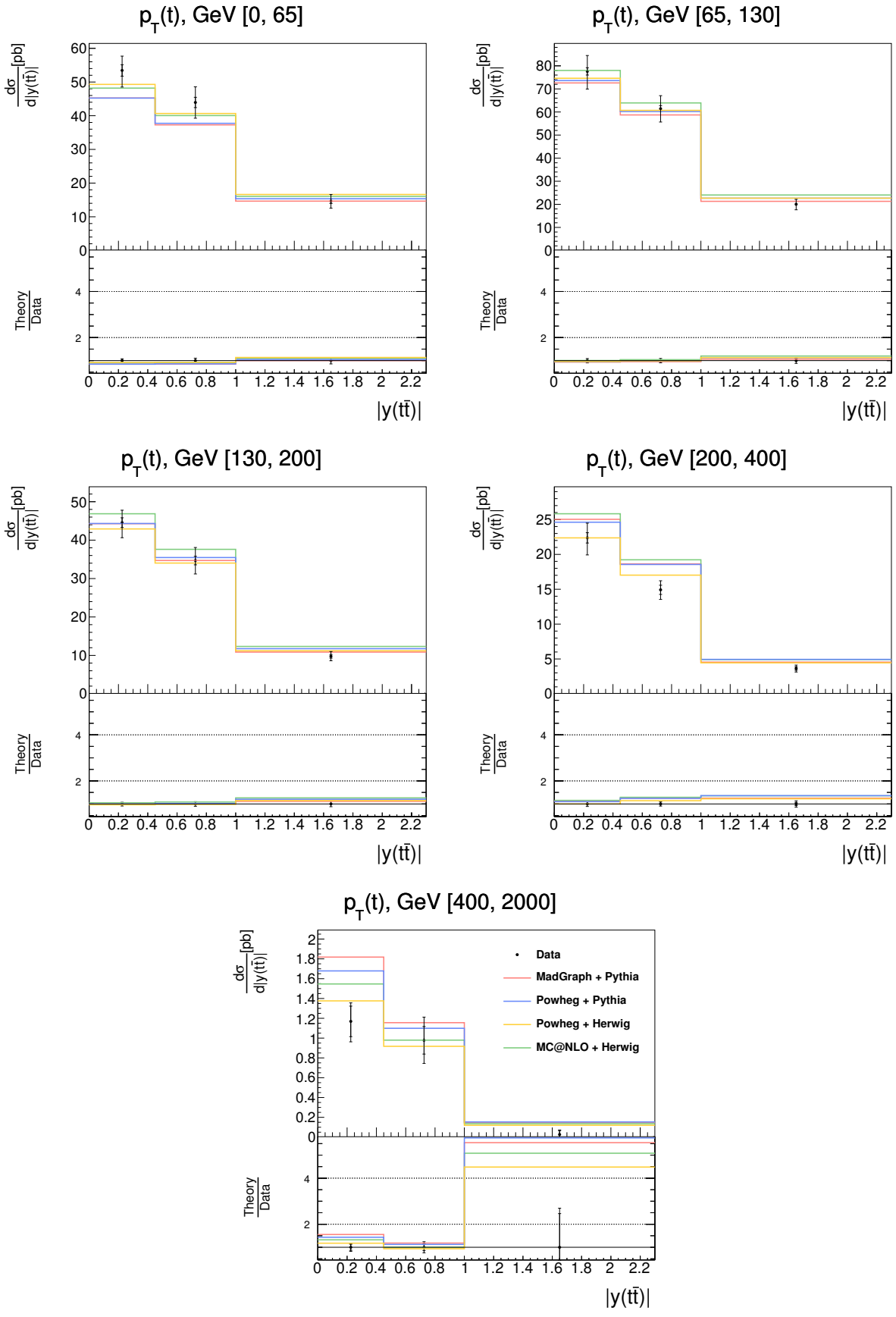

Figure H.2: The $t\bar{t}$ production cross sections in bins of $|y(t\bar{t})|$ for different $p_T(t)$ regions. The other details as in Fig. 11.3.

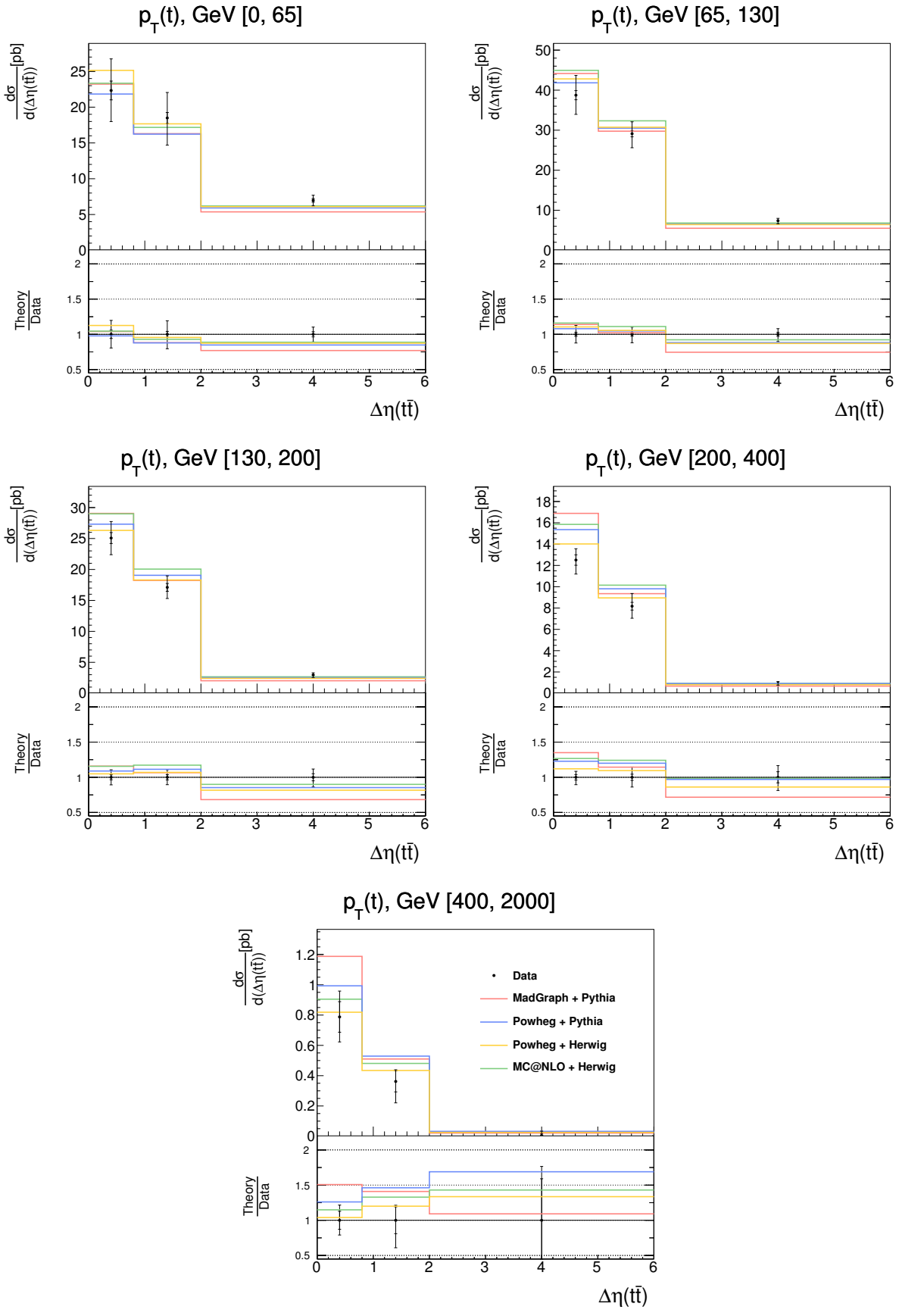

Figure H.3: The $t\bar{t}$ production cross sections in bins of $\Delta\eta(t\bar{t})$ for different $p_T(t)$ regions. The other details as in Fig. 11.3.

$p_T(t)$	0 to 65	65 to 125	125 to 200	200 to 290	290 to 400	400 to 2000
$\frac{d\sigma}{dp_T(t)}$ $[pb]$	1.06	1.51	0.76	0.216	0.04	0.0007
Source	Uncertainty [%]					
Statistical	1.6	1	1.4	2.8	6.1	8.3
Total Syst. Pos.	7.9	9	8.5	7.4	12.7	7.3
Total Syst. Neg.	9.3	9.5	9.4	8.7	13.5	8.9
Lumi Up	-2.6	-2.7	-2.7	-2.7	-2.7	-2.8
Down	2.8	2.9	2.8	2.8	2.8	2.9
b-tagging Pos.	0.8	0.8	0.9	0.9	1	1.4
Neg.	1.2	0.7	0.8	1	0.9	1.4
Kin Up	-0.33	-0.34	-0.35	-0.35	-0.37	-0.38
Down	0.33	0.35	0.35	0.35	0.37	0.38
JER Up	0.7	0	0.1	0.5	-0.1	1.3
Down	-1.1	-0.7	-0.7	-1.5	-0.2	-2.7
JES Up	-1.7	-2.9	-3.1	-1.5	-1	-1.9
Down	1.4	3.8	3	1.2	1.2	2.1
PU Up	0.5	0.4	0.3	-0	-0.1	0
Down	-0.5	-0.4	-0.3	-0.1	0.1	-0
Trig Up	-1.3	-1.3	-1.3	-1.2	-1.2	-1.2
Down	1.3	1.3	1.3	1.2	1.2	1.2
Lept Up	-2.1	-2.2	-2.2	-2.1	-2.2	-2.2
Down	2.2	2.3	2.2	2.2	2.3	2.3
BG Up	3.5	1.7	2	2.4	2.5	1.6
Down	-6.2	-4.3	-4.5	-5.2	-5.5	-5.1
DY Up	-0.8	-0.3	-0.1	-0.1	-0	0
Down	0.9	0.3	0.1	0.1	0	-0
Mass Up	0.1	-1.4	-1.3	0.3	0.3	1.1
Down	-0.1	1.9	1.3	-0.8	-0.6	-0.6
Match Up	0.3	-0.4	0.3	1.4	1	0.6
Down	1.1	-1.7	-0	-0	0.6	-0.8
Scale Up	1.9	2.9	1.7	0.1	1	-3.1
Down	-1.3	-2.6	-2.1	0.3	-0.8	3.3
PDF Pos.	0.5	0.3	0.2	0.2	0.3	0.1
Neg.	0.5	0.3	0.2	0.2	0.3	0.2
Hadronization	-2.5	-4	-4.8	-4.1	-10.7	-3.4
Hard Scat.	4.6	4.3	3.9	3.7	4.4	2

Table H.1: Cross sections as a function of $p_T(t)$. The table also presents the summary of all the statistical and systematic uncertainties.

$p_T(t)$ [GeV]	0 to 65			65 to 130			130 to 200			200 to 400			400 to 2000		
$\lvert y(t)\rvert$	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5
$\frac{d\sigma}{d\lvert y(t)\rvert}$ $[pb]$	43	37	15.5	63.1	49.2	21.7	36.2	26.8	11.4	18	12.4	3.7	0.9	0.7	0.1
Source	Uncertainty [%]														
Statistical	3.2	3.9	4.6	2	2.6	2.7	2.7	3.8	4.2	3.2	4.6	6.6	12.6	16.1	43.4
Total Syst. Pos.	7.9	10.4	10.9	8.8	10.1	8.4	7.4	11.6	8.4	8.1	9.3	10.4	23.4	10.3	17.7
Total Syst. Neg.	9.1	8.8	12.1	9.9	9.2	10.3	7.9	11.6	9.6	10.1	9.2	10.3	16.4	24.7	42.2
Lumi Up	-2.6	-3.2	-2.5	-2.4	-2.9	-2.3	-2.7	-2.9	-2.8	-2.7	-2.6	-2.6	-2.8	-2.9	-2.9
Down	2.7	3.4	2.6	2.5	3.1	2.5	2.9	3	2.9	2.8	2.7	2.8	2.9	3	3.1
b-tagging Pos.	0.9	1.1	0.7	0.8	1	0.9	0.9	1.2	1	1	1	0.7	2.1	1.8	2.4
Neg.	0.8	1.7	1.9	0.8	0.9	1.2	1	0.8	1.1	0.9	1.1	1	1.6	1.2	1.4
Kin Up	-0.31	-0.4	-0.31	-0.3	-0.39	-0.3	-0.35	-0.37	-0.34	-0.36	-0.34	-0.37	-0.37	-0.39	-0.36
Down	0.32	0.4	0.31	0.3	0.39	0.3	0.35	0.37	0.35	0.36	0.34	0.37	0.38	0.39	0.37
JER Up	0.6	1.7	-0	-0.4	0.3	0	-0	0.4	0.6	0.6	0.2	0.4	-1.9	-1	-15.7
Down	-0.7	-1.1	-1.1	-1.2	-0.1	-0.6	-1	-0.6	-1.8	-1.5	-0.3	-0.8	-2	-5.1	-14.6
JES Up	-0.2	-2.4	-3.1	-2.8	-4	-2.7	-2.2	-3.9	-3.7	-1.2	-0.5	0.2	-1.1	-1.8	-18.3
Down	0.9	2.4	1	2.9	5.1	3.6	2.6	3.7	3.5	0.8	1.5	2	-0.4	-1.6	-12.2
PU Up	0.1	0.8	0.7	0.3	0.3	0.5	0.4	0	0.5	-0.2	0.2	-0	0.2	0.1	-1.1
Down	-0.1	-0.7	-0.5	-0.3	-0.4	-0.6	-0.4	0	-0.5	0.1	-0.2	0.1	-0.2	0	0.6
Trig Up	-1	-1.5	-1.6	-1	-1.4	-1.5	-1.2	-1.2	-1.6	-1.2	-1.1	-1.5	-1.1	-1.2	-1.3
Down	1.1	1.5	1.7	1	1.4	1.5	1.2	1.3	1.6	1.2	1.1	1.6	1.2	1.3	1.3
Lept Up	-2	-2.6	-2	-1.9	-2.5	-1.9	-2.2	-2.3	-2.2	-2.2	-2.1	-2.2	-2.2	-2.3	-2.2
Down	2.1	2.7	2	2	2.6	2	2.2	2.4	2.2	2.2	2.1	2.3	2.3	2.4	2.2
BG Up	4	-2.4	5.3	4.8	-0.8	6	1.1	-0.6	0.8	2.6	3.7	3.3	2.6	1.5	0.5
Down	-6.2	-1.6	-7.7	-6.8	-2.4	-7.9	-3.6	-2.4	-3	-5.4	-6.4	-6.5	-6.5	-5.6	-3.9
DY Up	-0.6	-1.2	-0.9	-0.2	-0.3	-0.3	-0	-0.2	-0.4	-0.1	-0	0.1	0.1	0.1	-0
Down	0.6	1.2	0.9	0.2	0.3	0.3	0	0.2	0.4	0.1	0	-0	-0.1	-0.1	0
Mass Up	-0	0.1	0	-1	-1.9	-1.6	-1.2	-1.3	-1.2	0.5	0.5	2	2.6	1.1	-6.3
Down	-0.2	-0.5	0.6	2.1	2.6	1.2	1.7	2	0.8	-1.2	-0.8	-1	-1	-0.9	-6.2
Match Up	0.7	0.2	-1.5	-0.4	0.1	-1.7	0.3	1.8	0	-1.9	4.2	0.1	10.1	-4.8	-19.5
Down	0.4	2.8	5	-1	0.2	-1.8	0.7	-0.4	-1.1	-0.5	2.7	2.9	2.7	-13.4	-6.2
Scale Up	-1.1	5.5	1.5	1.9	4.5	2.1	3	-0.2	0.4	-2.5	3.1	4.2	5.6	-16.5	-14.2
Down	1	-1.7	-3.2	-1.8	-3.5	-3.7	-2.8	-0.2	-2.8	-1.2	0.6	1	14.5	-2.9	-21.4
PDF Pos.	0.7	0.8	0.4	0.2	0.5	0.2	0.2	0.4	0.5	0.2	0.3	0.2	0.6	0.3	0.6
Neg.	0.8	0.7	0.5	0.2	0.4	0.1	0.2	0.4	0.4	0.2	0.2	0.3	0.6	0.3	0.6
Hadronization	-4.7	-3.6	1	-3.6	-4.2	-1.2	-2.4	-7.7	-6	-5.7	-4.8	-5.1	-4.6	-9.1	-14.1
Hard Scat.	2.4	4.9	6.6	3.5	3.4	0.6	3.6	6.1	-1.6	3.1	2.5	4.5	13.5	0.1	-9.5

Table H.2: Cross sections as a function of $|y(t)|$ for different $p_T(t)$ bins. The table also presents the summary of all the statistical and systematic uncertainties.

$p_T(t)$ [GeV]	0 to 65			65 to 130			130 to 200			200 to 400			400 to 2000		
$\lvert y(t\bar{t})\rvert$	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3
$\frac{d\sigma}{d\lvert y(t\bar{t})\rvert}$ [pb]	53.4	43.9	14.6	77.6	61.4	20	44.6	34.7	9.8	22.4	14.9	3.6	1.2	1	0.03
Source	Uncertainty [%]														
Statistical	3.2	3.5	4.8	2	2.2	2.8	2.8	3.2	4.9	3.3	4.4	6.8	13.2	14.2	146.1
Total Syst. Pos.	7.3	10	12.6	8.6	9	10.3	6.7	9.2	11.7	8.9	7.5	11.7	9	19.1	86.2
Total Syst. Neg.	8.7	10.1	13.4	9.6	9.1	11.6	8.3	9.5	10.9	10.4	8.1	12	11.8	19.2	116
Lumi Up	-2.6	-3	-2.6	-2.5	-2.8	-2.4	-2.7	-2.8	-2.9	-2.6	-2.6	-2.6	-2.8	-2.8	-3.9
Down	2.8	3.2	2.7	2.6	3	2.5	2.9	3	3	2.8	2.8	2.8	2.9	2.9	4.2
b-tagging Pos.	1.1	0.8	1	0.5	0.9	1.2	0.8	1.1	1.2	1.2	0.6	1.1	2.1	1.7	10.7
Neg.	0.9	1.5	1.7	0.8	0.9	1.1	1	0.9	1.4	0.8	1.1	1.5	1.6	0.9	5.3
Kin Up	-0.32	-0.37	-0.33	-0.32	-0.36	-0.3	-0.35	-0.35	-0.37	-0.35	-0.36	-0.35	-0.38	-0.37	-0.58
Down	0.33	0.37	0.33	0.32	0.36	0.31	0.35	0.35	0.37	0.36	0.36	0.36	0.38	0.37	0.59
JER Up	0.1	2	0.6	0.1	-0.4	-0.4	0.5	-0.7	2.9	0.4	0.9	-1.3	0.5	-4.5	5
Down	-1.5	-0.7	-0.9	-0.4	-0.8	-1.2	-0.3	-1.7	-0.2	-1.4	-0.6	-1.2	-1.6	-5.2	-10.5
JES Up	-2.3	-0.4	-2.1	-3.2	-3.5	-2.8	-2	-3.7	-3.5	-1.6	0.2	-0.6	-0.8	-2.9	-16.8
Down	1.4	1.7	1	3.6	3.9	3.6	3	2.9	5	0.9	0.9	2.3	-0.9	1.1	-37.4
PU Up	0.5	0.4	0.8	0.3	0.4	0.5	0.3	0.3	0.3	-0.2	-0.2	0.3	0.2	0.3	-7.3
Down	-0.5	-0.3	-0.5	-0.3	-0.4	-0.6	-0.3	-0.2	-0.3	0.1	0.2	-0.3	-0.2	-0.2	5.6
Trig Up	-1.1	-1.3	-1.7	-1.1	-1.3	-1.5	-1.2	-1.2	-1.8	-1.1	-1.2	-1.5	-1.1	-1.2	-1.7
Down	1.1	1.4	1.7	1.1	1.3	1.5	1.2	1.2	1.8	1.1	1.2	1.5	1.2	1.2	1.8
Lept Up	-2.1	-2.4	-2.1	-2	-2.3	-2	-2.2	-2.2	-2.3	-2.1	-2.1	-2.1	-2.2	-2.2	-3.2
Down	2.2	2.5	2.1	2.1	2.4	2	2.3	2.3	2.4	2.2	2.2	2.2	2.3	2.3	3.3
BG Up	3	-0.3	5.1	4.2	0.7	6.1	1.1	-0.1	-0.3	2.7	3.2	3.3	2.3	2.7	-10.5
Down	-5.5	-3	-7.6	-6.3	-3.5	-8.1	-3.7	-2.7	-2.3	-5.5	-6	-6.5	-6	-6.7	6.7
DY Up	-0.5	-1.2	-1.1	-0.1	-0.3	-0.4	-0.1	-0.1	-0.4	-0.1	-0.1	-0	0.1	0.1	0.6
Down	0.5	1.1	1.2	0.1	0.3	0.4	0.1	0.1	0.4	0.1	0.1	0	-0.1	-0.1	-0.5
Mass Up	-0.4	0.2	0.5	-1.1	-1.7	-1.5	-1.1	-1.4	-1.2	0.3	1.3	1.5	2.2	-0	-9.5
Down	0.4	-0.5	-0.4	2.1	2.4	1.3	1.2	2	1.6	-0.7	-1.1	-1.6	-2.4	1.4	-26.2
Match Up	1.4	-1.9	0.7	-0.2	-0.4	-0.7	-0.2	2.3	0.7	-1.3	2.9	1	0.6	8	-73.4
Down	0.7	1	4.8	-1.2	-1.5	-1	0.3	1.5	-1.5	-1.2	1.3	4.6	-6.8	-2.9	17.7
Scale Up	1.1	2.2	1.3	3.2	2.7	1.5	2.2	0.7	-0.8	-1.7	2.3	2.9	-2.6	-5.7	15.4
Down	1.1	-0.9	-2.7	-3.4	-1.6	-2.4	-4.9	0.9	1.6	-0.4	0.3	-2.5	5.2	6.4	35
PDF Pos.	0.3	0.5	0.4	0.1	0.2	0.3	0.2	0.2	0.3	0.3	0.2	0.3	0.6	0.4	1.3
Neg.	0.4	0.5	0.3	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.5	0.4	1.4
Hadronization	-4.4	-5.1	2.3	-3.5	-4.6	-1.5	-3.3	-5.2	-5.6	-6.9	-2.7	-6.5	-3.8	-10.3	43.9
Hard Scat.	2	6.3	9.2	2.1	3.6	5.9	1	4.7	6.8	2.7	2.7	5.5	2.9	11.2	-61

Table H.3: Cross sections as a function of $|y(t\bar{t})|$ for different $p_T(t)$ bins. The table also presents the summary of all the statistical and systematic uncertainties.

$p_T(t)$ [GeV]	0 to 65			65 to 130			130 to 200			200 to 400			400 to 2000		
$\Delta\eta(t\bar{t})$	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6
$\frac{d\sigma}{d\Delta\eta(t\bar{t})}$ [pb]	22.3	18.5	7	38.7	29.1	7.4	25.1	17.1	3	12.5	8.2	0.94	0.8	0.36	0.02
Source	Uncertainty [%]														
Statistical	5.9	4.1	3.7	2.9	2.7	2.8	3.5	3.7	5.1	3.9	4.4	7.9	12.8	19	59
Total Syst. Pos.	18.9	18.8	9.6	12.4	9.8	7.5	10	10.2	10.6	7.5	13.9	14.6	17.5	9.6	48.7
Total Syst. Neg.	18.5	20.1	10.2	12	11.7	9.9	10.1	9.8	12.7	9.8	13.1	17	16.6	34.1	36
Lumi Up	-2	-3.7	-2.7	-2.3	-2.9	-2.5	-2.8	-2.8	-2.7	-2.7	-2.6	-2.7	-2.8	-2.9	-2.6
Down	2.1	3.9	2.8	2.4	3	2.7	3	2.9	2.8	2.8	2.7	2.9	2.9	3	2.7
b-tagging Pos.	1.2	1.1	1	1	0.9	0.4	1	1.1	1	1	0.9	0.9	2	1.5	2.3
Neg.	1.2	1.9	2	0.7	0.9	0.8	1	0.7	1.5	0.9	1.1	1.4	1.8	0.6	3.6
Kin Up	-0.23	-0.48	-0.32	-0.29	-0.38	-0.32	-0.34	-0.35	-0.36	-0.36	-0.35	-0.37	-0.37	-0.39	-0.35
Down	0.23	0.48	0.32	0.29	0.39	0.32	0.35	0.36	0.36	0.36	0.35	0.37	0.38	0.39	0.36
JER Up	2.1	0.2	1.3	-0	-0.6	-0.2	0.9	0.7	-0.8	0.3	0.1	1	-3.6	0.5	-3.2
Down	-0.3	-2.4	-0.5	-1.2	-1.4	0.5	-0.6	-0.9	-0.8	-1	-1	-2.1	-5.3	-1.9	2.5
JES Up	4.6	-2.5	-4.9	-1.8	-4	-3.8	-3.8	-2.2	-2.3	-0.3	-1.1	-2.7	-3.1	-1.4	-2
Down	-2.1	2	2.7	2.8	4.3	4.5	3.8	2.8	3.2	0.8	1.9	0.1	-1.7	-3.5	9.3
PU Up	0.7	0.6	0.1	0.5	0.3	0.3	0.2	0.4	0.6	-0.2	0.3	-1.4	-0.4	0.5	4.3
Down	-0.6	-0.5	-0.1	-0.5	-0.4	-0.4	-0.1	-0.4	-0.7	0.1	-0.3	1.3	0.3	-0.3	-5.2
Trig Up	-0.9	-1.8	-1.2	-1.1	-1.4	-1.3	-1.3	-1.2	-1.4	-1.2	-1.1	-1.4	-1.1	-1.3	-1.3
Down	0.9	1.8	1.3	1.1	1.4	1.3	1.3	1.3	1.5	1.3	1.2	1.5	1.2	1.3	1.3
Lept Up	-1.4	-3.1	-2.1	-1.8	-2.5	-2	-2.2	-2.2	-2.2	-2.2	-2.1	-2.2	-2.2	-2.3	-2
Down	1.5	3.2	2.2	1.9	2.5	2.1	2.2	2.3	2.2	2.3	2.1	2.3	2.2	2.4	2.1
BG Up	11.3	-6.9	2.2	6.9	0	2.9	-0	0.4	2	2.7	3.5	2	2.5	1.1	4.4
Down	-13.3	2.3	-4.5	-8.7	-3.1	-4.9	-2.5	-3.2	-4.6	-5.7	-6.2	-4.8	-6.3	-5.1	-7.8
DY Up	-0.9	-1.2	-0.7	-0.3	-0.3	-0.1	-0.3	-0.1	-0.1	0	-0.2	-0.1	-0	0.3	0.2
Down	0.9	1.2	0.7	0.3	0.3	0.1	0.3	0.1	0.1	-0	0.2	0.1	0	-0.3	-0.2
Mass Up	2.5	0.1	-1.4	-1.3	-1.7	-1.4	-1.9	-0.9	-0.6	0.4	1	0.6	1.3	0.3	2.3
Down	-3.6	-0.5	2.3	2.3	2.5	1.6	2	1.9	0.3	-1.4	-0.4	-1.8	-0.5	-2.8	1.3
Match Up	4.5	-1.2	-1.3	1.3	-0.9	-2.3	-0	3.1	-0.2	-1.6	3.7	-6	6.8	-6.1	-3.4
Down	6.9	-0.4	2.6	0.2	-5.5	-0.2	-0.2	-0.6	2.1	-2	5	-2.9	6.7	-26.6	31.4
Scale Up	-6.5	3.2	6.3	5.4	1	0.7	2.3	0.3	3.6	-3	4.4	-2.8	7.2	-17.8	-0.4
Down	6.4	-0.8	-5.4	-0.2	-2.6	-5.3	-1.8	0.8	-7.3	-1.8	0.8	-2.3	6.1	1.4	9
PDF Pos.	1.5	13.5	0.3	0.1	0.8	0.2	0.3	0.4	0.6	0.1	0.1	0.9	0.3	0.8	1.4
Neg.	1.4	13.8	0.3	0.1	0.9	0.2	0.2	0.3	0.6	0.1	0.1	0.8	0.3	0.8	1.7
Hadronization	-1.4	-8.8	-1.7	-2.9	-4.2	-3.4	-5.2	-3.4	-7	-5.1	-8.2	5.6	-7.6	-4.7	-32.4
Hard Scat.	9.7	6.6	2.7	6.6	5.8	0.2	5.6	7.3	-3.9	2.6	7.1	-12.5	11	-6.8	10.2

Table H.4: Cross sections as a function of $\Delta\eta(t\bar{t})$ for different $p_T(t)$ bins. The table also presents the summary of all the statistical and systematic uncertainties.

I | Cross Sections for Different Bins of $M(t\bar{t})$

In the following the tables with the numbers for cross sections shown in Sec. 11.4 are presented, both, for absolute and normalized cross sections. The corresponding uncertainties are also listed in the tables.

$M(t\bar{t})$ [GeV]	350 to 390			390 to 480			480 to 600			600 to 1000			1000 to 2000		
$\lvert y(t)\rvert$	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5
$\frac{d\sigma}{d\lvert y(t)\rvert}$ $[pb]$	35.3	28.1	9.5	67.1	47	17.5	38.4	27.8	12.6	18.6	18.3	10.2	0.7	1.3	1.1
Source	Uncertainty [%]														
Statistical	3.6	4.9	6.9	1.8	2.7	3.3	2.5	3.7	3.8	3.5	4	3.8	18.1	14.6	14.7
Total Syst. Pos.	7.6	9.6	13.8	8.1	11.2	9.2	6.5	10.9	7.9	14.1	10.8	9.3	33.7	15	7
Total Syst. Neg.	8.6	8.9	11.6	9.2	10.3	10.9	8.1	10.5	8.6	15.1	10.8	10.5	30.6	17.7	20.6
Lumi Up	-2.7	-3.2	-2.7	-2.4	-2.9	-2.2	-2.8	-3	-2.8	-2.6	-2.5	-2.6	-2.9	-2.8	-2.9
Down	2.9	3.3	2.8	2.5	3.1	2.3	2.9	3.2	2.9	2.8	2.7	2.7	3.1	2.9	3
b-tagging Pos.	1.2	1.1	1	0.7	1.2	1.1	0.9	0.9	0.7	1.1	0.5	0.6	2.1	1.9	1.6
Neg.	1	1.9	2.2	0.8	0.6	1.2	0.9	0.7	0.8	1	1.4	1.1	1.7	1.2	1.2
Kin Up	-0.34	-0.39	-0.33	-0.3	-0.38	-0.28	-0.35	-0.39	-0.34	-0.35	-0.34	-0.34	-0.39	-0.38	-0.4
Down	0.34	0.39	0.33	0.3	0.38	0.28	0.35	0.4	0.35	0.36	0.34	0.34	0.39	0.38	0.4
JER Up	-0.1	0.9	0.5	0.4	0.6	-0.5	-0.1	-0.4	0.9	0.2	0.9	0.7	0.7	0.1	-6.6
Down	-1.6	-0.3	-2.1	-0.7	-0.7	-1.5	-1.5	-0.4	-0.9	-1.5	-0.5	-0.5	2.3	-3.1	-2.6
JES Up	-0.2	-1.5	0.1	-2.5	-4.6	-4.3	-2.8	-3.5	-3.1	-1.1	-1.7	-0.7	1.9	2.1	-11.1
Down	-0.8	2	-0.7	3.3	5.3	3.5	2.4	4.2	4.3	0.5	1.5	1.7	2.2	1.1	0.3
PU Up	0.5	0.7	0.9	0.2	0.3	0.4	0.1	0.1	0.4	0	0.3	0.4	0.8	-0.2	-1.3
Down	-0.4	-0.7	-0.8	-0.2	-0.3	-0.5	-0.1	-0.1	-0.5	-0.1	-0.4	-0.4	-0.8	0.4	1.6
Trig Up	-1.1	-1.4	-2	-1	-1.4	-1.4	-1.2	-1.3	-1.5	-1.2	-1.2	-1.4	-1.3	-1.3	-1.7
Down	1.1	1.4	2	1	1.4	1.5	1.3	1.4	1.6	1.2	1.2	1.4	1.3	1.3	1.8
Lept Up	-2.2	-2.5	-2.1	-1.9	-2.4	-1.8	-2.2	-2.5	-2.2	-2.1	-2	-2.1	-2.3	-2.2	-2.3
Down	2.2	2.6	2.1	1.9	2.5	1.9	2.3	2.6	2.3	2.2	2.1	2.1	2.4	2.3	2.4
BG Up	2.8	-1.4	5.1	5.2	-0.1	7.3	0.4	-2.7	0.8	3.3	3.8	3.4	1.1	2.1	1.4
Down	-5.6	-2.7	-7.9	-7.1	-2.9	-8.9	-2.8	-0.5	-3.1	-6.3	-6.5	-6	-5.2	-5.7	-4.6
DY Up	-0.6	-1.2	-1.4	-0.2	-0.5	-0.2	-0.1	-0.1	-0.4	-0.1	-0.1	-0.2	0.1	0.1	0.2
Down	0.6	1.2	1.6	0.2	0.5	0.2	0.1	0.1	0.4	0.1	0.1	0.2	-0.1	-0.1	-0.2
Mass Up	-4	-3.7	-4.2	-0.2	-0.8	-0.5	0.4	-0.4	-0.1	2.4	1.6	1	4.3	3.2	-0.6
Down	4.8	4.7	3.8	0.6	1.2	0.6	-0.8	0.2	0.2	-2	-1.5	-1.3	-3.1	-1.8	-1.7
Match Up	-0.4	0.1	1.4	-0.3	1.2	-0.1	-0.4	0.5	-1.5	0.3	4.9	-0.1	6.9	0.7	-12
Down	-0.3	1.5	10.2	0	0.6	-1.9	-0.1	-0.8	-2.8	-1.3	1.7	1.7	3.2	-5.4	-11.4
Scale Up	1.4	3	3	1.9	3.9	2.3	0.1	4.2	-0.1	-1.4	2.4	1.9	2.3	-12.7	-6.5
Down	0.3	-1	-3.3	-2.1	-0.8	-0.8	-3.3	-3.1	-2.3	1.2	-1.7	-2.6	12.4	11.1	-3
PDF Pos.	1.1	1.1	1.2	0.2	0.3	0.2	0.2	0.7	0.3	0.4	0.3	0.1	0.4	0.6	0.4
Neg.	1.3	1	1.2	0.2	0.2	0.2	0.2	0.7	0.4	0.3	0.2	0.1	0.4	0.5	0.5
Hadronization	-1.4	-1.4	2.5	-3.4	-6.2	0.8	-3.5	-5.9	-3.5	-4.8	-5.1	-6.4	6.3	-5	-0
Hard Scat.	2.2	5.3	3	0.4	4.5	0.5	2.7	5.2	3.6	11.9	5.3	3.1	29	6.1	-4.9

Table I.1: Cross sections as a function of $|y(t)|$ for different $M(t\bar{t})$ bins. The table also presents the summary of all the statistical and systematic uncertainties.

$M(t\bar{t})$ [GeV]	350 to 390			390 to 480			480 to 600			600 to 1000			1000 to 2000		
$\lvert y(t\bar{t})\rvert$	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3
$\frac{d\sigma}{d\lvert y(t\bar{t})\rvert}$ [pb]	36	30.8	13	72.5	58.3	20.5	48.6	37.9	9.9	35	24.3	4.3	3.3	1.6	0.08
Source	Uncertainty [%]														
Statistical	4.3	4.6	4.8	2.1	2.2	2.6	2.6	2.9	4.7	2.8	3.6	7.8	9.7	17.8	75.8
Total Syst. Pos.	7.6	10.1	9.7	8.9	8.7	8.9	7.1	8.2	12.8	9.1	10.5	19.4	13	35.3	84.4
Total Syst. Neg.	8.7	9.7	10.3	9.5	8.6	10.3	9	8.4	12.9	10.3	11.4	19.5	12.9	35.5	92.7
Lumi Up	-2.7	-3	-2.8	-2.4	-2.8	-2.3	-2.8	-3	-2.9	-2.6	-2.6	-2.6	-2.7	-3	-2.8
Down	2.9	3.2	3	2.5	2.9	2.4	2.9	3.1	3	2.7	2.7	2.7	2.9	3.2	3
b-tagging Pos.	1.2	0.9	1	0.8	1	1.1	0.8	0.8	0.9	0.7	0.8	1.1	1.4	2.2	4.4
Neg.	1.1	1.7	1.6	0.7	0.7	1.1	0.8	0.8	0.9	0.8	1.2	1.9	1.6	0.6	2.9
Kin Up	-0.33	-0.38	-0.34	-0.3	-0.35	-0.3	-0.36	-0.37	-0.37	-0.35	-0.34	-0.35	-0.38	-0.4	-0.41
Down	0.34	0.38	0.35	0.31	0.35	0.3	0.36	0.37	0.37	0.35	0.35	0.35	0.38	0.41	0.42
JER Up	0.1	0.8	-0.4	0.1	0.5	0.6	0.4	-0.3	0.7	0.5	-0.1	1	2.7	-3.5	-18.7
Down	-2	-0.1	-1.8	-0.7	-0.7	-0.9	-0.7	-1.4	-0.8	-0.4	-1.8	-0.8	-0.2	-2.2	2.4
JES Up	-0.3	-0	-2.8	-3	-3.9	-3.5	-3.6	-2.6	-2.3	-1.8	-0.8	-1.2	0.9	-0.8	-12.3
Down	-1.3	1.6	0.7	4.1	3.9	4.3	2.8	3.3	4.2	1.8	0.8	0.6	1.6	1	-15.8
PU Up	0.4	0.7	0.9	0.3	0.3	0.2	0.1	0.3	0.2	0.2	0.2	0.7	0.1	-1.4	-2.5
Down	-0.3	-0.8	-0.8	-0.3	-0.2	-0.1	-0.2	-0.3	-0.3	-0.2	-0.2	-0.7	-0.2	2.1	1.3
Trig Up	-1.1	-1.2	-1.7	-1	-1.2	-1.5	-1.2	-1.3	-1.7	-1.2	-1.2	-1.5	-1.4	-1.5	-1.3
Down	1.1	1.3	1.8	1	1.3	1.5	1.2	1.4	1.8	1.2	1.2	1.5	1.4	1.5	1.3
Lept Up	-2.1	-2.4	-2.2	-2	-2.2	-1.9	-2.3	-2.3	-2.3	-2.1	-2.1	-2.1	-2.2	-2.4	-2.3
Down	2.2	2.5	2.3	2	2.3	2	2.3	2.4	2.4	2.1	2.1	2.2	2.3	2.5	2.4
BG Up	2.6	0.3	2.8	4.8	1.2	6.6	0.3	-1.8	-0.2	3.1	3.9	4.3	2.7	-0.6	4.7
Down	-5.3	-4.2	-5.8	-6.9	-3.8	-8.4	-2.9	-1.1	-2.4	-5.7	-6.8	-7.6	-6.4	-2.8	-10
DY Up	-0.7	-1.1	-1.2	-0.2	-0.4	-0.3	-0	-0.2	-0.4	-0.2	-0	-0.1	0.1	0.2	0.4
Down	0.6	1.1	1.3	0.2	0.5	0.4	0	0.2	0.3	0.2	0	0.1	-0.1	-0.2	-0.5
Mass Up	-4.4	-3.6	-3.9	-0.3	-0.9	-0.2	-0.2	-0.4	1	1.2	2.1	2.3	2.4	0.5	6.5
Down	4.8	5.1	4	1.1	1.1	0	0	-0.2	-0.8	-1.3	-1.3	-3.1	-2.3	-0.9	-6.3
Match Up	-0.1	-0.2	0.6	0.1	0.3	0.2	-0	-0	-0.9	1.1	3	1.9	-2.3	10.7	-30.7
Down	2.2	-1	4.6	0.2	-0.2	-1.9	-1.3	0.1	-3	-0.6	-0.5	5.2	-1.6	-11.1	-17.8
Scale Up	-0.4	3.1	4	2.6	3.7	1.3	2.6	-0.5	0.2	0.8	1.6	-2.3	-7.4	-10.1	26.1
Down	0.6	-0.6	-3.4	-2	-0.2	-1.8	-4.9	-1.6	-0.9	-1.3	-0	4.6	9.3	10.6	-23.3
PDF Pos.	0.3	1	0.2	0.1	0.2	0.3	0.2	0.2	0.6	0.2	0.2	0.6	0.5	0.5	5
Neg.	0.3	0.9	0.2	0.1	0.2	0.3	0.2	0.2	0.5	0.2	0.2	0.6	0.5	0.5	4.2
Hadronization	-1.8	-1.9	2.2	-4.1	-4.6	-0.1	-3.7	-4.7	-4.7	-5	-4.5	-4.8	-4.8	3.6	21.7
Hard Scat.	1.6	6.1	2.7	1.1	2.6	1.4	2.1	4	10.2	5.2	6.6	16.3	4.3	31.3	76.4

Table I.2: Cross sections as a function of $|y(t\bar{t})|$ for different $M(t\bar{t})$ bins. The table also presents the summary of all the statistical and systematic uncertainties.

$M(t\bar{t})$ [GeV]	350 to 390			390 to 480			480 to 600			600 to 1000			1000 to 2000		
$\Delta\eta(t\bar{t})$	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6
$\frac{d\sigma}{d\Delta\eta(t\bar{t})}$ [pb]	36.4	14.5	0.9	43.4	31.6	4.6	15.5	18.1	6	5.9	8.2	5.9	0.24	0.41	0.55
Source	Uncertainty [%]														
Statistical	4.1	5	30.2	2.6	2.5	4.7	4.6	3.6	3.1	5.7	4.3	2.9	26.5	17.2	11.1
Total Syst. Pos.	8.4	14.4	60.9	9.8	8.2	10.5	10.9	7.4	7.4	11.7	13	7.6	30.2	9.5	9.2
Total Syst. Neg.	11.3	15.6	31.8	9.8	9.3	11.2	11.8	7.5	11.6	14.8	13.1	9.6	13.9	19.2	19
Lumi Up	-2.4	-3.9	-1.5	-2.4	-2.8	-2.5	-2.9	-2.9	-2.8	-2.6	-2.6	-2.6	-3	-2.9	-2.7
Down	2.5	4.1	1.6	2.5	2.9	2.7	3.1	3	2.9	2.7	2.7	2.8	3.2	3	2.8
b-tagging Pos.	1	1.1	4.1	0.9	0.9	1	0.9	1.1	0.8	1.1	1	0.6	1.9	2.2	1.3
Neg.	1	2	4.2	0.6	0.7	1.6	1	0.9	1.2	0.9	1.2	1.4	2.2	0.9	1.6
Kin Up	-0.3	-0.47	-0.17	-0.31	-0.34	-0.31	-0.36	-0.37	-0.35	-0.35	-0.36	-0.34	-0.41	-0.38	-0.39
Down	0.3	0.48	0.17	0.32	0.35	0.31	0.36	0.38	0.36	0.35	0.36	0.34	0.41	0.38	0.39
JER Up	-0.1	-1.1	7	0.6	0.5	1.1	0.6	-0	-0.9	-0	0.3	0.7	-4.1	0.2	-4.6
Down	-1.2	-2.1	-2.3	-0.8	-0.7	0.7	-0.5	-1.5	-1.3	-1.4	-1.7	0	-3.6	2.4	-6.9
JES Up	-0.1	-1.4	5.5	-2.7	-3.5	-6.1	-2.4	-1.7	-4.9	0.5	-2	-1.6	-1.4	4	-10.1
Down	0.5	0.1	-12.2	3.9	4.7	5.5	1.5	3.4	3.5	-0.1	1	3	-1	-2.3	0.7
PU Up	0.6	0.8	0.9	0.4	0.3	-0	-0.1	0.4	0.2	-0	0.1	0.3	0.6	0.3	0.7
Down	-0.6	-0.7	0	-0.3	-0.3	-0.1	0.1	-0.4	-0.3	-0.1	-0.2	-0.3	-0.8	0	-0.5
Trig Up	-1.2	-1.7	-0.1	-1.2	-1.3	-1.1	-1.3	-1.3	-1.4	-1.1	-1.2	-1.4	-1.3	-1.2	-1.6
Down	1.2	1.8	0.1	1.2	1.3	1.1	1.3	1.4	1.4	1.2	1.2	1.4	1.3	1.2	1.7
Lept Up	-1.9	-3	-1.3	-2	-2.2	-2.1	-2.2	-2.3	-2.2	-2.1	-2.1	-2.1	-2.4	-2.2	-2.2
Down	2	3.1	1.3	2.1	2.3	2.1	2.3	2.4	2.3	2.1	2.2	2.1	2.5	2.3	2.3
BG Up	6.5	-8.6	18.3	5.4	1.1	3.5	-1	-0.5	0.6	4.3	3.8	2.4	-0.6	1.1	3.4
Down	-9	3.7	-20.1	-7.3	-3.8	-5.8	-1.7	-2.4	-2.9	-7.4	-6.6	-4.8	-3.5	-5	-6.5
DY Up	-0.5	-1.4	-2.9	-0.2	-0.3	-0.3	-0.1	-0.2	-0.2	0	0	-0.2	0	0	0.2
Down	0.5	1.4	3.1	0.2	0.3	0.4	0.1	0.2	0.2	-0	-0	0.3	-0	-0	-0.2
Mass Up	-3.5	-4	-8.3	1.1	-1	-2.4	0.7	0.7	-1.1	2.8	2.7	0.4	2.1	0.9	2.3
Down	2.6	6.5	11.4	-0.3	1.1	3.4	-2.2	-0.2	1.1	-2.7	-1.7	-0.9	-2.2	-4.2	-1.3
Match Up	-0.9	-0.6	16.3	1.5	0.4	-2	1	0.6	-3.3	-2.3	4.6	0	9.8	-2.7	-6.6
Down	0.9	-1.6	33.3	1.4	-4.3	0.2	0.2	-2.2	-1.4	-5.4	3.9	2	11.8	-9	-9.8
Scale Up	-3.5	3	41.1	3.4	1.9	4.4	1.9	1.6	0.4	-4.3	2.4	2.8	26.5	-13.4	-6.8
Down	-0	-1.2	-4	0.6	-1.8	-3.1	-3.9	-1.5	-6.9	1.5	1.2	-5.9	-9.4	0.8	6.4
PDF Pos.	1.9	1.3	-1.2	0.3	0.2	0.7	0.2	0.4	0.5	0.2	0.2	0.6	0.5	0.4	0.5
Neg.	1.8	1.3	-1.4	0.2	0.2	0.7	0.2	0.3	0.4	0.2	0.2	0.5	0.5	0.4	0.5
Hadronization	0.5	-5.4	11.1	-3.4	-3.7	-3.9	-6.6	-3.3	-4.9	-9	-7	-3.8	4.4	-6.5	-1.9
Hard Scat.	1.8	8.9	14.7	3.3	2.9	2	7	3.3	0.7	3.6	7.4	1	-5	1.5	-2.2

Table I.3: Cross sections as a function of $\Delta\eta(t\bar{t})$ for different $M(t\bar{t})$ bins. The table also presents the summary of all the statistical and systematic uncertainties.

$M(t\bar{t})$ [GeV]	350 to 390			390 to 480			480 to 600			600 to 1000			1000 to 2000		
$\|y(t)\|$	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5	0 to 0.6	0.6 to 1.2	1.2 to 2.5
$\frac{1}{\sigma}\ \frac{d\sigma}{d\|y(t)\|}$	0.15	0.119	0.04	0.285	0.2	0.074	0.163	0.118	0.053	0.079	0.078	0.043	0.003	0.0055	0.0046
Source	Uncertainty [%]														
Statistical	3.5	4.8	6.7	1.8	2.6	3.2	2.5	3.7	3.7	3.5	4	3.8	18.1	14.7	14.7
Total Syst. Pos.	5.4	5.7	12.6	4.1	5.3	7	2.8	6.6	2.9	9.7	6.2	4.5	32.1	14.8	8.9
Total Syst. Neg.	5.5	5.9	8.8	3.7	4.7	7.2	3.4	6.6	4.1	9.9	4.2	4.4	27.1	16.2	20.6
Lumi Up	-0	-0.5	0	0.3	-0.2	0.5	-0.1	-0.4	-0.1	0.1	0.1	0.1	-0.2	-0.1	-0.2
Down	0	0.5	-0	-0.3	0.2	-0.5	0.1	0.4	0.1	-0.1	-0.1	-0.1	0.2	0.1	0.2
b-tagging Pos.	1.1	0.7	0.8	0.6	0.6	0.8	0.6	0.6	0.3	0.8	0.2	0.3	2	1.5	1
Neg.	0.8	1.4	1.8	0.5	0.1	0.8	0.4	0.3	0.4	0.5	0.9	0.7	1.5	1	0.9
Kin Up	0	-0.05	0.01	0.04	-0.04	0.06	-0.01	-0.05	-0	-0.01	0	-0	-0.05	-0.04	-0.06
Down	0	0.05	-0.01	-0.04	0.04	-0.06	0.01	0.05	0	0.01	0	0	0.05	0.04	0.06
JER Up	-0.4	0.7	0.2	0.1	0.3	-0.7	-0.4	-0.6	0.7	-0.1	0.7	0.5	0.5	-0.2	-6.8
Down	-0.6	0.7	-1.1	0.3	0.3	-0.5	-0.5	0.6	0.1	-0.5	0.6	0.5	3.3	-2.2	-1.7
JES Up	2.3	1	2.7	0	-2.2	-1.9	-0.4	-1.1	-0.6	1.4	0.8	1.8	4.5	4.6	-8.9
Down	-3.3	-0.6	-3.2	0.7	2.6	0.9	-0.2	1.5	1.6	-2.1	-1.1	-0.9	-0.3	-1.5	-2.2
PU Up	0.2	0.4	0.5	-0.1	0	0.1	-0.2	-0.2	0.1	-0.3	-0	0.1	0.5	-0.5	-1.6
Down	-0.1	-0.4	-0.4	0.1	0	-0.1	0.2	0.2	-0.2	0.2	-0	-0	-0.5	0.7	1.9
Trig Up	0.2	-0.1	-0.7	0.3	-0.1	-0.1	0.1	-0.1	-0.2	0.1	0.1	-0.1	0	0	-0.4
Down	-0.2	0.1	0.7	-0.3	0.1	0.1	-0.1	0.1	0.2	-0.1	-0.1	0.1	-0	-0	0.4
Lept Up	-0	-0.3	0.1	0.3	-0.3	0.3	-0	-0.3	-0	0	0.1	0.1	-0.2	-0.1	-0.2
Down	0	0.3	-0.1	-0.3	0.3	-0.3	0	0.3	0	-0	-0.1	-0.1	0.2	0.1	0.2
BG Up	0.3	-3.8	2.6	2.7	-2.5	4.7	-2	-5	-1.6	0.9	1.4	0.9	-1.3	-0.3	-1
Down	-0.6	2.5	-3	-2.1	2.2	-4.1	2.3	4.7	2	-1.3	-1.5	-1	-0.1	-0.7	0.4
DY Up	-0.2	-0.9	-1	0.1	-0.1	0.2	0.3	0.3	0	0.3	0.3	0.2	0.5	0.5	0.6
Down	0.2	0.8	1.2	-0.1	0.1	-0.2	-0.3	-0.3	-0	-0.3	-0.3	-0.2	-0.5	-0.5	-0.6
Mass Up	-3.4	-3.1	-3.6	0.6	-0.1	0.3	1.2	0.4	0.7	3.3	2.4	1.8	5.3	4.1	0.2
Down	3.7	3.5	2.7	-0.3	0.2	-0.4	-1.7	-0.7	-0.7	-2.8	-2.3	-2.2	-3.9	-2.6	-2.5
Match Up	-0.5	-0.1	1.2	-0.5	1	-0.3	-0.6	0.3	-1.7	0.1	4.7	-0.3	6.7	0.5	-12.2
Down	-0.6	1.2	9.9	-0.2	0.3	-2.2	-0.3	-1.1	-3.1	-1.5	1.4	1.4	2.9	-5.7	-11.6
Scale Up	-0.5	1.1	1.1	0	2	0.4	-1.7	2.3	-1.9	-3.2	0.5	-0	0.4	-14.3	-8.3
Down	1.9	0.6	-1.8	-0.5	0.8	0.8	-1.8	-1.6	-0.7	2.9	-0.1	-1	14.2	12.9	-1.4
PDF Pos.	1	0.9	1.3	0.2	0.1	0.3	0.1	0.6	0.3	0.3	0.4	0.1	0.5	0.5	0.5
Neg.	1.2	0.9	1.4	0.2	0.1	0.3	0.1	0.5	0.4	0.2	0.3	0.1	0.4	0.5	0.6
Hadronization	1.9	1.8	5.9	-0.2	-3.1	4.2	-0.3	-2.8	-0.4	-1.7	-2	-3.3	9.9	-1.9	3.3
Hard Scat.	-1	1.9	-0.3	-2.8	1.2	-2.7	-0.5	1.8	0.3	8.3	1.9	-0.2	24.9	2.7	-7.9

Table I.4: Normalized cross sections as a function of $|y(t)|$ for different $M(t\bar{t})$ bins. The table also presents the summary of all the statistical and systematic uncertainties.

$M(t\bar{t})$ [GeV]	350 to 390			390 to 480			480 to 600			600 to 1000			1000 to 2000		
$\lvert y(t\bar{t})\rvert$	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3	0 to 0.45	0.45 to 1	1 to 2.3
$\frac{1}{\sigma}\frac{d\sigma}{d\lvert y(t\bar{t})\rvert}$	0.154	0.132	0.055	0.31	0.249	0.088	0.208	0.162	0.042	0.149	0.104	0.018	0.014	0.007	0.0004
Source	Uncertainty [%]														
Statistical	4.3	4.5	4.7	2.1	2.2	2.5	2.6	2.9	4.7	2.8	3.6	7.8	9.7	17.8	75.8
Total Syst. Pos.	5.9	5.7	8.2	4.2	3.3	6	3.2	4.6	7.4	3.5	5.7	15.1	12.2	31.9	79
Total Syst. Neg.	6.5	4.9	7.4	3.7	2.7	5.8	4.9	5.1	7.8	3.6	4.7	13.8	10.3	32.2	88.3
Lumi Up	-0.1	-0.3	-0.1	0.3	-0.1	0.4	-0.1	-0.3	-0.2	0.1	0.1	0.1	-0.1	-0.4	-0.2
Down	0.1	0.3	0.1	-0.3	0.1	-0.4	0.1	0.3	0.2	-0.1	-0.1	-0.1	0.1	0.4	0.2
b-tagging Pos.	1.1	0.5	0.7	0.7	0.5	0.8	0.4	0.4	0.8	0.3	0.3	1	0.9	2	4.9
Neg.	0.9	1.2	1.2	0.5	0.2	0.7	0.3	0.2	0.6	0.4	0.6	1.4	1.2	0.4	2.8
Kin Up	0.01	-0.04	-0	0.04	-0.01	0.04	-0.02	-0.03	-0.03	-0	-0	-0.01	-0.04	-0.06	-0.07
Down	-0	0.03	0	-0.04	0.01	-0.04	0.02	0.03	0.03	0	0	0.01	0.04	0.06	0.07
JER Up	-0.2	0.5	-0.7	-0.2	0.2	0.4	0.1	-0.6	0.4	0.2	-0.4	0.8	2.4	-3.8	-18.9
Down	-1.1	0.9	-0.9	0.3	0.3	0	0.3	-0.5	0.2	0.6	-0.8	0.2	0.7	-1.3	3.4
JES Up	2.2	2.5	-0.4	-0.5	-1.4	-1.1	-1.1	-0.1	0.2	0.6	1.7	1.3	3.4	1.7	-10
Down	-3.9	-1.1	-1.9	1.4	1.3	1.6	0.2	0.7	1.5	-0.9	-1.8	-2	-1	-1.6	-18
PU Up	0.1	0.4	0.6	-0	-0	-0.2	-0.2	-0	-0.2	-0.2	-0.1	0.4	-0.2	-1.7	-2.8
Down	0	-0.5	-0.5	0	0.1	0.2	0.1	0	0.1	0.2	0.1	-0.4	0.1	2.5	1.6
Trig Up	0.2	0.1	-0.4	0.3	0.1	-0.2	0.1	-0	-0.4	0.1	0.1	-0.2	-0.1	-0.2	-0
Down	-0.2	-0.1	0.5	-0.3	-0.1	0.2	-0.1	0	0.4	-0.1	-0.1	0.2	0.1	0.2	0
Lept Up	0	-0.2	-0	0.2	-0.1	0.2	-0.1	-0.2	-0.2	0.1	0.1	0.1	-0.1	-0.3	-0.2
Down	-0	0.2	0	-0.2	0.1	-0.2	0.1	0.2	0.2	-0.1	-0.1	-0.1	0.1	0.3	0.2
BG Up	0.2	-2.1	0.4	2.4	-1.2	4.1	-2	-4.1	-2.5	0.7	1.5	1.8	0.3	-3	2.2
Down	-0.3	0.9	-0.9	-2	1.3	-3.6	2.3	4.1	2.8	-0.7	-1.8	-2.7	-1.4	2.4	-5.2
DY Up	-0.3	-0.7	-0.9	0.2	-0.1	0.1	0.4	0.2	0	0.2	0.4	0.3	0.5	0.6	0.8
Down	0.2	0.7	0.9	-0.2	0.1	-0	-0.4	-0.2	-0.1	-0.2	-0.4	-0.3	-0.5	-0.5	-0.9
Mass Up	-3.8	-3	-3.3	0.5	-0.1	0.6	0.6	0.4	1.9	2	3	3.2	3.3	1.3	7.6
Down	3.6	3.9	2.8	0.1	0.1	-0.9	-0.9	-1.1	-1.7	-2.2	-2.1	-3.9	-3.1	-1.8	-6.9
Match Up	-0.4	-0.5	0.3	-0.2	-0	-0.2	-0.4	-0.4	-1.2	0.7	2.6	1.5	-2.6	10.3	-30.9
Down	2.3	-0.9	4.7	0.3	-0.1	-1.8	-1.2	0.2	-3	-0.5	-0.4	5.3	-1.5	-11	-17.8
Scale Up	-2.1	1.3	2.2	0.8	1.9	-0.4	0.9	-2.2	-1.5	-0.9	-0.1	-3.9	-9	-11.7	23.9
Down	1.9	0.7	-2.2	-0.7	1.1	-0.5	-3.6	-0.3	0.4	-0	1.3	6	10.7	12	-22.3
PDF Pos.	0.3	0.8	0.3	0.1	0.2	0.2	0.1	0.2	0.4	0.2	0.3	0.6	0.6	0.5	4.9
Neg.	0.4	0.7	0.3	0.1	0.1	0.2	0.1	0.2	0.3	0.2	0.3	0.6	0.6	0.5	4.1
Hadronization	1.3	1.2	5.4	-1.1	-1.5	3.1	-0.7	-1.7	-1.7	-2	-1.5	-1.8	-1.8	6.9	25.6
Hard Scat.	-2.1	2.2	-1.1	-2.6	-1.2	-2.3	-1.6	0.2	6.1	1.3	2.7	12	0.5	26.4	69.9

Table I.5: Normalized cross sections as a function of $|y(t\bar{t})|$ for different $M(t\bar{t})$ bins. The table also presents the summary of all the statistical and systematic uncertainties.

$M(t\bar{t})$ [GeV]	350 to 390			390 to 480			480 to 600			600 to 1000			1000 to 2000		
$\Delta\eta(t\bar{t})$	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6	0 to 0.8	0.8 to 2	2 to 6
$\frac{1}{\sigma}\ \frac{d\sigma}{d\Delta\eta(t\bar{t})}$	0.151	0.06	0.004	0.181	0.132	0.0191	0.064	0.075	0.0249	0.025	0.034	0.0247	0.001	0.0017	0.0023
Source	Uncertainty [%]														
Statistical	3.9	4.9	30.3	2.6	2.4	4.7	4.6	3.6	3.1	5.7	4.3	2.8	26.5	17.2	11.1
Total Syst. Pos.	7.5	12	58.6	5.1	2.4	5.1	6.1	3.3	3.7	8.7	8.8	3.7	29.5	9.2	10.9
Total Syst. Neg.	9.4	13.1	30.3	3	4.4	5.2	6.9	3.6	7.4	11.2	6.6	5	15.6	19.2	17.5
Lumi Up	0.3	-1.2	1.2	0.3	-0.1	0.2	-0.2	-0.2	-0.1	0.1	0.1	0.1	-0.3	-0.2	0
Down	-0.3	1.2	-1.2	-0.3	0.1	-0.2	0.2	0.2	0.1	-0.1	-0.1	-0.1	0.3	0.2	-0
b-tagging Pos.	0.5	0.7	4.6	0.6	0.4	0.9	0.8	1.1	0.5	0.8	0.8	0.4	1.4	1.9	0.9
Neg.	0.5	1.5	4.2	0.2	0.2	1.3	0.6	0.4	0.8	0.4	0.9	1.1	1.7	0.9	1.3
Kin Up	0.04	-0.13	0.18	0.03	-0	0.03	-0.01	-0.03	-0.01	-0.01	-0.01	0.01	-0.07	-0.04	-0.04
Down	-0.04	0.13	-0.18	-0.03	0	-0.03	0.01	0.03	0.01	0.01	0.01	-0.01	0.07	0.04	0.04
JER Up	-0.4	-1.3	6.7	0.4	0.3	0.8	0.4	-0.3	-1.1	-0.2	0.1	0.5	-4.3	-0	-4.8
Down	-0.2	-1.2	-1.4	0.1	0.2	1.6	0.5	-0.5	-0.3	-0.4	-0.8	1	-2.7	3.4	-6.1
JES Up	2.5	1.1	8.2	-0.1	-1	-3.6	0.1	0.8	-2.4	3.1	0.5	0.9	1.1	6.7	-7.8
Down	-2.1	-2.5	-14.5	1.2	2	2.8	-1.2	0.7	0.7	-2.7	-1.7	0.3	-3.6	-4.8	-1.9
PU Up	0.2	0.5	0.6	0	-0.1	-0.4	-0.5	0.1	-0.1	-0.3	-0.2	0	0.3	-0.1	0.4
Down	-0.3	-0.4	0.4	0	0.1	0.2	0.4	-0.1	0	0.3	0.2	-0	-0.5	0.4	-0.2
Trig Up	0.1	-0.5	1.2	0.1	0	0.2	0	-0	-0.1	0.1	0.1	-0.1	0	0.1	-0.3
Down	-0.1	0.5	-1.2	-0.1	-0	-0.2	-0	0	0.1	-0.1	-0.1	0.1	-0	-0.1	0.4
Lept Up	0.3	-0.8	0.9	0.2	-0	0.1	-0.1	-0.2	-0.1	0.1	0.1	0.1	-0.3	-0.1	-0.1
Down	-0.3	0.8	-0.9	-0.2	0	-0.1	0.1	0.2	0.1	-0.1	-0.1	-0.1	0.3	0.1	0.1
BG Up	4.3	-10.5	15.8	3.2	-1	1.4	-3.1	-2.6	-1.5	2.1	1.6	0.3	-2.7	-1	1.2
Down	-4.5	8.9	-16.1	-2.6	1	-1.1	3.2	2.5	2	-2.7	-2	0	1.3	-0.3	-1.8
DY Up	-0.1	-1	-2.6	0.2	0.1	0	0.3	0.2	0.2	0.4	0.4	0.1	0.4	0.4	0.6
Down	0.1	1	2.7	-0.2	-0.1	0	-0.3	-0.2	-0.2	-0.4	-0.4	-0.1	-0.4	-0.4	-0.6
Mass Up	-2.8	-3.4	-7.9	2	-0.2	-1.7	1.6	1.6	-0.3	3.8	3.7	1.2	3.1	1.8	3.2
Down	1.4	5.1	9.7	-1.3	-0	2.2	-3.1	-1.2	0.1	-3.5	-2.6	-1.9	-3.1	-5	-2.3
Match Up	-1	-0.6	16.2	1.4	0.3	-2.1	0.9	0.6	-3.4	-2.3	4.6	-0	9.7	-2.8	-6.7
Down	1.1	-1.5	33.5	1.6	-4.1	0.4	0.4	-2	-1.3	-5.2	4.1	2.1	12	-8.8	-9.6
Scale Up	-5.4	0.9	38.3	1.4	-0.1	2.4	-0.1	-0.4	-1.6	-6.2	0.4	0.7	24	-15.1	-8.7
Down	2.1	0.9	-2	2.7	0.3	-1.1	-1.9	0.5	-4.9	3.6	3.3	-3.9	-7.5	2.9	8.6
PDF Pos.	2.1	1	-1.4	0.1	0.1	0.5	0.1	0.1	0.2	0.4	0.3	0.3	0.7	0.6	0.6
Neg.	2.1	1.1	-1.6	0.1	0.1	0.5	0.1	0.1	0.2	0.4	0.2	0.3	0.8	0.7	0.5
Hadronization	4.2	-1.9	15.2	0.1	-0.2	-0.4	-3.2	0.3	-1.4	-5.6	-3.5	-0.2	8.2	-3	1.7
Hard Scat.	-1.5	5.4	11	-0	-0.4	-1.3	3.5	-0	-2.6	0.3	3.9	-2.2	-8	-1.8	-5.3

Table I.6: Normalized cross sections as a function of $\Delta\eta(t\bar{t})$ for different $M(t\bar{t})$ bins. The table also presents the summary of all the statistical and systematic uncertainties.

J | Correlation Matrices

The correlation matrices for all the unfolded distributions for which the cross sections are measured are presented in Fig. J.5-J.8.

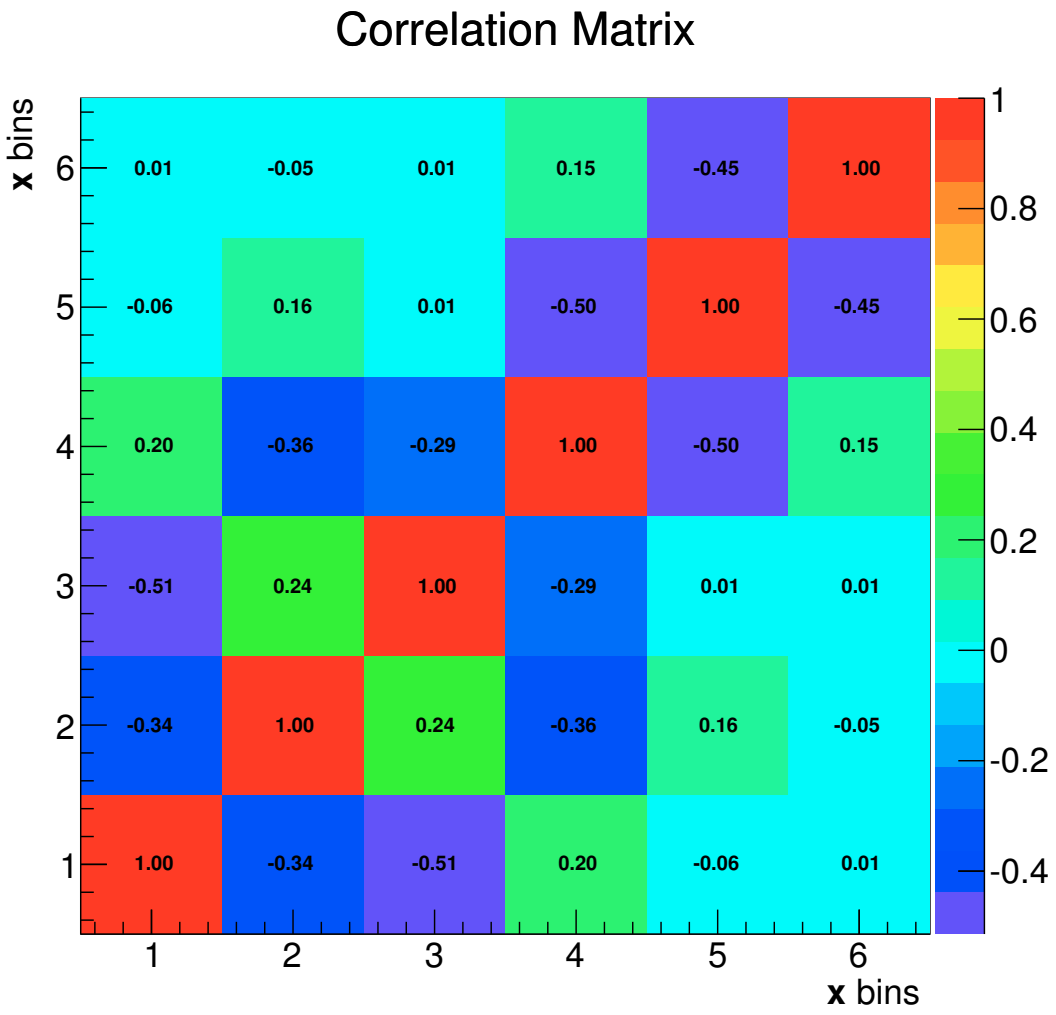

Figure J.1: Correlation matrix $\mathbf{V}_{xx}$ for the unfolded $\mathbf{x}$ distribution which corresponds to the $p_T(t)$. The binning is the following: [0, 65, 125, 200, 290, 400, 2000] GeV.

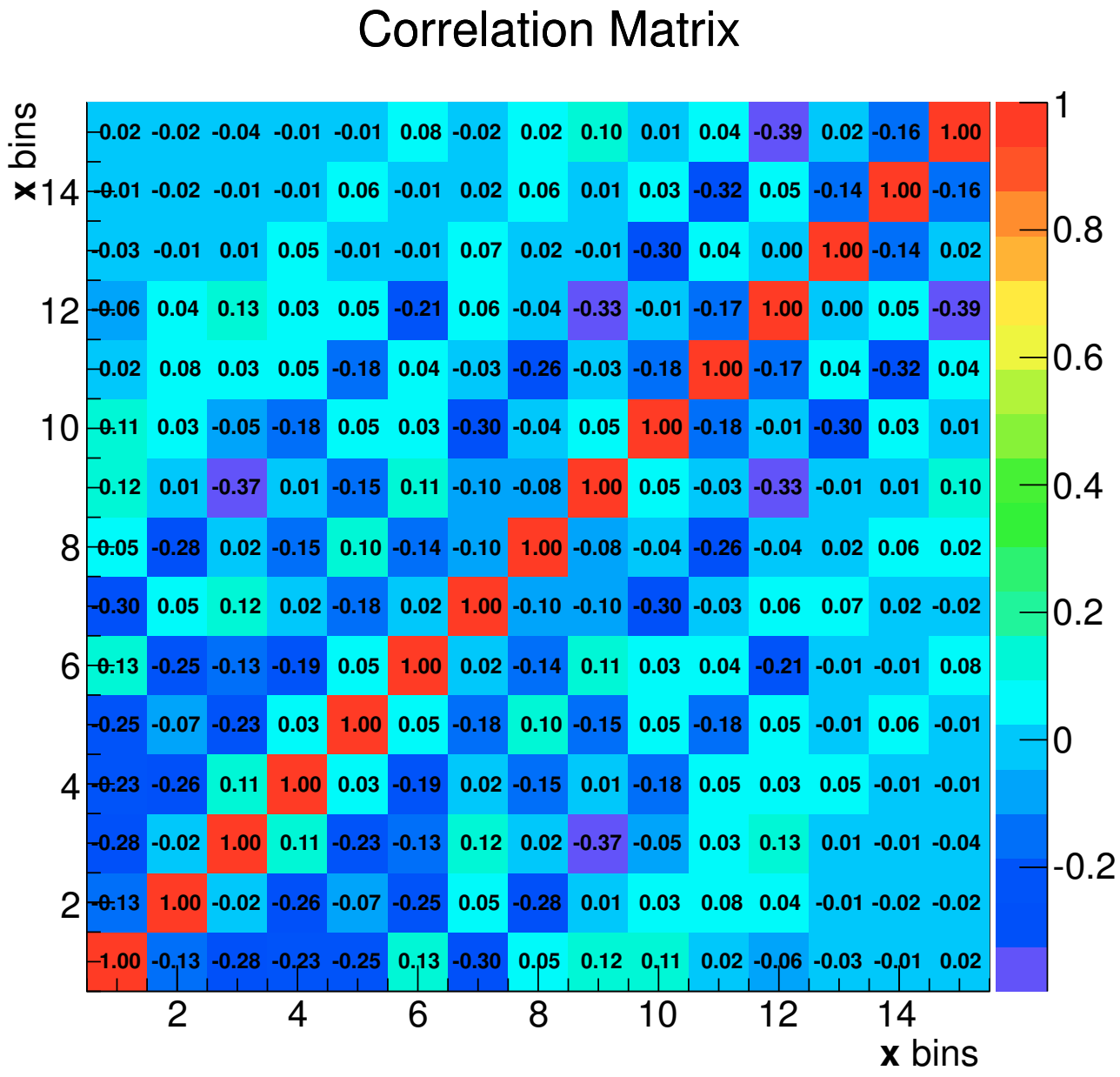

Figure J.2: Correlation matrix $\mathbf{V}_{xx}$ for the unfolded $\mathbf{x}$ distribution which corresponds to the $|y(t)|$ in bins of $p_T(t)$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $p_T(t)$ – [0, 65, 130, 200, 400, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t)|$ – [0, 0.6, 1.2, 2.5].

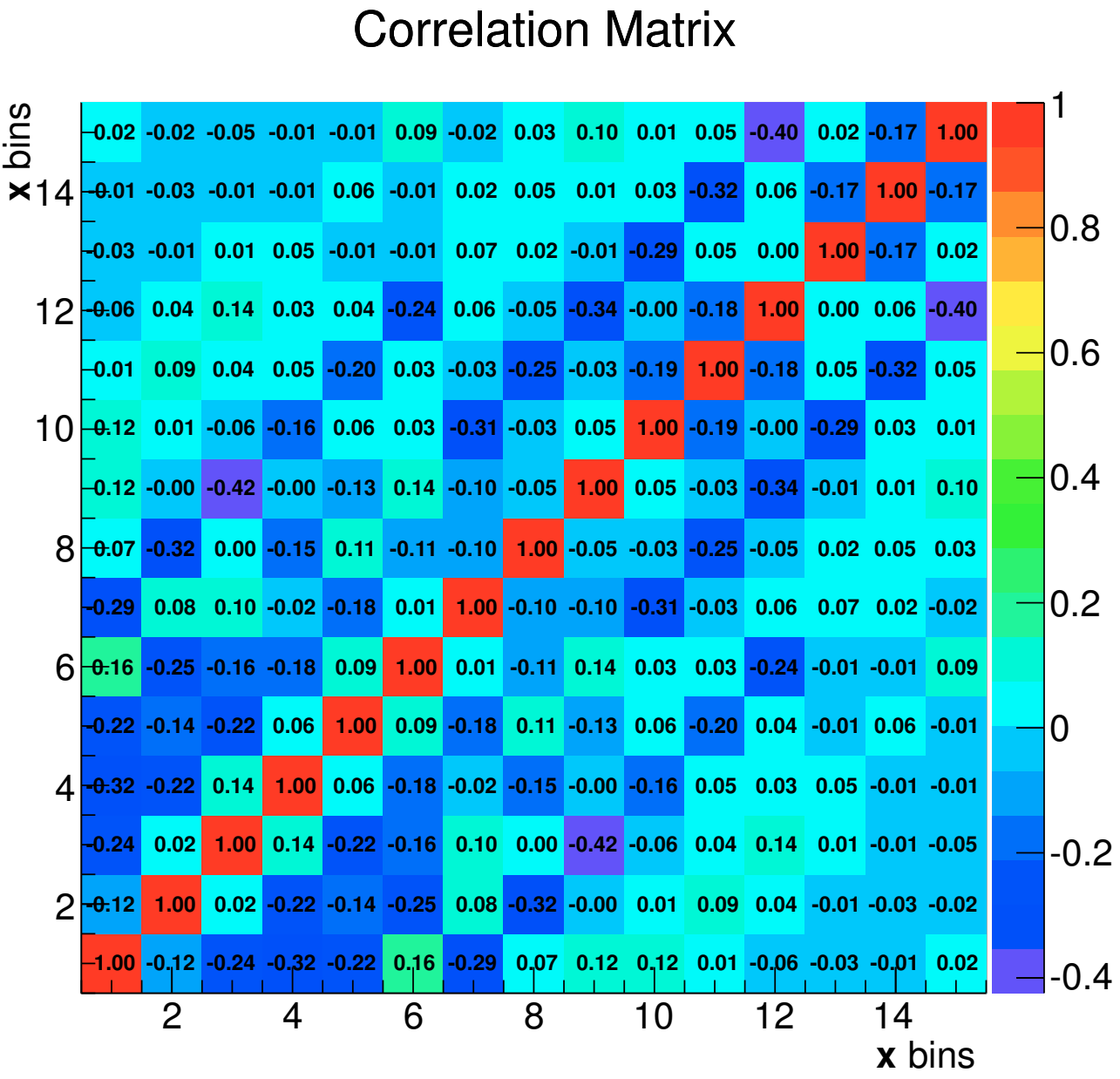

Figure J.3: Correlation matrix $\mathbf{V}_{xx}$ for the unfolded $\mathbf{x}$ distribution which corresponds to the $|y(t\bar{t})|$ in bins of $p_T(t)$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $p_T(t)$ – [0, 65, 130, 200, 400, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t\bar{t})|$ – [0, 0.45, 1.0, 2.3].

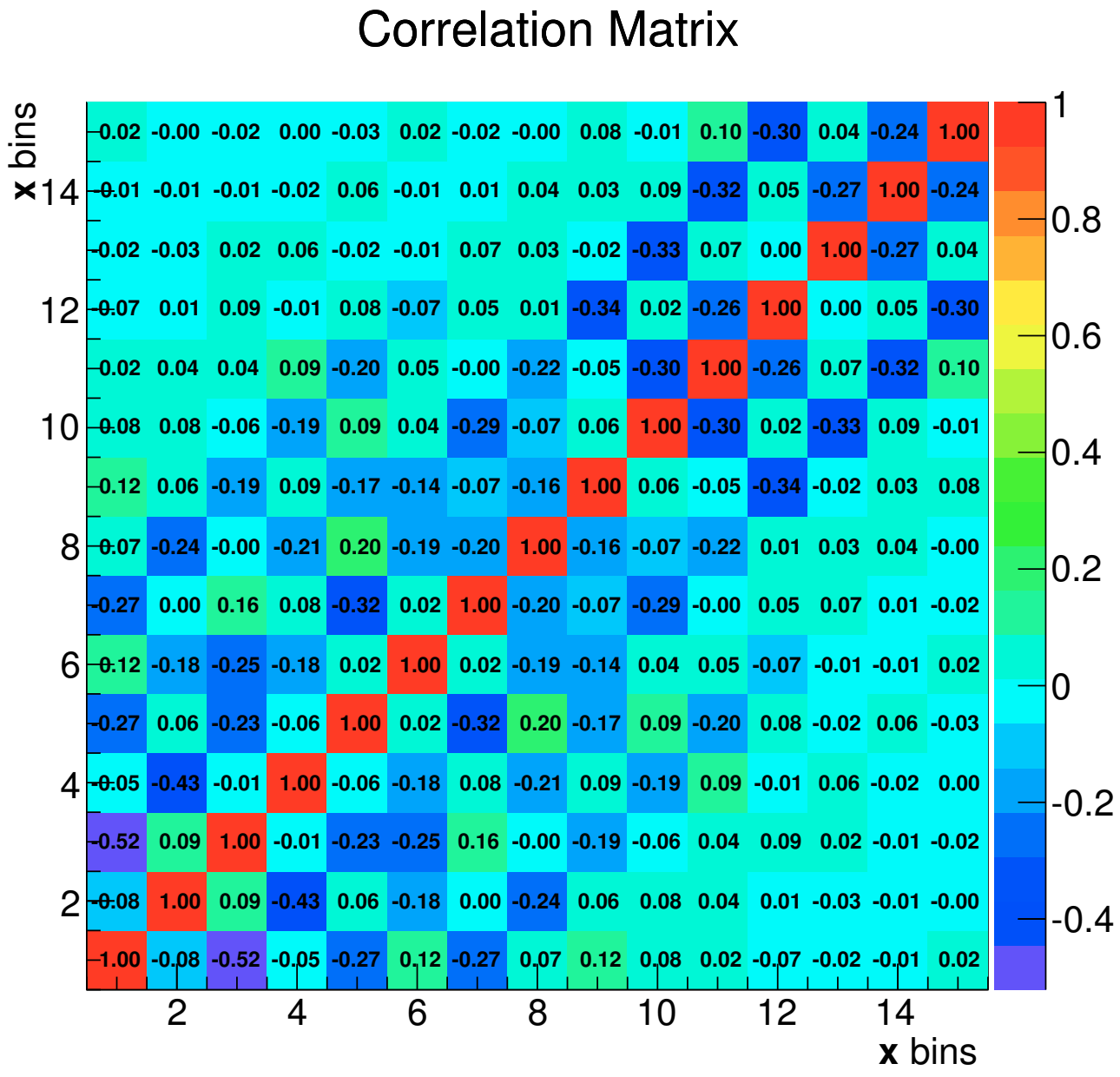

Figure J.4: Correlation matrix $\mathbf{V}_{xx}$ for the unfolded $\mathbf{x}$ distribution which corresponds to the $\Delta\eta(t\bar{t})$ in bins of $p_T(t)$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $p_T(t)$ – [0, 65, 130, 200, 400, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t\bar{t})|$ – [0, 0.8, 2.0, 6.0].

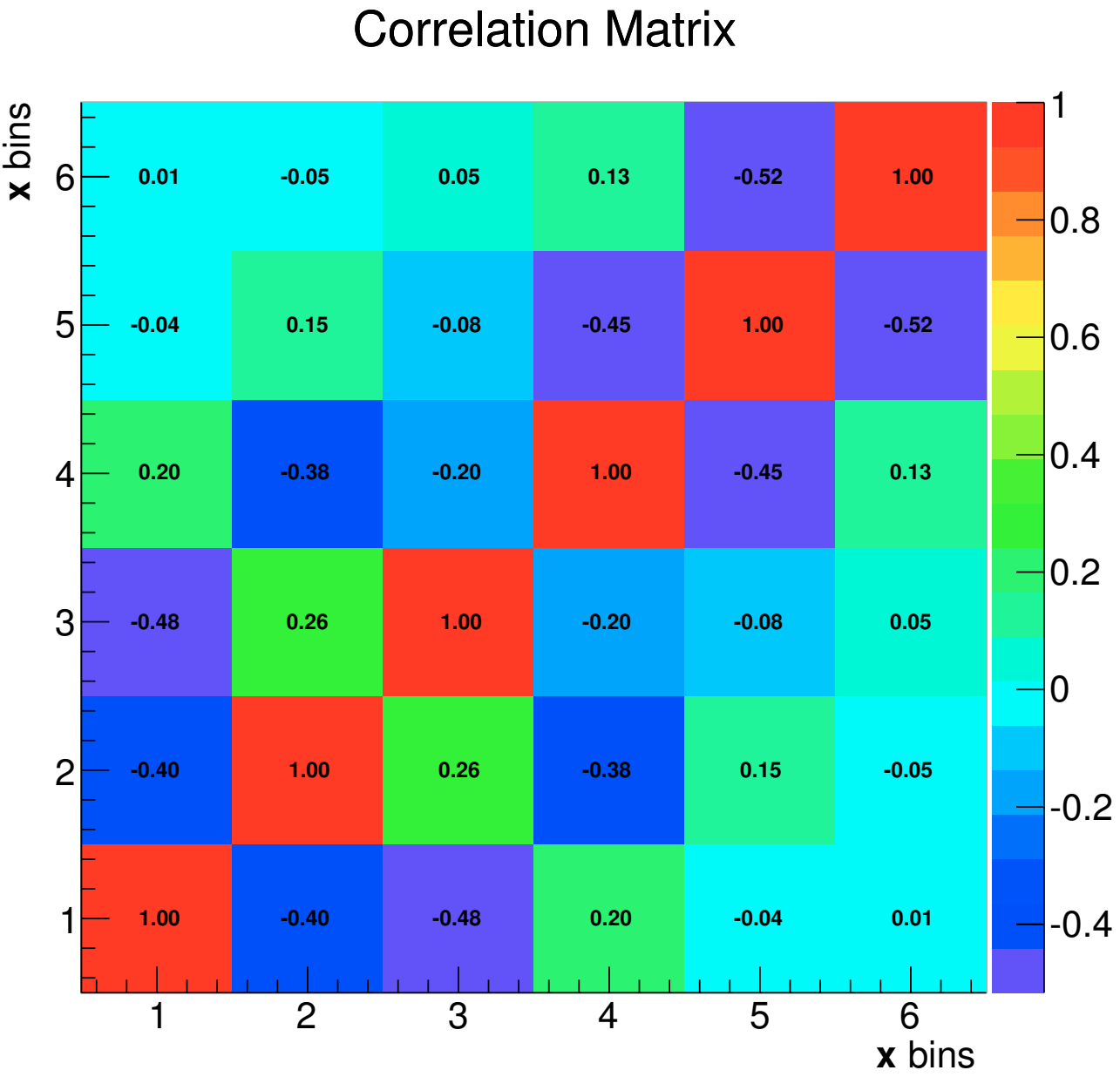

Figure J.5: Correlation matrix $\mathbf{V}_{xx}$ for the unfolded $\mathbf{x}$ distribution which corresponds to the $M(t\bar{t})$. The binning is the following: [340, 380, 470, 620, 820, 1100, 2200] GeV.

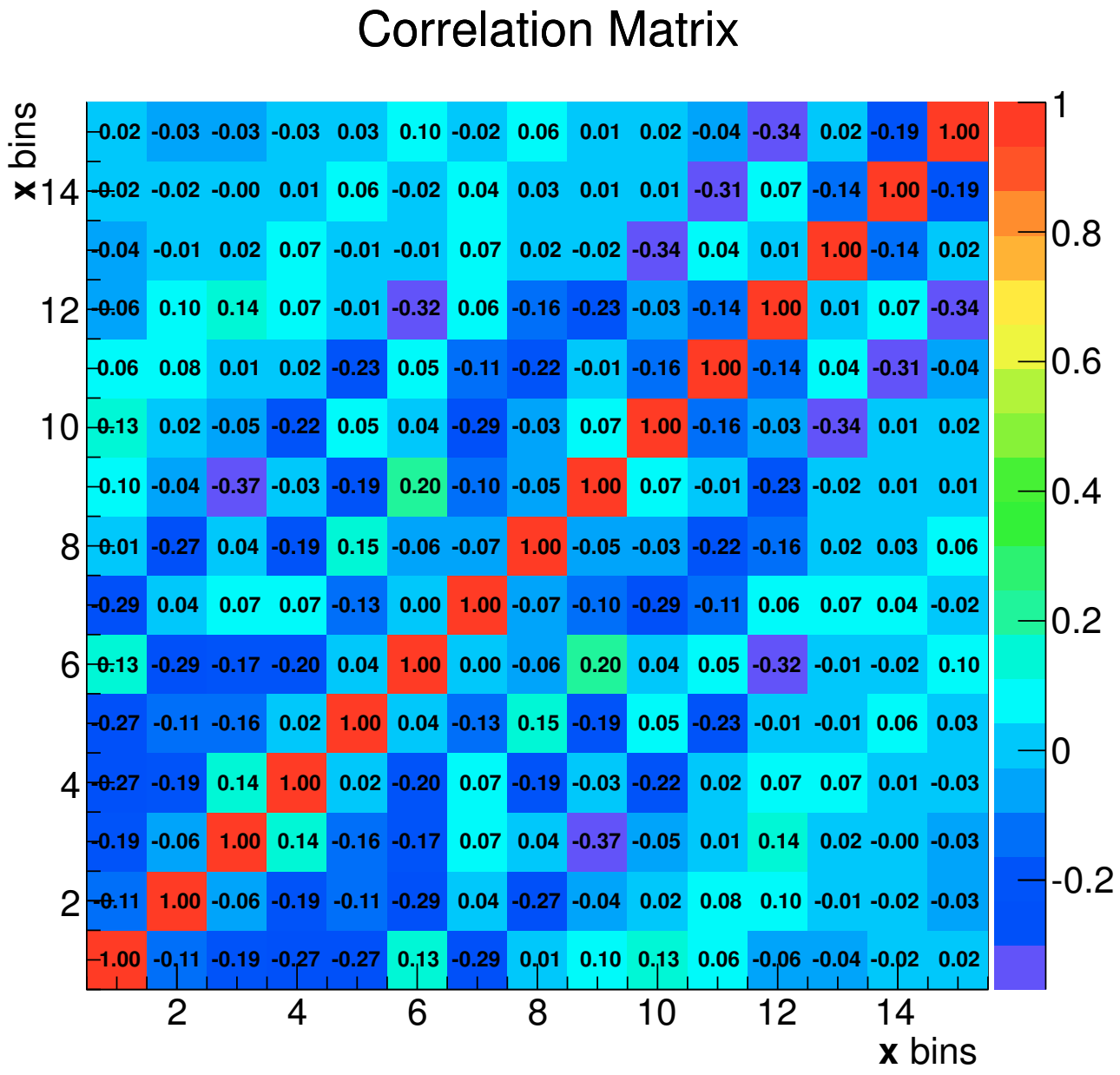

Figure J.6: Correlation matrix $\mathbf{V}_{xx}$ for the unfolded $\mathbf{x}$ distribution which corresponds to the $|y(t)|$ in bins of $M(t\bar{t})$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $M(t\bar{t})$ – [350, 390, 480, 600, 1000, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t)|$ – [0, 0.6, 1.2, 2.5].

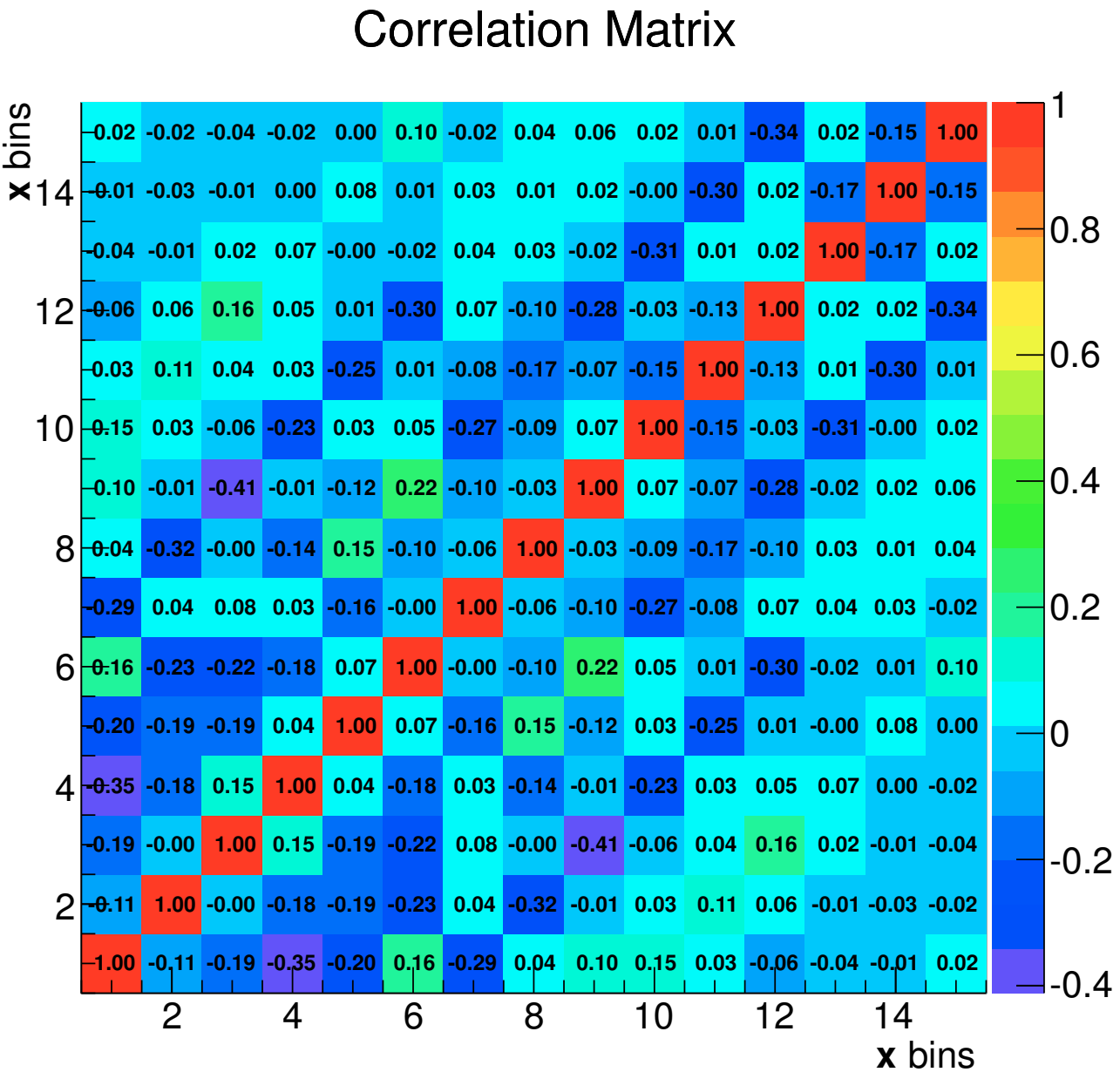

Figure J.7: Correlation matrix $\mathbf{V}_{xx}$ for the unfolded $\mathbf{x}$ distribution which corresponds to the $|y(t\bar{t})|$ in bins of $M(t\bar{t})$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $M(t\bar{t})$ – [350, 390, 480, 600, 1000, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t\bar{t})|$ – [0, 0.45, 1.0, 2.3].

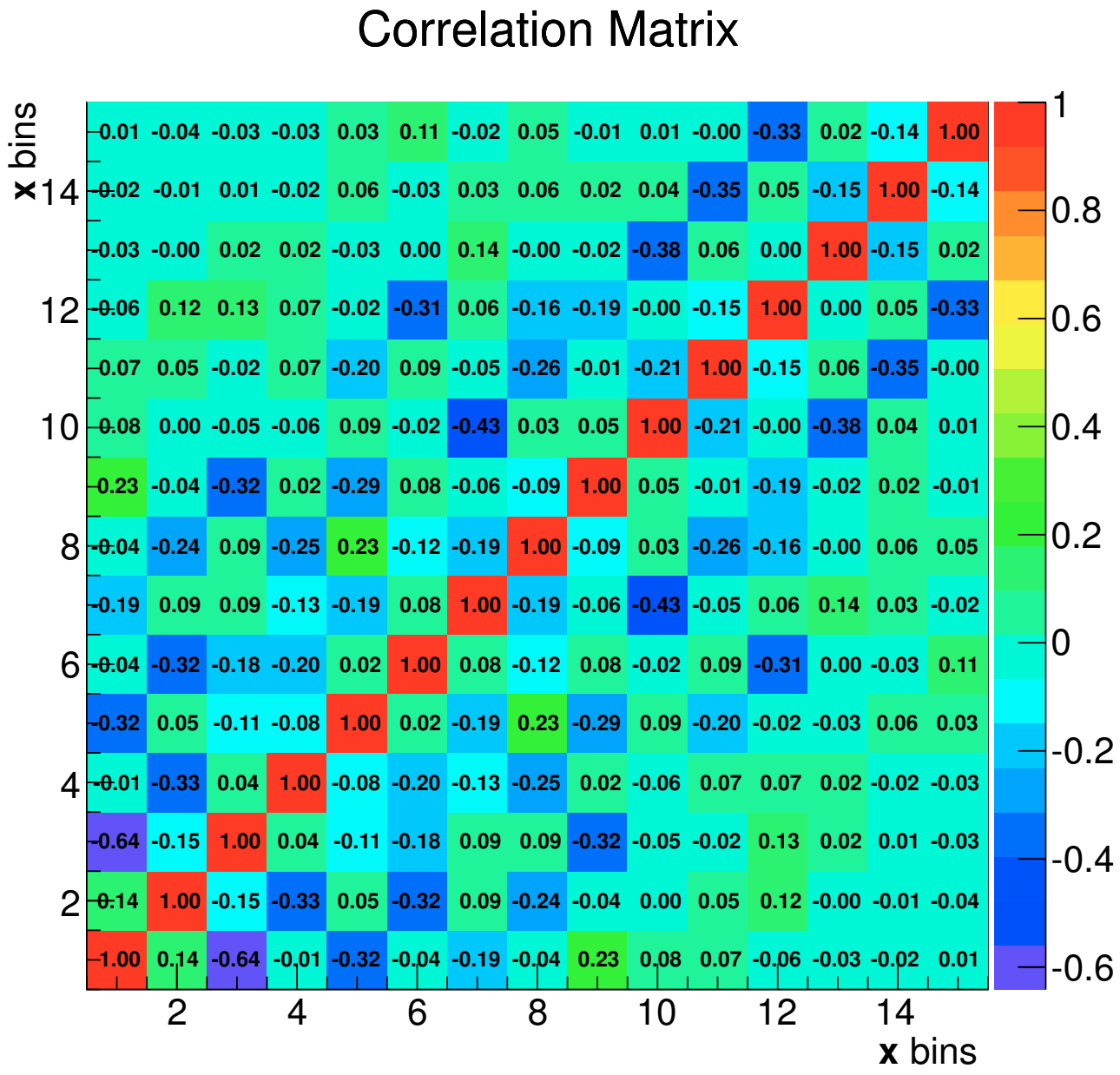

Figure J.8: Correlation matrix $\mathbf{V}_{xx}$ for the unfolded $\mathbf{x}$ distribution which corresponds to the $\Delta\eta(t\bar{t})$ in bins of $M(t\bar{t})$. The binning is the following: sequences of three bins (1-3, 4-6, 7-9, 10-12, 13-15) correspond to the bins of $M(t\bar{t})$ – [350, 390, 480, 600, 1000, 2000] GeV. The three bins in one sequence, which corresponds to one $p_T(t)$ bin are the bins of $|y(t\bar{t})|$ – [0, 0.8, 2.0, 6.0].

Bibliography

[1] E. D. Bloom and D. H. Coward and H. DeStaebler and J. Drees and G. Miller and L. W. Mo and R. E. Taylor and M. Breidenbach and J. I. Friedman and G. C. Hartmann and H. W. Kendall, *High-Energy Inelastic e-p Scattering at 6^o and 10^o*, *Phys. Rev. Lett.* **23** (1969) 930–934.

[2] M. Breidenbach and J. I. Friedman and H. W. Kendall and E. D. Bloom and D. H. Coward and H. DeStaebler and J. Drees and L. W. Mo and R. E. Taylor, *Observed Behavior of Highly Inelastic Electron-Proton Scattering*, *Phys. Rev. Lett.* **23** (1969) 935–939.

[3] J.-E. Augustin et. al., *Discovery of a Narrow Resonance in e^+e^- Annihilation*, *Physical Review Letters* **33** (1974) 1406–1408.

[4] J. J. Aubert et. al, *Experimental Observation of a Heavy Particle J*, *Physical Review Letters* **33** (1974) 1404–1406.

[5] L. M. Lederman, *The Upsilon Particle*, *Scientific American* **239** (1978) 72–81.

[6] **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.

[7] **D0** Collaboration, S. Abachi et al, *Search for High Mass Top Quark Production in $p\overline{p}$ Collisions at $\sqrt{s} =$ 1.8 TeV*, *Phys. Rev. Lett.* **74** (1995) 2422–2426.

[8] G. Aad et al., *Measurement of jet shapes in top-quark pair events at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *The European Physical Journal C* **73** (2013), no. 12 1–31.

[9] I. Korol, *Measurement of Double Differential $t\bar{t}$ Production Cross Sections with the CMS Detector*, Ph.D. thesis, Universität Hamburg (2016).

[10] "Elementary Particles and the World of Planck Scale ." https://universe-review.ca/F15-particle.htm

[11] "Standard Model." http://www.physik.uzh.ch/groups/serra/StandardModel.html

[12] G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Physics Letters B* **716** (2012), no. 1 1 – 29.

[13] **CMS** Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett.* **B716** (2012) 30–61, [arXiv:1207.7235].

[14] P. Dirac, *The Principles of Quantum Mechanics*. International series of monographs on physics. Clarendon Press (1981).

[15] Barger, V. and Marfatia, D. and Whisnant, K.L., *The Physics of Neutrinos*. Princeton University Press (2012).

[16] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, *Prog. Theor. Phys.* **49** (1973) 652–657.

[17] **Particle Data Group** Collaboration, J. Beringer et al., *Review of particle physics*, *Phys. Rev. D* **86** (2012) 010001.

[18] S. M. Bilenky and J. Hosek, *Glashow-Weinberg-Salam Theory of Electroweak Interactions and the Neutral Currents*, *Phys. Rept.* **90** (1982) 73–157.

[19] Higgs, P. W., *Broken Symmetries and the Masses of Gauge Bosons*, *Physical Review Letters* **13** (oct, 1964) 508–509.

[20] D. Griffiths, *Introduction to Elementary Particles*. Physics textbook. Wiley (2008).

[21] S. Bethke, *World Summary of α_s (2012)*, *Nucl.Phys.Proc.Suppl.* **234** (2013) 229–234, [arXiv:1210.0325].

[22] S. D. Drell and T.-M. Yan, *Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies*, *Phys. Rev. Lett.* **25** (1970) 316–320.

[23] F.-P. Schilling, *Top Quark Physics at the LHC: A Review of the First Two Years*, *Int. J. Mod. Phys.* **A27** (2012) 1230016, [arXiv:1206.4484].

[24] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl. Phys.* **B126** (1977) 298.

[25] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, *Sov. Phys. JETP* **46** (1977) 641–653. [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[26] V. N. Gribov and L. N. Lipatov, *Deep inelastic ep scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438–450. [Yad. Fiz.15,781(1972)].

[27] **ZEUS, H1** Collaboration, F. D. Aaron et al., *Combined Measurement and QCD Analysis of the Inclusive $e^{\pm}p$ Scattering Cross Sections at HERA*, *JHEP* **01** (2010) 109, [arXiv:0911.0884].

[28] **New Muon** Collaboration, M. Arneodo et al., *Measurement of the proton and deuteron structure functions, F2(p) and F2(d), and of the ratio σ-L/σ-T*, *Nucl. Phys.* **B483** (1997) 3–43, [hep-ph/9610231].

[29] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadopky, and W. K. Tung, *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07** (2002) 012, [hep-ph/0201195].

[30] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J.* **C75** (2015), no. 5 204, [arXiv:1412.3989].

[31] H. Abramowicz et. al., *Combination of Measurements of Inclusive Deep Inelastic $e^{\pm}p$ Scattering Cross Sections and QCD Analysis of HERA Data*, . *Temporary entry*.

[32] **Particle Data Group** Collaboration, K. A. Olive et al., *Review of Particle Physics*, *Chin. Phys.* **C38** (2014) 090001.

[33] "PARTON LUMINOSITY AND CROSS SECTION PLOTS." http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html

[34] **ATLAS** Collaboration, K. Aaboud et al., *Measurement of top quark pair differential cross-sections in the dilepton channel in pp collisions at $\sqrt{s}$ = 7 and 8 TeV with ATLAS*, arXiv:1607.0728.

[35] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J.* **C75** (2015), no. 11 542, [arXiv:1505.0448].

[36] **D0** Collaboration, V. M. Abazov et al., *Measurement of the $t\bar{t}$ production cross section using dilepton events in $p\bar{p}$ collisions*, *Phys. Lett.* **B704** (2011) 403–410, [arXiv:1105.5384].

[37] **CDF** Collaboration, T. Aaltonen et al., *Combination of CDF top quark pair production cross section measurements with up to 4.6 fb^{-1}*, *CDF note* **9913** (2009).

[38] **CDF, D0** Collaboration, T. A. Aaltonen et al., *Combination of measurements of the top-quark pair production cross section from the Tevatron Collider*, *Phys. Rev.* **D89** (2014), no. 7 072001, [arXiv:1309.7570].

[39] **D0** Collaboration, V. M. Abazov et al., *$t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$TeV*, *Phys. Rev. D* **67** (2003) 012004.

[40] **CDF** Collaboration, T. Affolder et al., *Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, *Phys. Rev. D* **64** (2001) 032001.

[41] **ATLAS** Collaboration, *Statistical combination of top quark pair production cross-section measurements using dilepton, single-lepton, and all-hadronic final states at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2012-024, CERN, Geneva, 2015.

[42] **ATLAS** Collaboration, *Measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s} = 8$ TeV using $e\mu$ events with b-tagged jets*, Tech. Rep. ATLAS-CONF-2013-097, CERN, Geneva, 2013.

[43] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **11** (2012) 067, [arXiv:1208.2671].

[44] **CMS** Collaboration, *Top pair cross section in e/mu+jets at 8 TeV*, Tech. Rep. CMS-PAS-TOP-12-006, CERN, Geneva, 2012.

[45] **CMS** Collaboration, *Top pair cross section in dileptons*, Tech. Rep. CMS-PAS-TOP-12-007, CERN, Geneva, 2012.

[46] "Useful Diagrams of Top Signals and Backgrounds." http://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html

[47] L. Evans and P. Bryant, *LHC Machine*, *Journal of Instrumentation* **3** (2008), no. 08 S08001.

[48] "CERN|Accelerating science." http://www.cern.ch

[49] A. Blas et al., *The PS complex as proton pre-injector for the LHC: Design and implementation report*, Tech. Rep. CERN-2000-003, CERN-2000-03, 2000.

[50] "CERN Optic Accelerators Site." http://cern-accelerators-optics.web.cern.ch/cern-accelerators-optics/

[51] L. Arnaudon and P. Baudrenghien and M. Baylac et al., *Linac4 Technical Design Report*, Tech. Rep. CERN-AB-2006-084. CARE-Note-2006-022-HIPPI, CERN, Geneva, 2006.

[52] *LEP design report*, tech. rep.

[53] **ALICE** Collaboration, *ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC*. LHC Tech. Proposal. CERN, Geneva (1995).

[54] **LHCb** Collaboration, R. Antunes-Nobrega et al., *LHCb reoptimized detector design and performance: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva (2003).

[55] **ATLAS** Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.

[56] **CMS** Collaboration, S. Chatrchyan et al., *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[57] **TOTEM** Collaboration, G. Latino, *The TOTEM Experiment at the LHC*, arXiv:0905.2936.

[58] A. Tricomi et al., *The LHCf experiment at the LHC: Physics Goals and Status*, *Nucl. Phys. B, Proc. Suppl.* **196** (2009) 30–35.

[59] "CERN - CMS prepares for the future." http://orbiterchspacenews.blogspot.de/2013/04/cern-cms-prepares-for-future.html

[60] D. Sprenger, M. Weber, R. Adolphi, R. Brauer, L. Feld, K. Klein, A. Ostaptchuk, S. Schael, and B. Wittmer, *Validation of Kalman Filter alignment algorithm with cosmic-ray data using a CMS silicon strip tracker endcap*, *JINST* **5** (2010) P06007, [arXiv:1003.5460].

[61] **CMS** Collaboration, *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, *Journal of Instrumentation* **9** (2014), no. 10 P10009.

[62] **CMS** Collaboration, S. Chatrchyan et al., *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, *JINST* **9** (2014), no. 10 P10009, [arXiv:1405.6569].

[63] B. Isildak, *Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s} = 7$ TeV*, arXiv:1308.6064.

[64] **Particle Data Group** Collaboration, S. Eidelman et al., *Review of particle physics. Particle Data Group*, *Phys. Lett.* **B592** (2004) 1–1109.

[65] P. Adzic et al., *Energy resolution of the barrel of the CMS electromagnetic calorimeter*, *JINST* **2** (2007) P04004.

[66] "XDAQ - CMS Online Software project page." http://xdaq.web.cern.ch/xdaq/setup/images/HCAL.png

[67] V. D. Elvira, *Measurement of the Pion Energy Response and Resolution in the CMS HCAL Test Beam 2002 Experiment*, Tech. Rep. CMS-NOTE-2004-020, CERN, Geneva, 2004.

[68] "CMS-solenoid." http://irfu.cea.fr/en/Phocea/Vie_des_labos/Ast/ast_sstechnique.php?id_ast=2261

[69] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [hep-ph/0603175].

[70] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV*, *JHEP* **09** (2011) 109, [arXiv:1107.0330].

[71] T. Gleisberg and S. Höche and F. Krauss and M. Schönherr and S. Schumann and F. Siegert and J. Winter, *Event generation with SHERPA 1.1*, *Journal of High Energy Physics* **2009** (2009), no. 02 007.

[72] B. Andersson, *The Lund model*, *Nuclear Physics A* **461** (1987), no. 1 513 – 520.

[73] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber, *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *JHEP* **01** (2001) 010, [hep-ph/0011363].

[74] B. R. Webber, *Fragmentation and hadronization*, *Int. J. Mod. Phys.* **A15S1** (2000) 577–606, [hep-ph/9912292]. [577(1999)].

[75] *ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11*, Tech. Rep. ATL-PHYS-PUB-2011-009, CERN, Geneva, 2011.

[76] F. Maltoni and T. Stelzer, *MadEvent: Automatic event generation with MadGraph*, *JHEP* **02** (2003) 027, [hep-ph/0208156].

[77] S. Mrenna and P. Richardson, *Matching matrix elements and parton showers with HERWIG and PYTHIA*, *JHEP* **05** (2004) 040, [hep-ph/0312274].

[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].

[79] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan, *New parton distributions for collider physics*, *Phys. Rev.* **D82** (2010) 074024, [arXiv:1007.2241].

[80] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029, [hep-ph/0204244].

[81] "Introduction to Geant4." http://geant4.web.cern.ch/geant4/UserDocumentation/UsersGuides/IntroductionToGeant4/html/index.html

[82] P. Billoir, *Progressive track recognition with a Kalman like fitting procedure*, *Comput. Phys. Commun.* **57** (1989) 390–394.

[83] P. Billoir and S. Qian, *Simultaneous pattern recognition and track fitting by the Kalman filtering method*, *Nucl. Instrum. Meth.* **A294** (1990) 219–228.

[84] R. Mankel, *A Concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system*, *Nucl. Instrum. Meth.* **A395** (1997) 169–184.

[85] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, *Nucl. Instrum. Meth.* **A262** (1987) 444–450.

[86] K. Rose, *Deterministic annealing for clustering, compression, classification, regression, and related optimization problems*, *Proceedings of the IEEE* **86** (1998), no. 11 2210–2239.

[87] New world record - first pp collisions at 8 TeV . http://cms.web.cern.ch/news/new-world-record-first-pp-collisions-8-tev

[88] **CMS** Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*, Tech. Rep. CMS-PAS-PFT-09-001, 2009.

[89] "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.gi

[90] **CMS** Collaboration, S. Chatrchyan et al., *Performance of CMS muon reconstruction in pp collision events at* $\sqrt{s} = 7$ *TeV*, *JINST* **7** (2012) P10002, [arXiv:1206.4071].

[91] **CMS** Collaboration, *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*, Tech. Rep. CMS-PAS-PFT-10-001, 2010.

[92] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*, Tech. Rep. CMS-NOTE-2005-001, CERN-CMS-NOTE-2005-001, 2005.

[93] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois, *Electron reconstruction in CMS*, *Eur. Phys. J.* **C49** (2007) 1099–1116.

[94] G. P. Salam and G. Soyez, *A Practical Seedless Infrared-Safe Cone jet algorithm*, *JHEP* **05** (2007) 086, [arXiv:0704.0292].

[95] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev.* **D48** (1993) 3160–3166, [hep-ph/9305266].

[96] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001, [hep-ph/9707323].

[97] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **04** (2008) 063, [arXiv:0802.1189].

[98] R. Frühwirth, W. Waltenberger, and P. Vanlaer, *Adaptive vertex fitting*, *J. Phys.* **G34** (2007) N343.

[99] *b-tagging in dense environments*, Tech. Rep. ATL-PHYS-PUB-2014-014, CERN, Geneva, 2014.

[100] **CMS** Collaboration, *Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, $t\bar{t}$ and boosted topology events*, Tech. Rep. CMS-PAS-BTV-13-001, 2013.

[101] **CMS** Collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS*, *Journal of Instrumentation* **6** (2011) 11002, [arXiv:1107.4277].

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

[103] "Met regression in high pile-up environment." https://twiki.cern.ch/twiki/bin/viewauth/CMS/MVAMet

[104] I. Asin Cruz, *Measurement of Top-Quark-Pair Differential Cross Sections in Proton-Proton Collisions at $\sqrt{s} = 8$TeV with the CMS Experiment*, Ph.D. thesis, Universität Hamburg (2014).

[105] "Offline DQM Shift Instructions." https://twiki.cern.ch/twiki/bin/viewauth/CMS/OfflineDQMShifts

[106] "JSON File." /afs/cern.ch/cms/CAF/certification/Collisions12/8TeV/929Reprocessing/Cert_190456-208686_8TeV_22Jan2013ReReco_930Collisions12_JSON.txt

[107] **CMS** Collaboration, C. Diez Pardos and J. Kieseler, *Dilepton trigger and lepton identification efficiencies for the top quark pair production cross section measurements at 8 TeV in the dilepton decay channel, CMS-AN-12-389 [Internal Note].*

[108] HCAL Noise Library . https://twiki.cern.ch/twiki/bin/viewauth/CMS/HcalNoiseInfoLibrary

[109] **CMS** Collaboration, *Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update*, Tech. Rep. CMS-PAS-SMP-12-008, 2012.

[110] "Estimating systematic errors due to pileup modeling." https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupSystematicErrors Accessed: 2015-07-28.

[111] "Standard model cross sections for cms at 8 tev." https://twiki.cern.ch/twiki/bin/view/CMS/StandardModelCrossSectionsat8TeV Accessed: 2016-06-25.

[112] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the $t\bar{t}$ 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].

[113] S. Fleischmann, *Boosted top quark techniques and searches for $t\bar{t}$ resonances at the LHC*, *Journal of Physics: Conference Series* **452** (2013), no. 1 012034.

[114] Uncertainty on the differential shapes. BTV Twiki Page . https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagPOG#2012_Data_and_MC_EPS13_prescript

[115] L. Sonnenschein, *Analytical solution of $t\bar{t}$ dilepton equations*, *Phys. Rev. D* **73** (2006) 054015.

[116] L. Sonnenschein, *Erratum: Analytical solution of $t\bar{t}$ dilepton equations [Phys. Rev. D **73** , 054015 (2006)]*, *Phys. Rev. D* **78** (2008) 079902.

[117] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of differential top-quark pair production cross sections in pp colisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J.* **C73** (2013), no. 3 2339, [arXiv:1211.2220].

[118] **CMS** Collaboration, *Measurement of the differential ttbar production cross section for high-pt top quarks in e/mu+jets final states at 8 TeV*, .

[119] S. Schmitt, *TUnfold: an algorithm for correcting migration effects in high energy physics*, *JINST* **7** (2012) T10003, [arXiv:1205.6201].

[120] A. N. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, *Soviet Math. Dokl.* **4** (1963) 1035–1038.

[121] A. Hocker and V. Kartvelishvili, *SVD approach to data unfolding*, *Nucl. Instrum. Meth.* **A372** (1996) 469–481, [hep-ph/9509307].

[122] **CMS** Collaboration, *Performance of b tagging at $\sqrt{s}$=8 TeV in multijet, $t\bar{t}$ and boosted topology events*, Tech. Rep. CMS-PAS-BTV-13-001, CERN, Geneva, 2013.

[123] **CMS** Collaboration, *CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update*, Tech. Rep. CMS-PAS-LUM-13-001, CERN, Geneva, 2013.

[124] **ATLAS** Collaboration, G. Aad et al., *Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in $\sqrt{s}$ = 8 TeV proton-proton collisions using the ATLAS detector*, *Phys. Rev.* **D93** (2016), no. 3 032009, [arXiv:1510.0381].

[125] P. Langacker, *The Physics of Heavy Z′ Gauge Bosons*, *Rev. Mod. Phys.* **81** (2009) 1199–1228, [arXiv:0801.1345].

[126] **CMS** Collaboration, V. Khachatryan et al., *Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s}$ = 8 TeV*, *Phys. Rev.* **D93** (2016), no. 1 012001, [arXiv:1506.0306].

[127] **CMS** Collaboration, *Search for ttbar resonances in dilepton+jets final states in pp collisions at 8 TeV*, Tech. Rep. CMS-PAS-B2G-12-007, CERN, Geneva, 2014.

Acknowledgements

I would like to express my thankfulness to the people, who were supporting and helping me throughout the whole path towards the competition of this work.

My first acknowledgement goes to Olaf Behnke, my supervisor and main support in this work. I am very thankful to you, Olaf, for your ideas, knowledge and experience which you shared with me all the time. I really enjoyed the discussions with you. You are always emitting endless scientific curiosity, which was very important and inspiring for me every time I felt like falling into routine work. Thank you also for being informal and cheerful.

I would also like to thank Peter Schleper, who was reading through my thesis very thoroughly. Your comments were very deep and useful.

My further acknowledgement goes to Christian Sander, with whom we had numerous discussions on the analysis (especially on the kinematic reconstruction part). Thank you also for reading through my thesis and giving wise suggestions.

I am also very thankful to Stefan Schmitt for helping me with understanding and implementing the TUnfold package to my analysis.

I would also like to express my gratitude towards Daniel Pitzl, who was guiding me during my technical task related to the Phase I CMS pixel detector upgrade. You are always there ready to share your knowledge with the students. Thank you for making us do the things correct and from the first try.

I would like to thank all the people from the DESY Top Group. You were always there, ready to have fruitful and informative discussions on my analysis and on physics in general. I have really enjoyed the environment I was working in.

My gratitude also goes to my wonderful office mates - Nazar Bartosik and Jasone Garay. You guys are so different, but this made a perfect balance. It was never boring in our office. Our numerous discussions on all the possible and impossible topics were a real pleasure to participate.

And the last but not least is my family, a perfect and powerful background for me. First of all, I would like to thank my parents and grandmother who were always a live example of hard working people to me. Thank you for highly prioritizing my education. And now my newest and biggest support is my husband and my wonderful children. No matter what happens in my life and in my work, I will always keep smiling, because I have you around.