Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Společná laboratoř optiky UP
a FzÚ AV ČR v Olomouci

DIPLOMOVÁ PRÁCE

Studium kinematiky párů top kvarků v experimentu ATLAS

Vypracoval:	**Jan Palička**
Studijní program:	N1701 Fyzika
Studijní obor:	1702T001 / Aplikovaná fyzika
Forma studia:	Prezenční
Vedoucí diplomové práce:	Mgr. Jiří Kvita, Ph.D.
Termín odevzdání práce:	Srpen 2017

Prohlášení

Prohlašuji, že jsem předloženou diplomovou práci vypracoval samostatně pod vedením Jiřího Kvity a že jsem použil zdrojů, které cituji a uvádím v seznamu použitých pramenů.

V Olomouci dne 7. 8. 2017

.....

Jan Palička

Bibliografická identifikace

Jméno a příjmení autora	Jan Palička
Název práce	Studium kinematiky párů top kvarků v experimentu ATLAS
Typ práce	Diplomová
Pracoviště	Společná laboratoř optiky Univerzity Palackého v Olomouci a Fyzikálního ústavu Akademie věd ČR
Vedoucí práce	Mgr. Jiří Kvita, Ph.D.
Rok obhajoby práce	2017
Abstrakt	Cílem diplomové práce je seznámit se s analýzou dat ve fyzice vysokých energií, a to od simulace po zpracování reálných dat s experimentem ATLAS na urychlovači LHC v laboratoři CERN se zaměřením na fyziku top kvarku, konkrétně na události, kdy vznikají páry top kvarků s velkým příčným impulzem, které umožňují studium fyziky top kvarku při velkých přenesených hybnostech, a srovnat naměřená data s teoretickými předpověďmi.
Klíčová slova	Top kvark, ATLAS, simulace, MadGraph, boostovaný režim, studium výběrových kritérií, fyzika vysokých energií.
Počet stran	67
Počet příloh	1
Jazyk	Anglický

Bibliographical identification

Autor's first name and surname	Jan Palička
Title	Study kinematics pairs of top quarks in ATLAS
Type of thesis	Master
Department	Joint Laboratory of Optics of Palacky Univerzity in Olomouc and Institute of Physics of the Czech Academy of Sciences
Supervisor	Mgr. Jiří Kvita, Ph.D.
The year of presentation	2017
Abstract	The aim of the diploma thesis is to get acquainted with the data analysis in high energy physics with the ATLAS experiment at the LHC collider at CERN, from simulation to real data, focusing on the physics of top quark pairs with the high transverse momentum by studying events with high momentum transfer and comparing the measured spectra to theoretical predictions.
Keywords	The top quark, ATLAS, simulation, MadGraph, boosted region, selection and optimisation study, high energy physics.
Number of pages	67
Number of appendices	1
Language	English

Contents

Introduction

This diploma thesis is focused on the analysis of the top quark pair production in pp collisions at the Large Hadron Collider using the ATLAS detector. The thesis is split into three main parts. In the first part basics facts about the Standard Model particles and the top quark are presented. The second part is devoted to an analysis based on the Monte-Carlo generator MadGraph and the detector simulator Delphes using generated $t\bar{t}$ signal samples and simple cuts. The last part is about work on an ATLAS analysis with main results of measured data and Monte-Carlo comparison presented, and a study of correlation between measured variables.

Chapter 1

What is the Top Quark

The top quark is a standard model particle discovered by CDF and D0 experiments at the Tevatron accelerator in 1995. It is the heaviest elementary particle in the Standard Model (SM). The mass of the top quark is $173.34 \pm 0.27(\text{stat}) \pm 0.71(\text{syst})$ GeV [10], measured by Tevatron and LHC experiments. Current all discovered SM particles are displayed in Figure 1.1. There are the leptons and the corresponding neutrinos, three families of a quarks, intermedial bosons mediating interactions and the Higgs boson, Higgs mechanism is essential to explain the property of "mass" of particles.

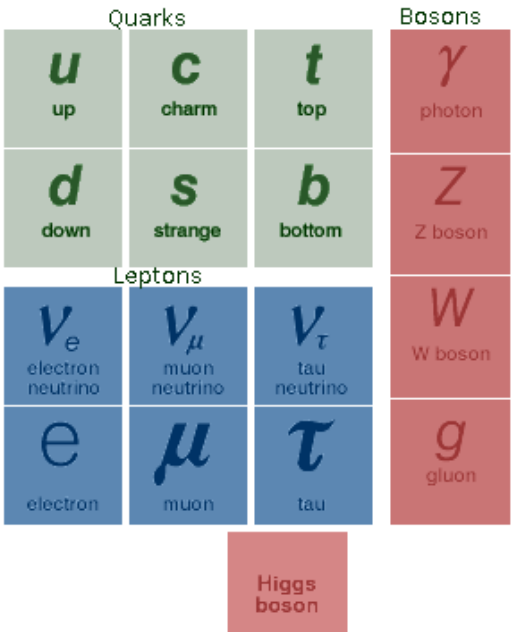

Figure 1.1: The Standard Model particles.

1.1 Top quark production

At the leading order (LO) of quantum chromodynamics (QCD) which describes the strong interactions of quarks and gluons, in proton-proton (pp) collisions at LHC at the center-of-mass energy $\sqrt{s}$ = 13 TeV the top quarks are produced by the quarks annihilation ∼10% [9] the Feynman diagram in Figure 1.2 (a), or the dominant production process of the gluon fusion ∼90% [9] the Feynman diagrams in Figures 1.2 (b), (c) and (d). With the increasing center of mass energy $\sqrt{s}$ the quarks annihilation fraction decreases while the gluon-gluon fusion increases. Also with larger center-of-mass energy the cross section increases as it is indicate in Table 1.1, for center-of-mass $\sqrt{s}$ = 13 TeV the predicted cross section is $815.96^{+19.37}_{-28.61}$ pb [15].

Collider	$\sigma_{t\bar{t}}$ [pb] NNLO	NNLO+NNLL
Tevatron, $\sqrt{s} = 1.96$ TeV	$7.009^{+0.259\ +0.169}_{-0.374\ -0.121}$	$7.164^{+0.111\ +0.169}_{-0.200\ -0.122}$
LHC, $\sqrt{s} = 7$ TeV	$167.0^{+6.7\ +4.6}_{-10.7\ -4.7}$	$172.0^{+4.4\ +4.7}_{-5.8\ -4.8}$
LHC, $\sqrt{s} = 8$ TeV	$239.1^{+9.2\ +6.1}_{-14.8\ -6.2}$	$245.8^{+6.2\ +6.2}_{-8.4\ -6.4}$
LHC, $\sqrt{s} = 14$ TeV	$933.0^{+31.8\ +16.1}_{-51.0\ -17.6}$	$953.6^{+22.7\ +16.2}_{-33.9\ -17.8}$

Table 1.1: The theoretical top quark cross sections at LHC (the center-of-mass energy $\sqrt{s} = 7$ TeV, $\sqrt{s} = 8$ TeV, $\sqrt{s} = 14$ TeV) and Tevatron ($\sqrt{s} = 1.96$ TeV), edited from [9].

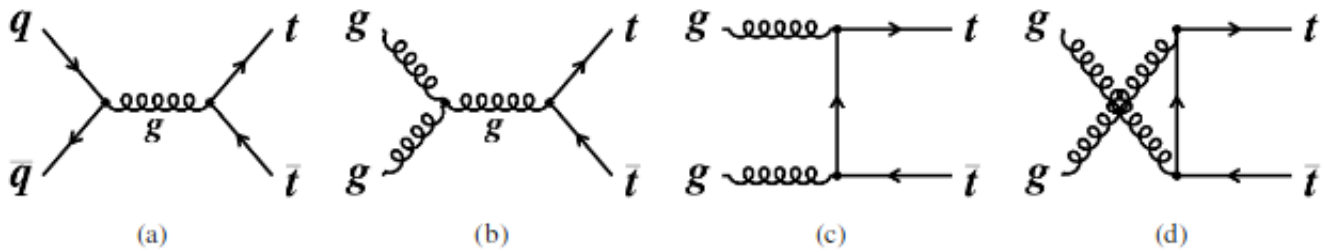

Figure 1.2: Top quark pair production Feynman diagrams at the leading order of QCD, edited from [9].

1.2 Top quark decay

In the SM the top quark decays in 99% cases to a b quark and a W boson. Decay modes are displayed in Figure 1.3. The top quark pair final states can be divided by the individual top quark decay modes to the all-hadronic (in Figure 1.4: "alljets") channel where the both W bosons decay to quark pairs, the semi-leptonic channel where one of the W bosons decay to a quarks pair and the other one to

a charged lepton and a neutrino, and the di-lepton channel where both W bosons decay to leptons and neutrinos. There are three charged leptons in the Standard Model and the leptonic channel is further divided based on their flavours as in Figure 1.4.

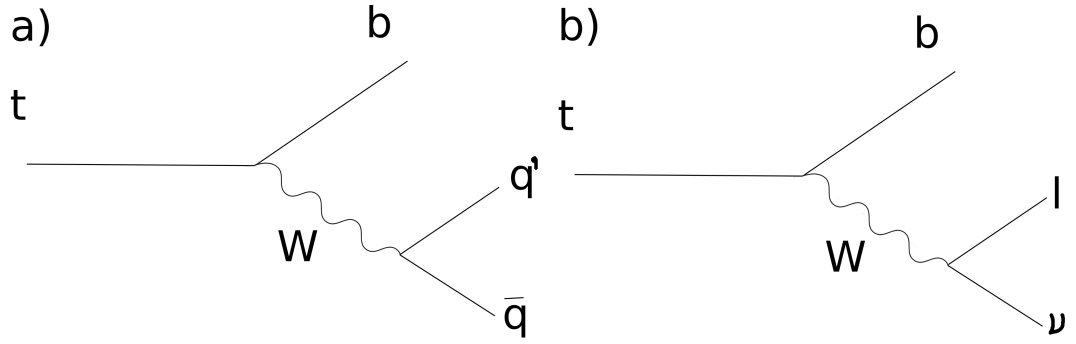

Figure 1.3: Top quark decay modes a) hadronic b) leptonic.

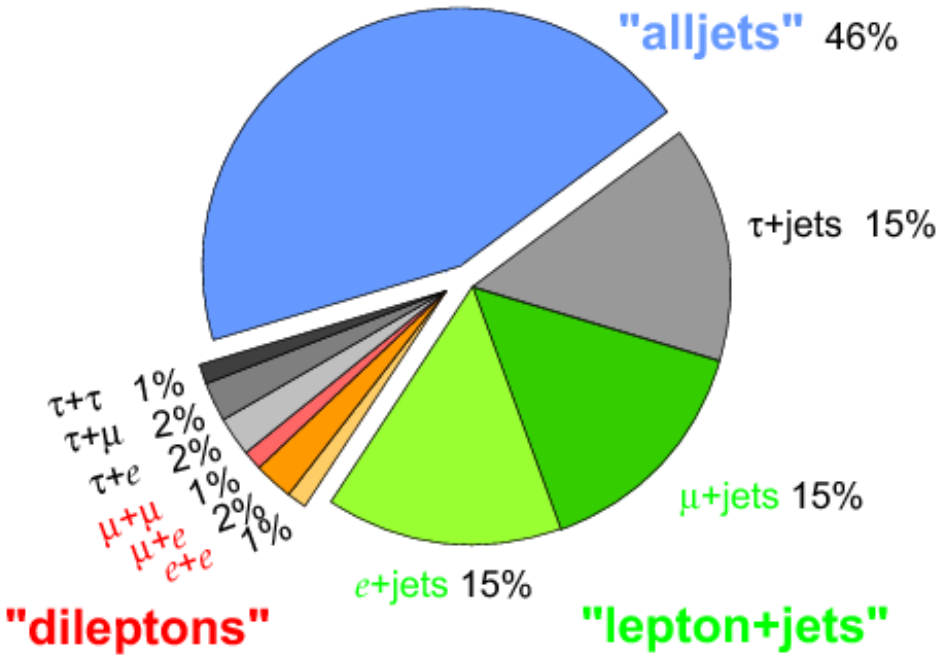

Figure 1.4: Top quark decay branching ratio [3].

From Figure 1.4 it is evident that the largest fraction is the "alljets" channel, where the W boson decays to quark anti-quark pair and 3 jets from the W boson and a b quark are expected, or in boosted topology, one large-R jet with some sub-jets substructure inside.

1.3 Top quark signatures in a detector

A particle jet is defined by a cone of final state particles after hadronisation, then in the calorimeter the jet is defined as a collimated deposit of energy in the detectors. When particles enter the calorime-

ter, they lose energy by ionisation and radiation processes and the primary particle is stopped. The jet is characterized by a four vector of the information from the tracking detectors and the calorimeters. The raw data from the detector are stored and then are calibrated and reduce for a faster analysis. The energy deposits from calorimeters are reduced to a jet, which is a four-vector with energy and momentum (E,p_x,p_y,p_z). Jets also have a finite size in $\eta - \phi$ plane. By a conventional choice, small jets are used with $R = 0.4$ and large-R jets with $R = 1.0$.

Specifically for the all-hadronic top quark pair decay mode in which top quark decays to a W boson and a b quark and the W boson decays to two quarks, there are two topologies. First topology is called resolved where after hadronisation there are three resolved (angularly separated) deposits of energy, jets. In contrast, the boosted topology is characterized by a very high top quark p_T, the jets are highly boosted and merged and there is only one large-R jet, the difference is illustrated in Figure 1.5. The large-R jet is then checked for substructure that can be described by a τ variables, explained in Chapter 2.4.2.

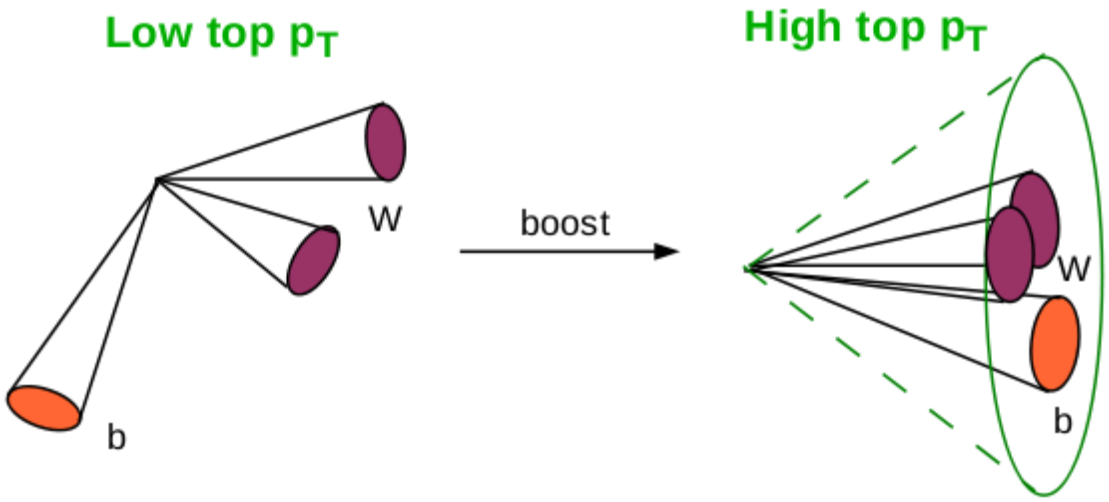

Figure 1.5: Resolved and boosted topology of the top quark decay products in the laboratory frame, edited from [2].

Chapter 2

Private analysis

This Chapter is focused on a simulation of the top quark pair production at the $\sqrt{s} = 13$ TeV energy of the Large Hadron Collider (LHC) and detected by a simplified model simulation of the ALTAS detector using a MadGraph, Pythia8, Delphes and ROOT frameworks. The main motivation for using MadGraph is simplicity and variability. Anyone can use it for generating events of various collisions and energies with simple configuration files. One can extract any information about the event. It is a useful framework for theoretical studies of the top quark pair kinematic. This part of thesis is focused on a simple analysis of the top quark pair production in pp collisions.

2.1 MadGraph

MadGraph [11] is Monte-Carlo event generator with aplications to hadron collider physics. It is a $C++$ framework controlled from the Linux terminal. With the special command or the text file with the commands, MadGraph generates events and their full kinematics. These can be generated in pp collisions, e^+e^- or the proton anti-proton ($p\bar{p}$) collisions, at any energy. It can be used from a web interface or on a local computer. For process generation it can work with Standard Model and also with some theories beyond, like Supersymmetry. The model configuration files define the particles interactions and properties including their invariant masses and life times. The events are generated on parton level, then passed to a parton shower framework Pythia8 [12] the detector level simulation the framework Delphes [7] is used. Those extensions of Delphes and Pythia8 are automatically loaded to MadGraph and use similar configurations files. MadGraph can safely generate 50 000 events, for more statistic one needs to merge samples together.

MadGraph is easy to use, after download and extraction it can be started immediately and with the command *install* extensions can be added for example the aforementioned Pythia8 for showering and Delphes for detector simulation. For the study of the top-antitop pair production in the all-hadronic

channel the following commands for MadGraph generator were used [20]

`generate` $pp > tt\sim, t > jjb, t\sim > jjb$ @0

`add process` $pp > tt\sim j, t > jjb, t\sim > jjb$ @1

which generate the top quark and antitop quark at NLO prediction and decays them hadronicaly, so in the detector we expects 6 jets in the resolved topology or 2 large-R jets in the boosted topology. In order to promote the boosted topology, requirements on the p_T of each top quark is placed to be larger than 200 GeV and the anti-kt jet finder with the $\Delta R = 1.0$ is used. These conditions are set in the main MadGraph configuration file *run_card.dat*. In MadGraph the top quark mass is $m_t = 172$ GeV.

2.2 Showering

After the MadGraph parton-level production there is the step called showering, to account for the particles radiating photons or gluons before their hadronize. For showering, Pythia8 is used. These two steps produce files for further detector simulation. 10 samples each with the 50 000 events were generated, amounting to a 500 000 events signal sample for the demonstration of the analysis below.

2.3 Detector simulation (Delphes)

Delphes [7] si a $C++$ framework for fast detector simulation. The detector in simulation includes a tracking system with a magnetic field, calorimeters and muon chambers. There is a set of detector resolution, non-homogeneity in the magnetic field, gaps between detector segments. It provides information about particles like p_T, tracks, energy, missing E_T and many others.
Additional information about the generated particles, even those that cannot be measured by real detector, is stored. This information is defined in a separate file, Delphes then stores the required information to a ROOT-file for an analysis with the ROOT framework.

2.4 Analysis

The analysis is performed within the ROOT framework [14]. The ROOT is an object oriented C++ framework used in high energy physics. It is designed for petabytes of data analysis in the most effective and fastest way possible. The data are stored in ROOT-files, for the most effective and fast reading and lowest disk size. ROOT framework contains predefined statistical functions, linear

algebra classes or numerical algorithms for fast and easy coding and analysis. The ROOT then can produce plots, provides an easy fitting tool and then exports to various output formats like PDF, JPG or GIF. Mostly, analyzers use three steps of the analysis: the main storage to a ROOT-file, the analysis storage of analyzed histograms to a ROOT-file and then the printing of histograms.
All the original Figures in this thesis with the histograms are produced within the ROOT framework. In the analysis a default 200 GeV p_T cut for large-R jet was set, because the lower p_T is not even generated by MadGraph. This selection is used for MadGraph study of the top quark pairs kinematics in this thesis.

2.4.1 Selection

Motivation for the p_T cuts comes from the Equation 2.1. If we are want to select boosted top quark jets, we need a larger p_T. Then with the increasing p_T cut the lower-mass jets and background do not survive the selection. The proof that the higher p_T cuts help us to get a clearer signal region is in the Figures 2.1 and 2.2. There are visible the W boson peaks at 80 GeV and the top quark peaks around 172 GeV.
In Figure 2.1 there is the large-R jet invariant mass after different p_T cuts on the large-R jet. Figures 2.1 and 2.2 show why at boosted region high p_T cuts are used, because they promote top jets. We can them later cut-off the W boson jets peaking at $\sim$80 GeV. The leading jet p_T have the highest p_T in the event and the sub-leading jet have the second highest p_T in the event.

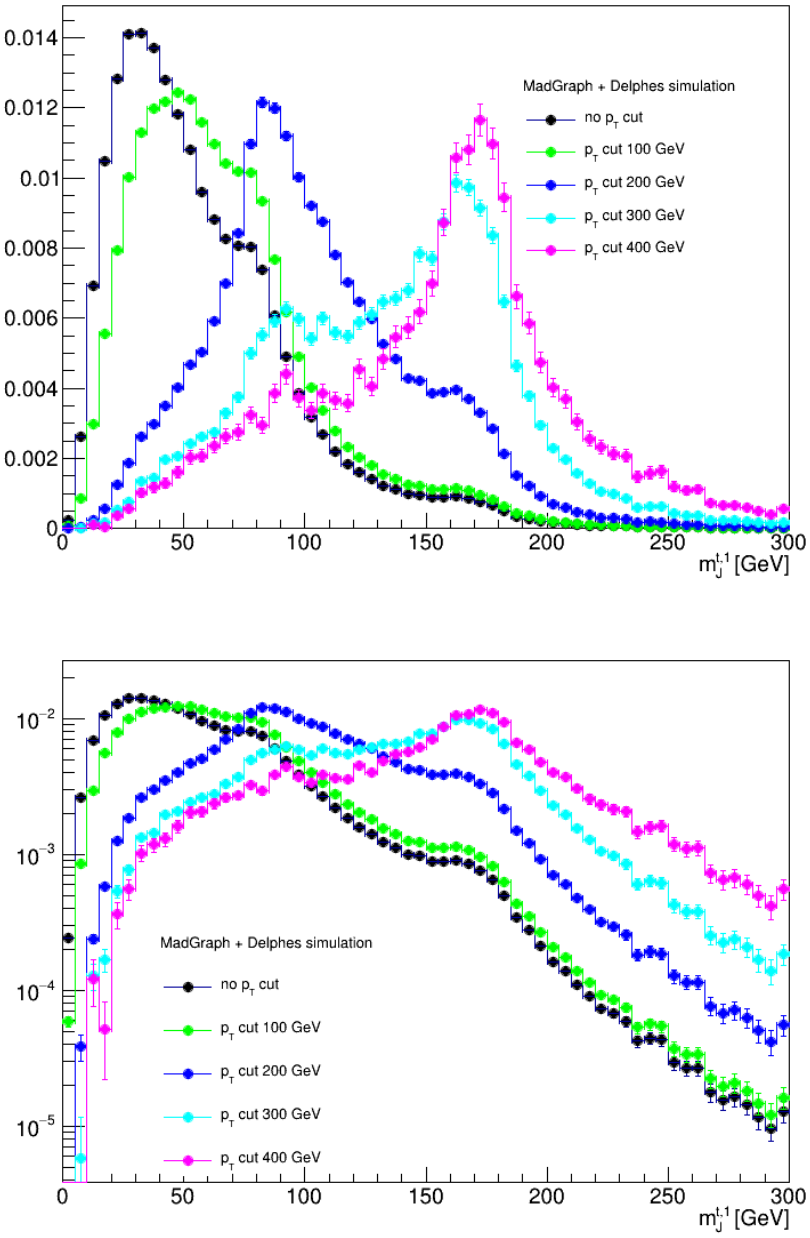

Figure 2.1: Leading jet mass shape in log and non-log scale. Normalized spectra plotted for various p_T cuts on the leading top large-R jet from the Monte-Carlo $t\bar{t}$ all-hadronic sample generated by MadGraph, setup in Chapter 2.1.

A visible peak shape shows two things, first one is that with higher cuts we have a cleaner Monte-Carlo $t\bar{t}$ signal sample, the second one is that the higher cuts also remove a large part of signal, that means there is a lower statistics.

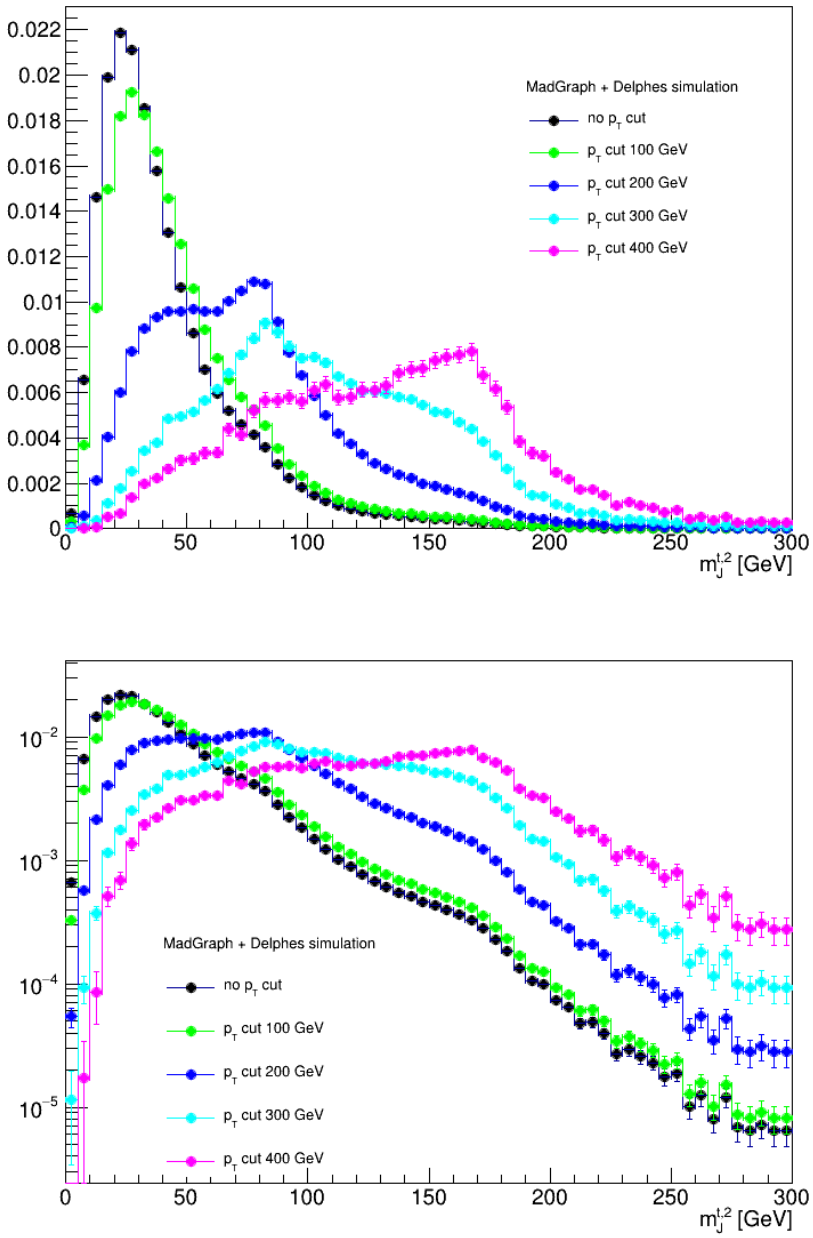

Figure 2.2: Subleading jet mass shape in log and non-log scale. Normalized spectra plotted for various p_T cuts on the sub-leading top large-R jet from the Monte-Carlo $t\bar{t}$ all-hadronic sample generated by MadGraph, setup in Chapter 2.1.

This study of the mass peaks structure of normalized histograms ratings choice of the p_T cut, is a compromise between the signal purity and statistic. The p_T cut is the main cut in the analysis and defines the phase space of the detector and the particle level.

2.4.2 Variables

The Monte-Carlo uses the mathematically well-defined theory (mostly the Standard model) as a model for physics processes event generation. One has to choose variables for data and Monte-Carlo comparison because while one variable could fit the data perfectly, another variable at the same time could be worse. The variables choice is explained in Chapter 3.

For the particle identification, we need to know both momentum and energy informations included in the basic equation (with $c = \hbar = 1$)

$$E^2 = p^2 + m^2, \tag{2.1}$$

The main basic spectra p_T, $m_{t\bar{t}}$, $p_T^{t\bar{t}}$, $y_{t\bar{t}}$, m_J and other, are used for the theory and data comparison. The Monte-Carlo generator works with the theoretical models, but there could be some differences from the real measured data, then some of the free parameters of the MC generator must be tunned for better agreement with the data. The analysis is made for the data and the MC predictions. Firstly we measure the data, than we setup the MC generator for the parameters what we need, the central-mass energy $\sqrt{s}$, the pp collision for the LHC and the ATLAS detector. When we have data and MC predictions we use the same code for the analysis. The final comparison is the ratio of the MC prediction and data for each spectrum, that ratio should be 1 within the experimental errors if the theory is correct. There are many aspects why the ratio will never be a 1, the main reason is the resolution of the detector or missing dynamics in the theoretical models. If the MC over data are in a within the statistical and systematical limits we could say that the theory is consistent with data.

Transverse momentum p_T

The transversal momentum p_T is one of the most important measured variable because $p_T = \sqrt{p_x^2 + p_y^2}$ we measure directly and it is well defined in theoretical models. It is important for the theory and data comparison and also searching for a new physics where the top quark is a background.

In Figure 2.3 there is a side view of a reconstructed event in the ATLAS detector, where in the x-y plane the $p_T = (p_x, p_y)$ is measured, from known tracks (blue lines usually start from the center in the Figure 2.3) using the well-known magnetic field inside the detector. The lines and the information about energy from the calorimeters (dark blue rectangles in Figure 2.3) are used to identify and define jets. The examples of the p_T distributions is in Figure 2.4.

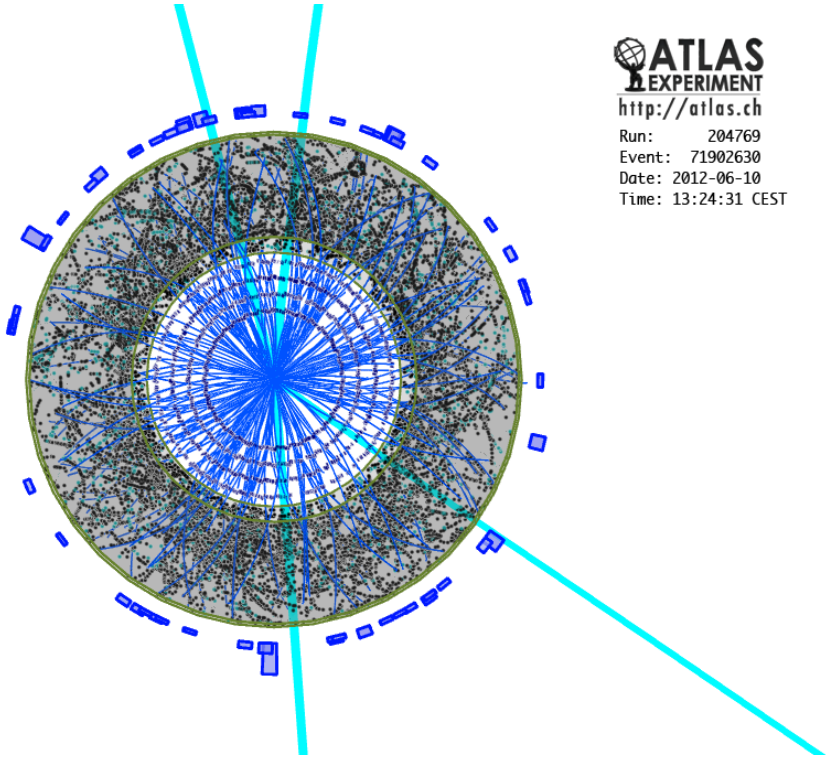

Figure 2.3: An example of the ATLAS detector side event view. Edited from [6].

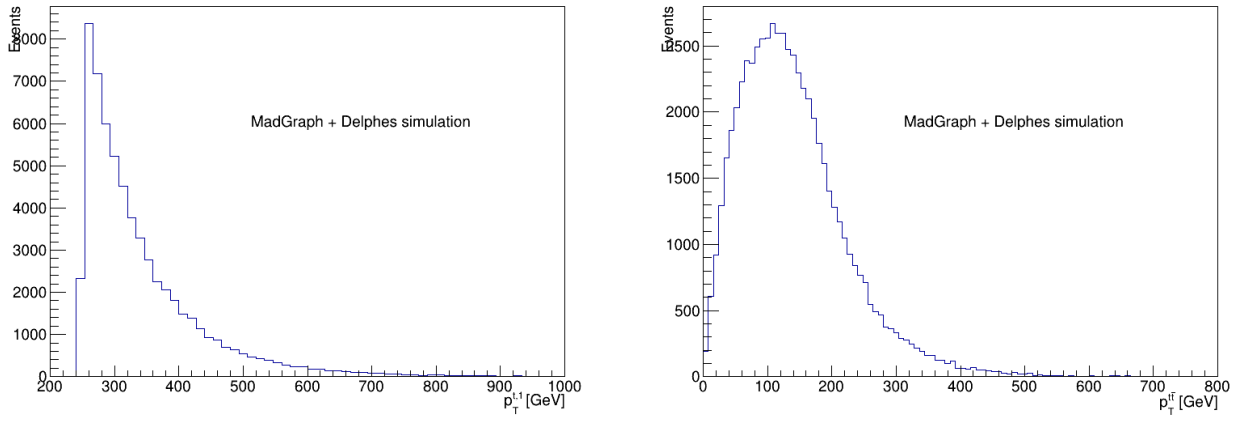

Figure 2.4: The shape comparison between $p_T^{t\bar{t}}$ of the $t\bar{t}$ system (right) and leading large-R jet $p_T^{t,1}$ (left) Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

The difference of the slopes of the two spectra in Figure 2.4 is clear, because the leading large-R jet $p_T^{t,1}$ distribution have the exponential behavior starting at 200 GeV, because there is a p_T cut in the generator. The $p_T^{t\bar{t}}$ has a shape of Poisson distribution at lower starts from zero and peaked around 100 GeV.

Invariant mass m

Invariant mass is a Lorentz transformation invariant parameter. It is an important spectrum for a large-R jet, m_J, but the most important is the invariant mass of the $t\bar{t}$ system, $m_{t\bar{t}}$, where we are searching potentially new particles beyond-the-standard-model (BSM) decaying to a $t\bar{t}$ pair. The most famous BSM theory is Supersymetry (SUSY). If there is a new particle decaying to $t\bar{t}$ pair, we would see a peak corresponding to its mass over continuum.

We can measure the energy of lepton (except muon or neutrino), in the ATLAS detector with the electromagnetic calorimeter and from tracking detectors we can compute a transversal momentum and because there are segmented calorimeters, we know the angle. With this informations and with the Equation 2.1, we find

$$p^2 = p_T^2 + p_z^2, \tag{2.2}$$

$$p^2 = p_T^2 + p^2 \cos^2\theta, \tag{2.3}$$

and with simple algebraic modifications there is a final equation for invariant mass of a particle

$$m = \sqrt{E^2 - \frac{p_T^2}{\sin^2\theta}}. \tag{2.4}$$

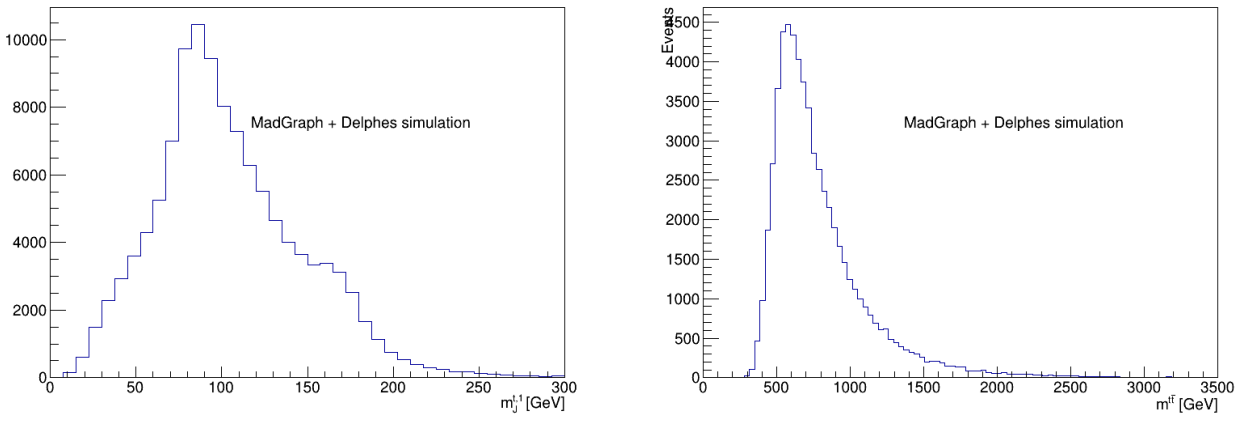

Figure 2.5: The leading large-R jet mass $m^{t,1}$ (left) and mass $m_{t\bar{t}}$ of the $t\bar{t}$ (right) with p_T cut 200 GeV Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

The invariant mass of the jet or the $t\bar{t}$ system is important in the search for a new physics and a new particles. If there are a new particle, there will be visible a peak. Peak at the invariant mass, most probably Gaussian or Poisson in shape, has some parameters and from the fit of this peak we can extract additional information about the particle like lifetime or Γ, although also affected by experimental resolution.

Rapidity y

The rapidity is a dimensionless variable, which is a measure of longitudinal space movement (along the z-axis). The rapidity is defined as

$$y \equiv \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{2.5}$$

where the E is the Energy and p_z is a component of $|p| = (p_x, p_y, p_z)$ along the beam the z axis. Motivation for studying the rapidity is the data and Monte-Carlo cross check. We also know that the rapidity must be symmetric around zero. If it is not we know that there is some issue like bad selection or reconstruction.

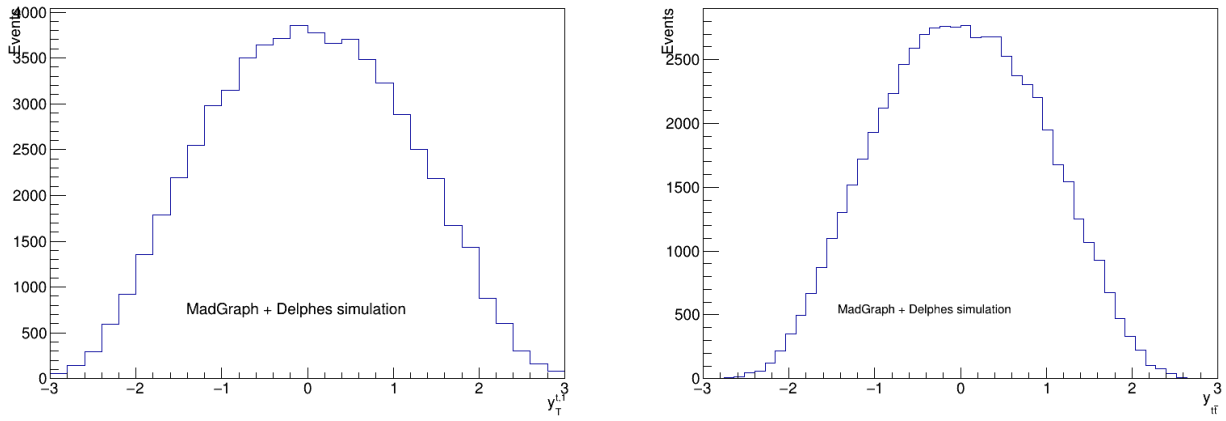

Figure 2.6: The rapidity $y_{t\bar{t}}$ of the $t\bar{t}$ system (right) and leading large-R jet $y^{t,1}$ (left) Monte-Carlo generated with MadGraph without analysis cuts, setup in Chapter 2.1.

The $y_{t\bar{t}}$ and $y^{t,1}$ distributions are usually of the Gaussian form.

Cos θ*

The θ^* is the angle between the jet and the z axis in the central mass system (C. M. S.). Let's have a simple process $q\overline{q}$ to a $t\bar{t}$ as in Figure 2.7. The θ^* angle can be used for identification which processes displayed in Figure 1.2 are dominant.

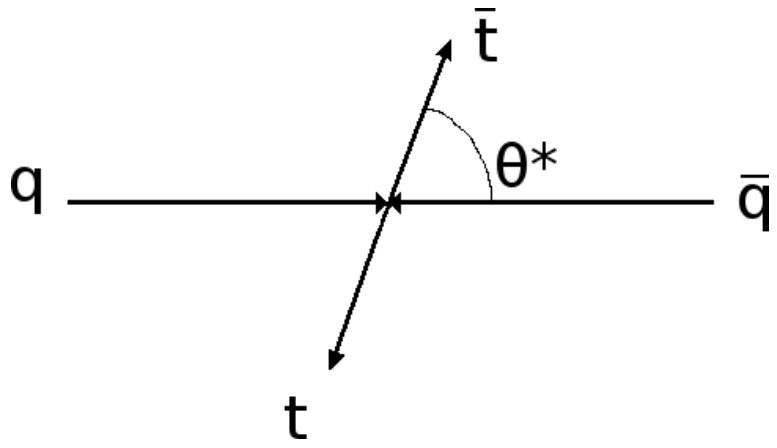

Figure 2.7: The $q\overline{q}$ to $t\bar{t}$ process kinematic at the C. M. S.

The Feynman diagram for this process is in Figure 1.2 (a). The whole energy of the collision is

$$\sqrt{\hat{s}} = E_q + E_{\overline{q}} = E_t + E_{\bar{t}}, \tag{2.6}$$

where $\hat{s}$ is one of the Mandelstam variables [10]. There are three such: the s, t, and u, they are invariant against the Lorentz transformation. The s-channel productions are in Figure 1.2 (a) and (b), there are also t-channel in Figure 1.2 (c) and (d).

The real data are not measured at C. M. S., usually the events are boosted along the z (beam) axis and consequently the E_q from the Equation 2.6 does not have the same value as the $E_{\overline{q}}$. From the histogram of $\cos\theta^*$ one could deduce what process is dominant (the s-channel or the t-channel). The $|\cos\theta^*|$ distribution is displayed in Figure 2.8. The $|\cos\theta^*|$ distribution are different for each of the channels (s, t, u).

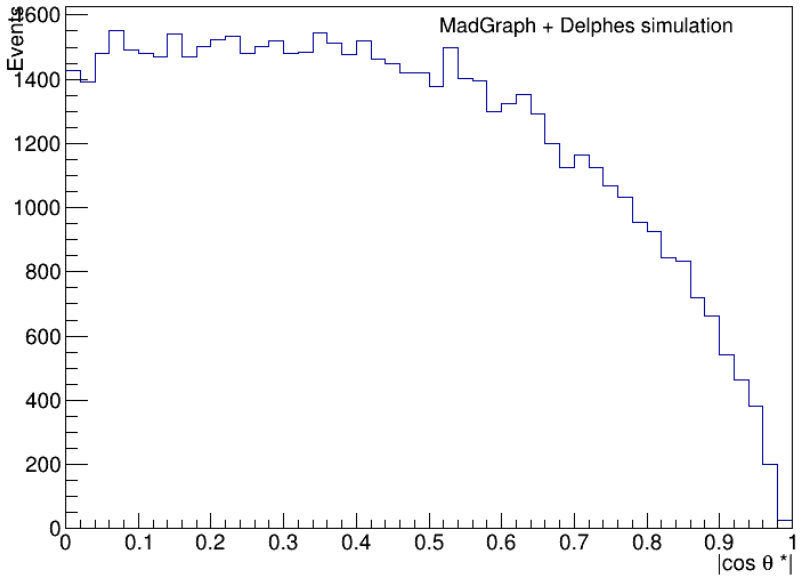

Figure 2.8: The $|\cos\theta^*|$ distribution of the $t\bar{t}$ signal sample generated by Monte-Carlo with MadGraph, setup in Chapter 2.1.

Jet substructure

Jet substructure is important for the top-tagging and for the signal region selection as it is shown in Figure 2.9. The τ_N is a N-subjettines variable, when we have a large-R jet with anti-k_t $R = 1.0$, then there is a search for sub-jets with anti-k_t $R = 0.4$. The τ_N is how much N sub-jets are in the large-R jets event then for the event selection is then important the ratio $\tau_{XY} = \tau_X/\tau_Y$ on which cuts are used to cut-off the background and to have cleaner top jets.

The cuts on the τ variables are important and used in the analysis, because they allows better separation of signal from background. Specifically the $\tau_{32} = \tau_3/\tau_2$ is the ratio of a τ with three sub-jets and a τ with two sub-jets and similarly is for the τ_{21}.

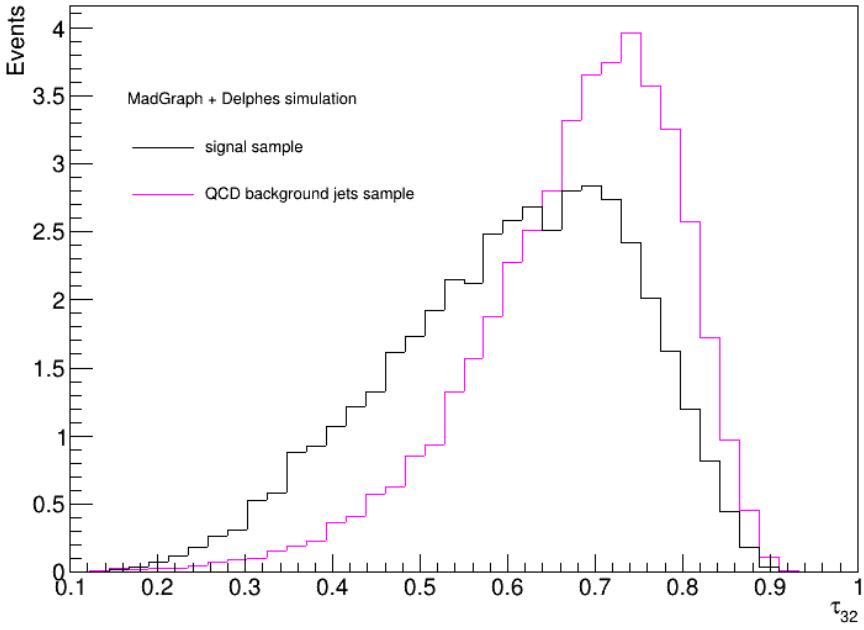

Figure 2.9: The distributions of the $\tau_{32}^{t,1}$ variable for the leading large-R jet with the $p_T^{t,1}$ cut of 400 GeV on the jets for the QCD background sample (purple) and the signal sample (black) Monte-Carlo $t\bar{t}$ signal generated with MadGraph, setup in Chapter 2.1.

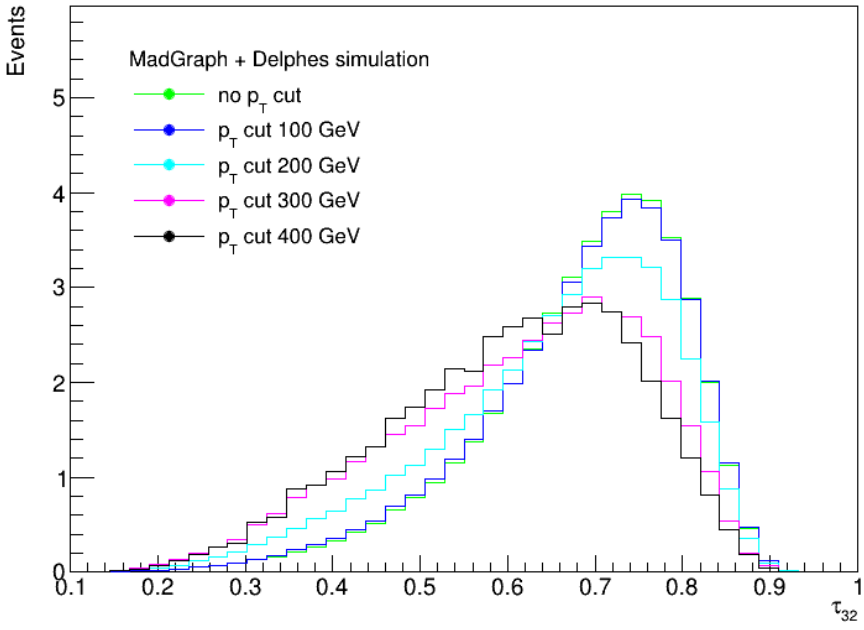

Figure 2.10: The normalized distribution of $\tau_{32}^{t,1}$ of the leading large-R jet with the different $p_T^{t,1}$ cuts (100, 200, 300 and 400 GeV) on the signal sample Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

With the cuts on the τ_{32} variable, we could cut-off the background and get a clearer signal region. The population changes with the p_T cuts is obvious in Figures 2.11 and 2.12.

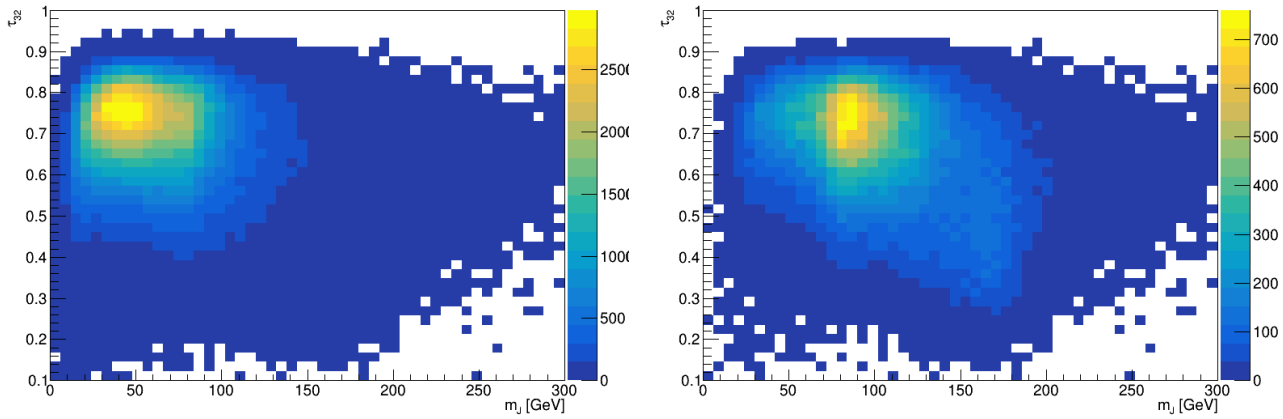

Figure 2.11: The leading large-R jet $\tau_{32}^{t,1}$ vs the mass of the leading top jet $m_J^{t,1}$ with $p_T^{t,1}$ cuts (left) of 100 GeV and 200 GeV (right), Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

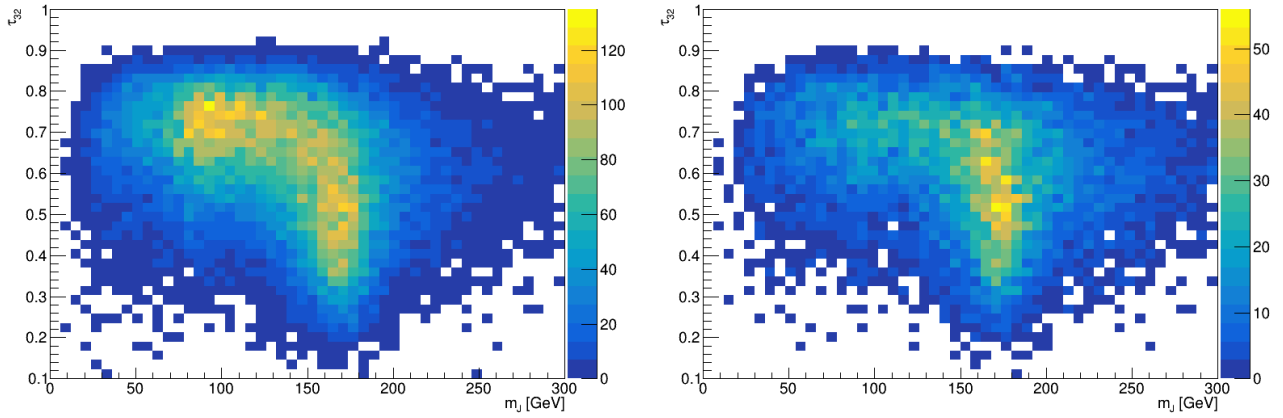

Figure 2.12: The leading large-R jet $\tau_{32}^{t,1}$ vs the mass of the leading top jet $m_J^{t,1}$ with $p_T^{t,1}$ cuts (left) of 300 GeV and 400 GeV (right), Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

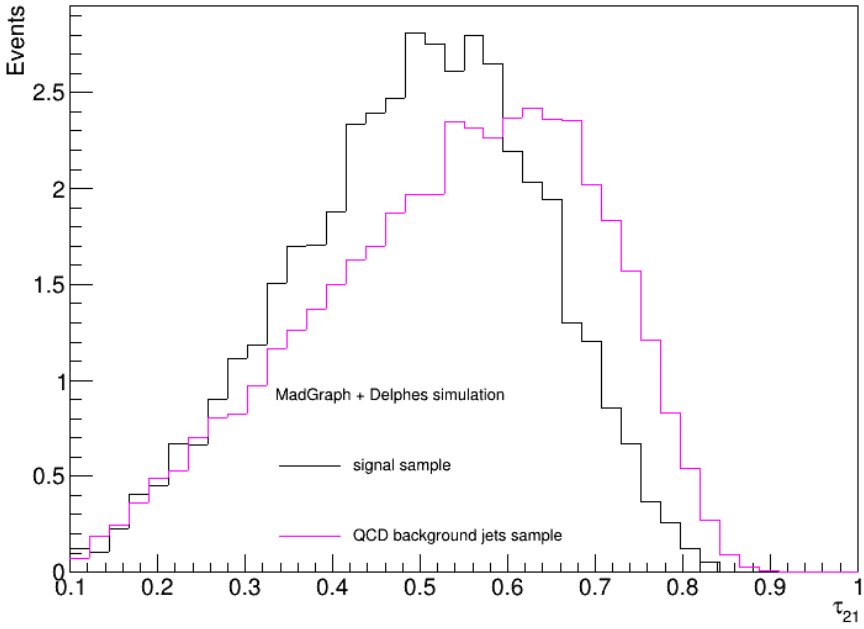

Figure 2.13: The distributions of $\tau_{21}^{t,1}$ of the leading large-R jet with the $p_T^{t,1}$ cut of 400 GeV on the jets only for the QCD background sample (purple) and the signal sample (black) Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

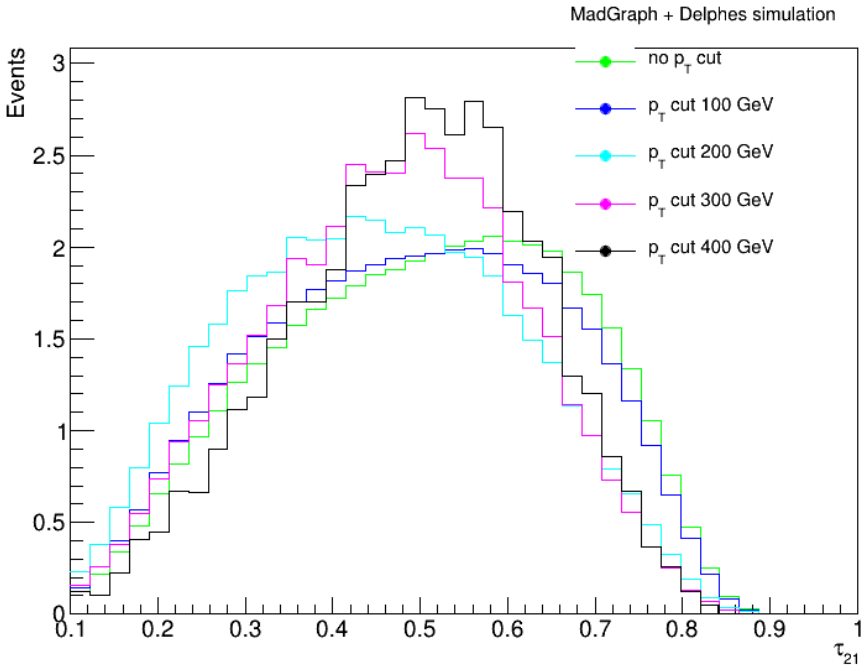

Figure 2.14: The normalized distribution of $\tau_{21}^{t,1}$ of the leading large-R jet with the different $p_T^{t,1}$ cuts (100, 200, 300 and 400 GeV) on the signal $t\bar{t}$ sample generated with MadGraph, setup in Chapter 2.1.

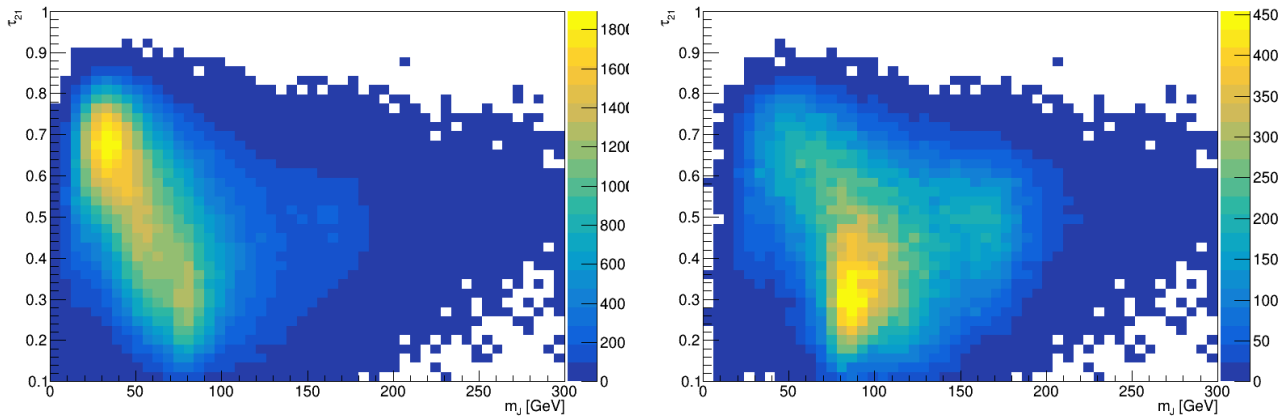

Figure 2.15: The leading large-R jet $\tau_{21}^{t,1}$ vs the mass of the leading top jet $m_J^{t,1}$ with $p_T^{t,1}$ cuts (left) of 100 GeV and 200 GeV (right), Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

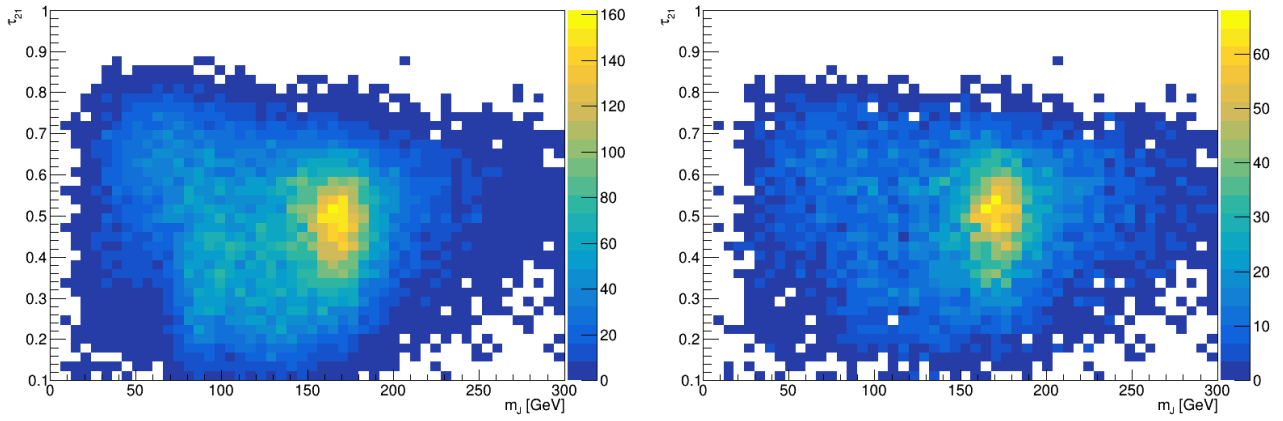

Figure 2.16: The leading large-R jet $\tau_{21}^{t,1}$ vs the mass of the leading top jet $m_J^{t,1}$ with $p_T^{t,1}$ cuts (left) of 300 GeV and 400 GeV (right), Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

The difference between the τ_{32} and τ_{21} is visible in the Figure 2.12 on the right and the Figure 2.16 on the right. The top quark, with the same p_T cut, at τ_{32} is peaking around 0.5 at Y axis and 172 GeV at X axis meanwhile at the τ_{21} is the top quark peaking at 0.45 at Y axis and 172 GeV at X axis. In contrast, the W peak is around the 80 GeV in the τ_{32} distributions is peaking around the 0.7, the W peak is around the 80 GeV in τ_{21} distributions at the 0.3. This difference of the top quark peak and W boson peak for the τ cuts study is used below. In the ATLAS analysis τ_{32} cut is mostly used but with the cut on τ_{32} and τ_{21} it would be possible to get even clearer signal region without background from jets not originating from top quarks.

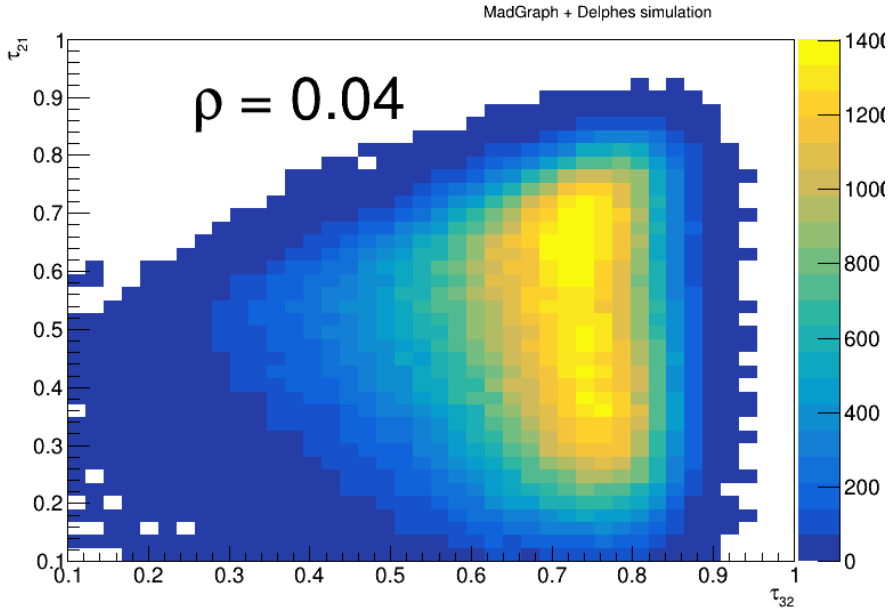

Figure 2.17: The leading large-R jet $\tau_{21}^{t,1}$ vs $\tau_{32}^{t,1}$ with $p_T^{t,1}$ cuts of 100 GeV Monte-Carlo generated with MadGraph, setup in Chapter 2.1. The ρ is a correlation coefficient.

Pout

The $\left|p_{out}^{t\bar{t}}\right|$ variable is the absolute value of the out-of-plane momentum, the top quark vector projection to the plane defined by the another top quark (anti-top) in the same event and the beam z axis, this describe the following Equation 2.7 from [16]

$$\left|p_{out}^{t\bar{t}}\right| = \left|p^{t,1} \cdot \frac{p^{t,2} \times \hat{z}}{|p^{t,2} \times \hat{z}|}\right|. \tag{2.7}$$

The geometrical illustration of the $\left|p_{out}^{t\bar{t}}\right|$ variable is shown on the Figure 2.18.

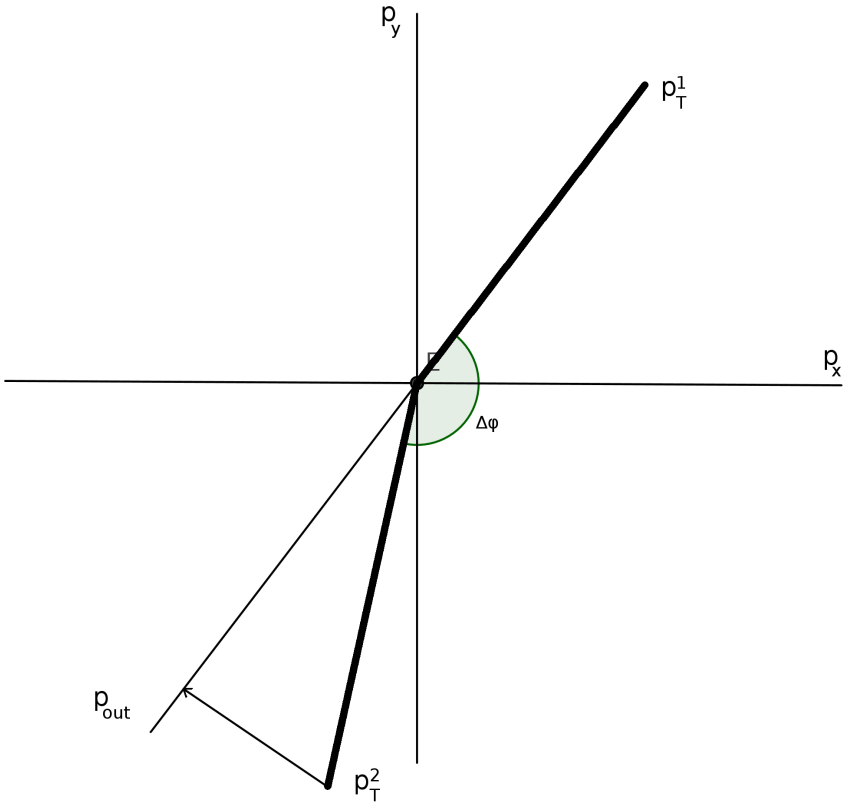

Figure 2.18: The $p_{out}^{t\bar{t}}$ at the transversal, x-y plane, schema.

This variable is sensitive to p_T imbalance in the transverse plane, to the emission of radiation connected with the $t\bar{t}$ production. The angle between the top anti-top pair is sensitive to a hypothetical new particles in the t-channel and also radiation.

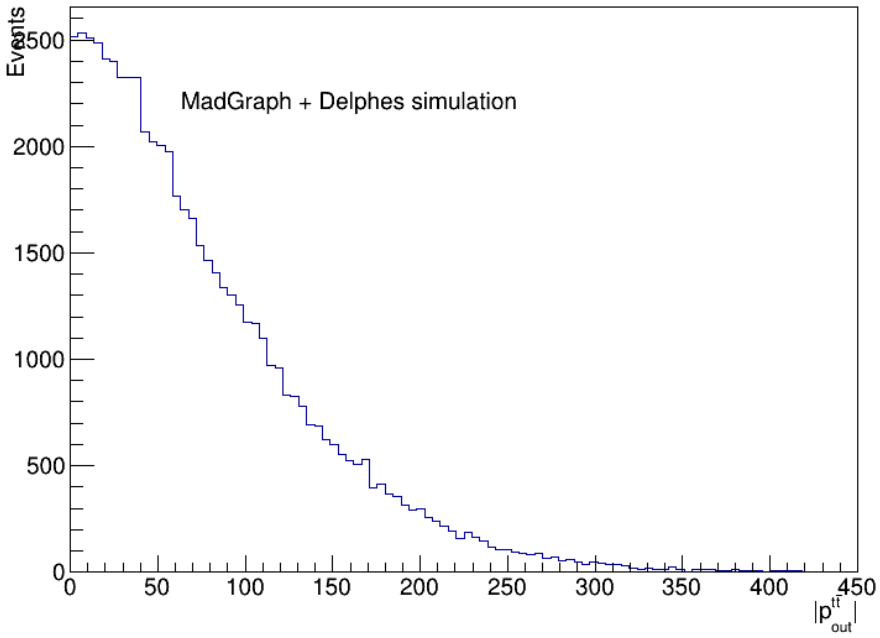

Figure 2.19: The $p^{t\bar{t}}_{out}$ on the Monte-Carlo generated with MadGraph without analysis cuts, setup in Chapter 2.1.

The jet substructure cuts study

This study is for the τ_{32} and τ_{21} cuts, it shows how this cut can improve analysis and make the signal region clearer. In Figure 2.20 it is obvious the purpose of the τ_{XY} cuts, the signal region becomes clearer, top mass peak more visible although with lower statistics. There is shown that the τ_{32} cut reduces the statistics at all distributions almost evenly, but the τ_{21} cut extremely reduces the W mass peak at 80 GeV. After both cuts there are only the top quark peak at 172 GeV which we desire.

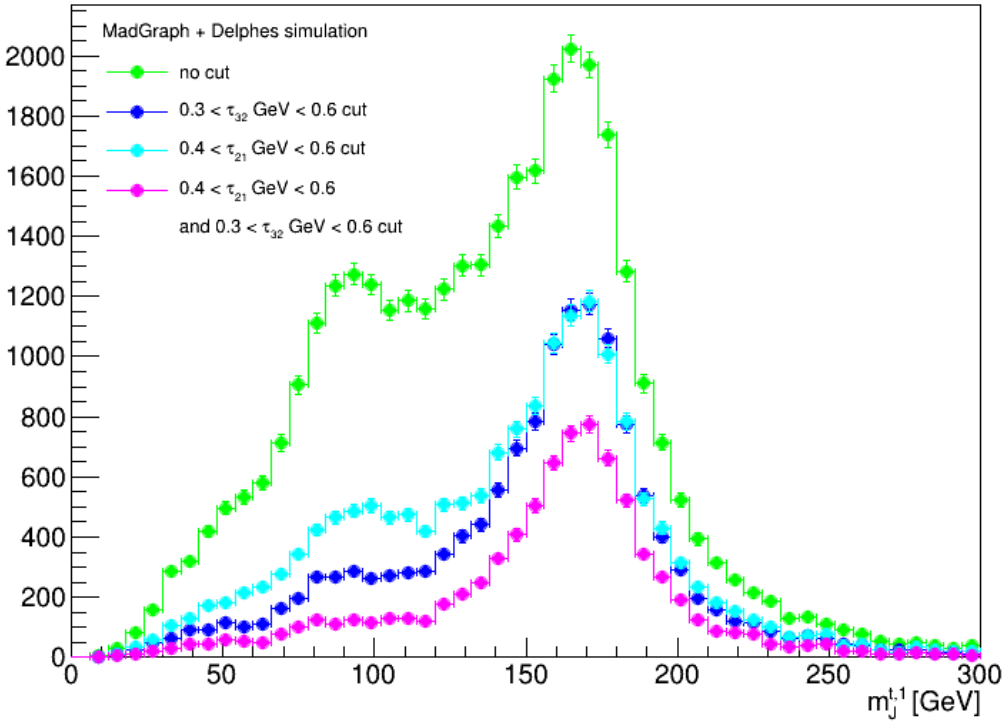

Figure 2.20: The leading large-R jet mass $m_J^{t,1}$ with τ_{32} and τ_{21} cuts and with p_T cut on 300 GeV, signal Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

Leading and sub-leading jets

The event with the $t\bar{t}$ and the two reconstructed large-R jets can be divided to a leading jet with the larger p_T and a second one the sub-leading jet. The p_T distributions of these two are primarily used for the theoretical model and real measured data comparison. There are visible some differences in the Figures 2.21 and 2.22. The main difference is in the masses of the leading and sub-leading jets. In Figure 2.22 (left) the leading jet $m_J^{t,1}$ there is a visible large W boson mass peak and a smaller top quark peak, except the sub-leading jet $m_J^{t,2}$ there is visible the continuum and the W peak but the top quark peak is almost insignificant indicating that the subleading top jet is less boosted as expected. In the Figure 2.23 there is not visible difference in the shape of rapidities.

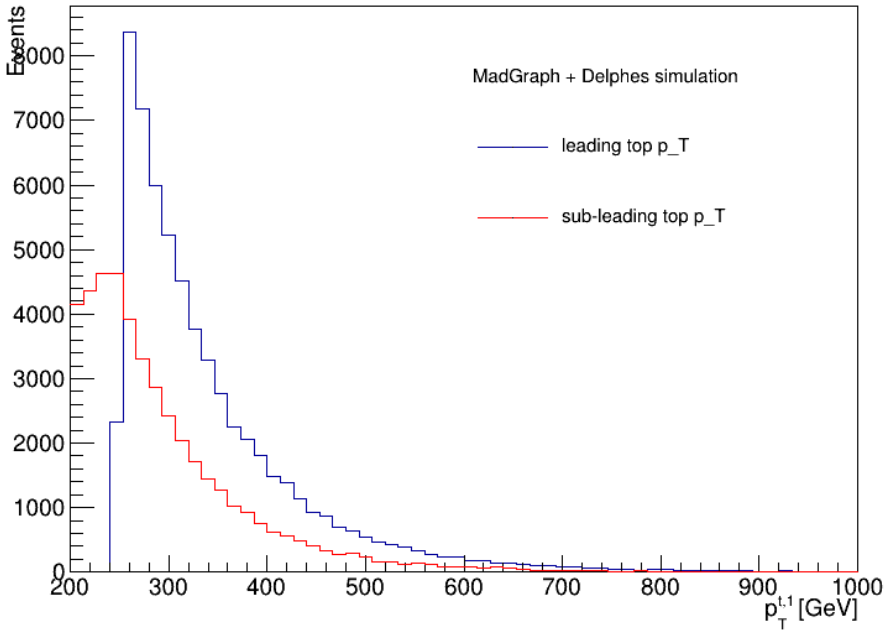

Figure 2.21: The leading top large-R jet $p_T^{t,1}$ (blue) and sub-leading top large-R jet $p_T^{t,2}$(red) for the signal Monte-Carlo generated with MadGraph without cuts except the setup in Chapter 2.1.

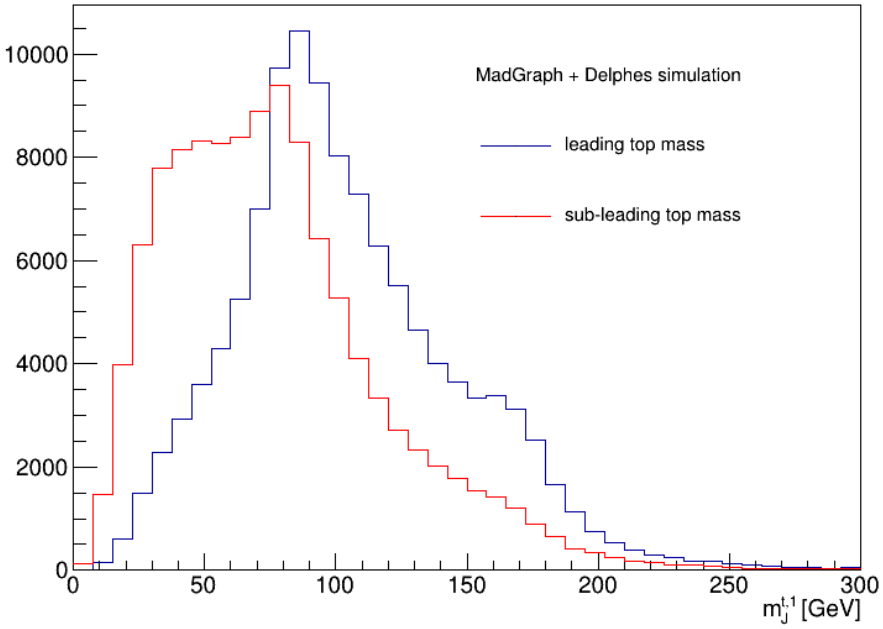

Figure 2.22: The leading top large-R jet $m_J^{t,1}$ (blue) and sub-leading top large-R jet $m_J^{t,2}$(red) for the signal Monte-Carlo generated with MadGraph without cuts except the setup in Chapter 2.1.

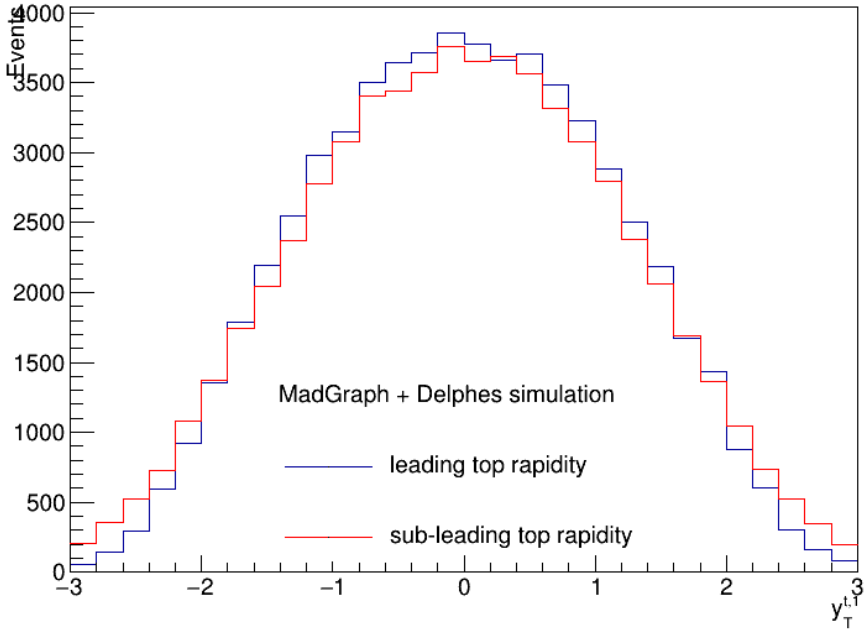

Figure 2.23: The leading top large-R jet rapidity $y_J^{t,1}$ (blue) and sub-leading top large-R jet rapidity $y_J^{t,2}$(red) for the signal Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

After the cut studies we chose the following cuts $0.6 > \tau_{32} > 0.3$, $0.6 > \tau_{21} > 0.4$, $p_T^{t,1} > 400$ GeV and $p_T^{t,2} > 300$ GeV to select top quark jets in the MadGraph and Delphes analysis. The final results are below in Figures 2.24, 2.25 and 2.26.

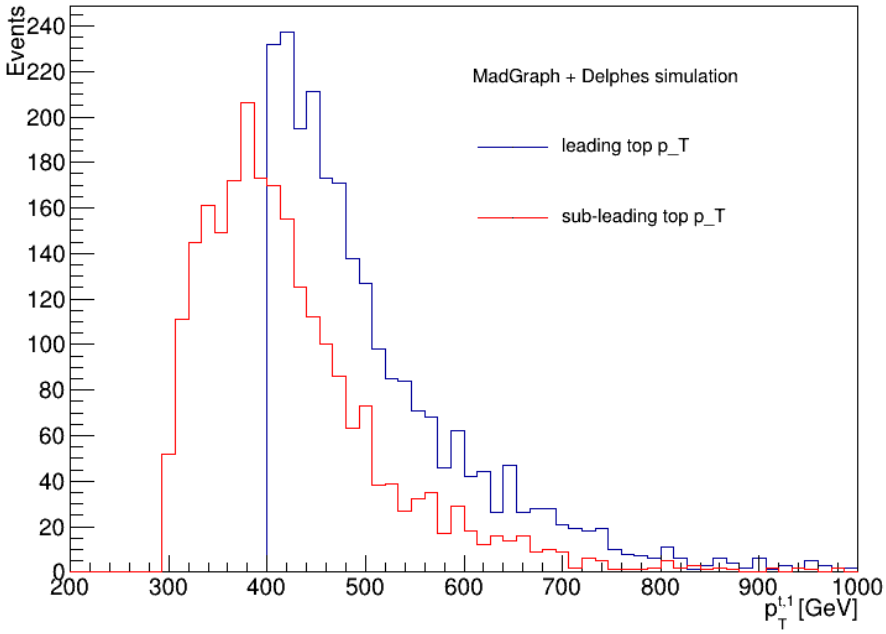

Figure 2.24: The leading top large-R jet $p_T^{t,1}$ (blue) and sub-leading top large-R jet $p_T^{t,2}$ (red) for the signal Monte-Carlo generated with MadGraph after cuts, the setup in Chapter 2.1.

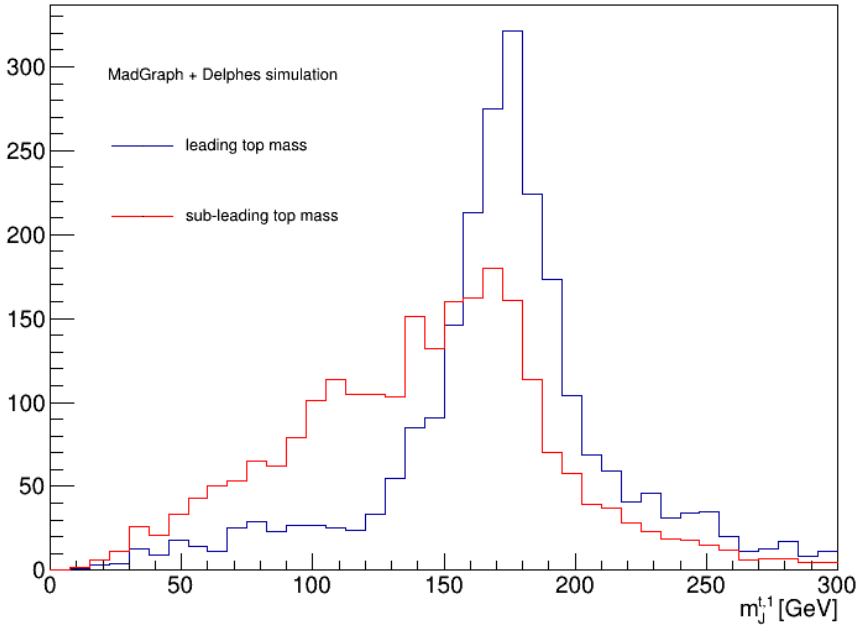

Figure 2.25: The leading top large-R jet $m_J^{t,1}$ (blue) and sub-leading top large-R jet $m_J^{t,2}$(red) for the signal Monte-Carlo generated with MadGraph after cuts, the setup in Chapter 2.1.

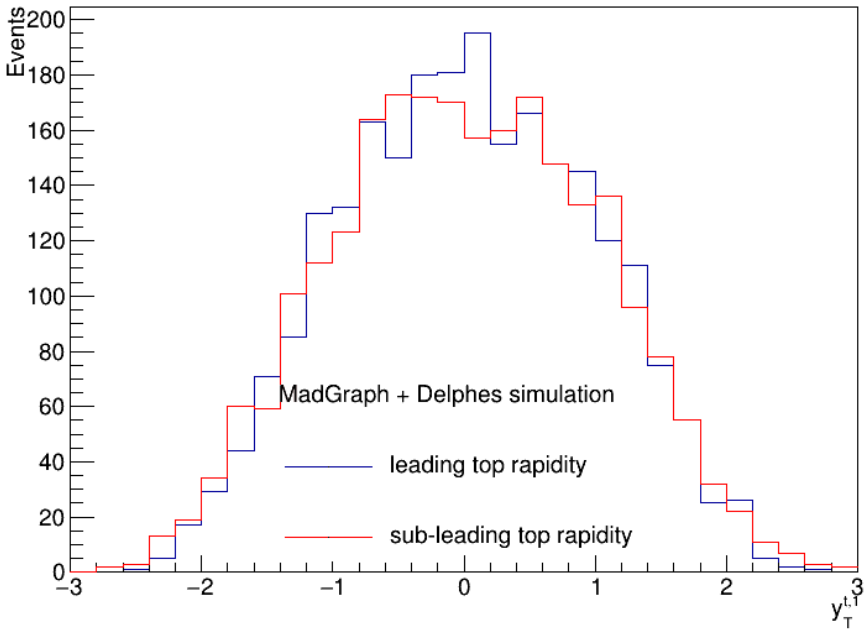

Figure 2.26: The leading top large-R jet $y_J^{t,1}$ (blue) and sub-leading top large-R jet $y_J^{t,2}$ (red) for the signal Monte-Carlo generated with MadGraph after cuts, the setup in Chapter 2.1.

In the comparison of the spectra without cuts and after them a big decrease of statistic is obvious as expected but also we get clearer signal region made of mostly the top quark jets as it is visible in Figure 2.25.

The top anti-top system

The $t\bar{t}$ system is a sum of the top quark and anti-top quark four-vectors of the large-R jets in one event. There possibly could be a new particle, which decays to $t\bar{t}$, this hypothetical particle is called the *Z prime* (Z'). The main motivation for analysis of the all-hadronic channel with boosted topology is the searches for the beyond-the-standard model particles.
In Figures 2.27, 2.28, 2.29 and 2.34 are displayed spectra without analysis cuts (blue) and with them (purple), there are also the normalized spectra for the shape comparison.

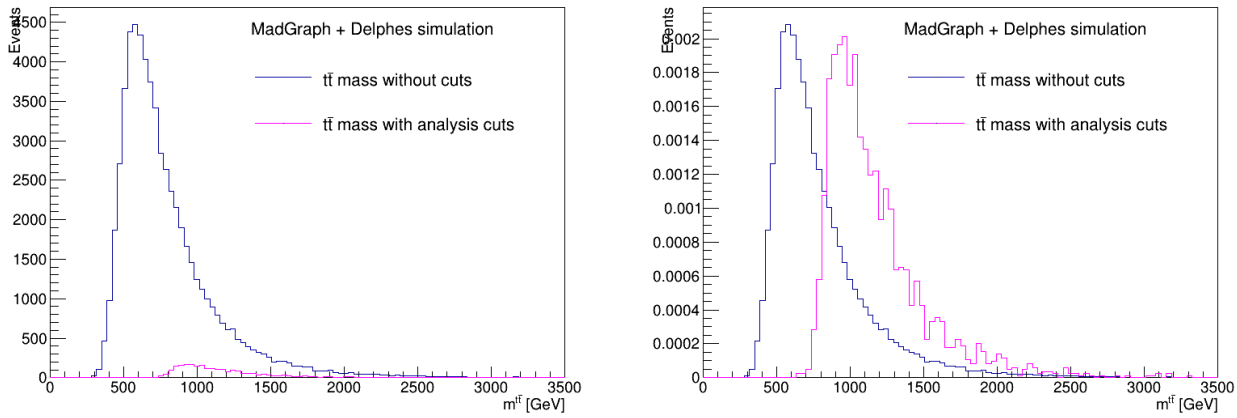

Figure 2.27: The mass of the $t\bar{t}$ system $m_{t\bar{t}}$ without analysis cuts (blue) and with analysis cuts (purple) as non-normalized (left) and normalized (right) distributions Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

Specially in Figure 2.27 is visible the shape to the larger mass of the $t\bar{t}$ system with the analysis cuts except in Figures 2.28, 2.29 and 2.34 is visible only the decrease of statistic. That means, if we are searching for the new particles we need higher p_T cuts to look for higher mass and also our chosen cuts are correct.

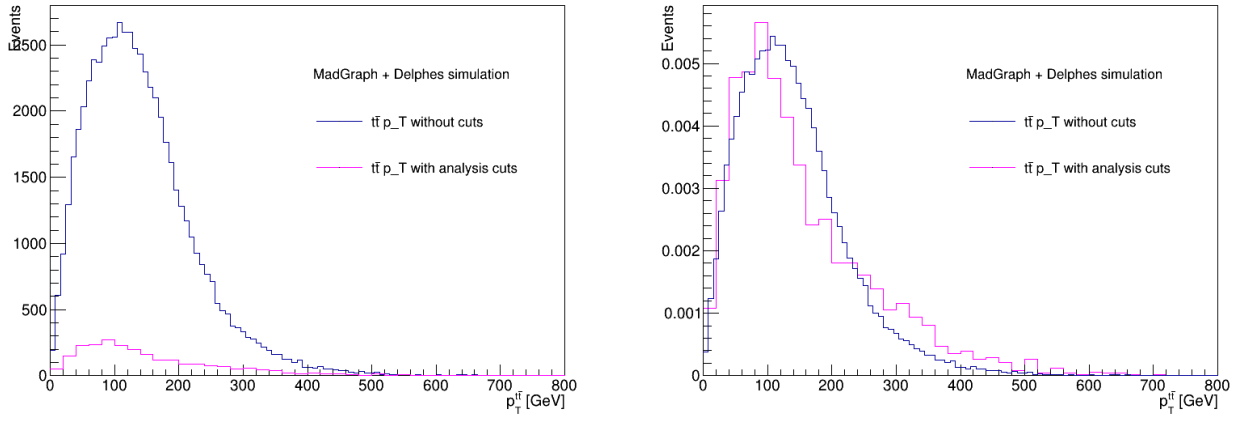

Figure 2.28: The transverse momentum of the $t\bar{t}$ system $p_T^{t\bar{t}}$ without analysis cuts (blue) and with analysis cuts (purple) as non-normalized (left) and normalized (right) distributions Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

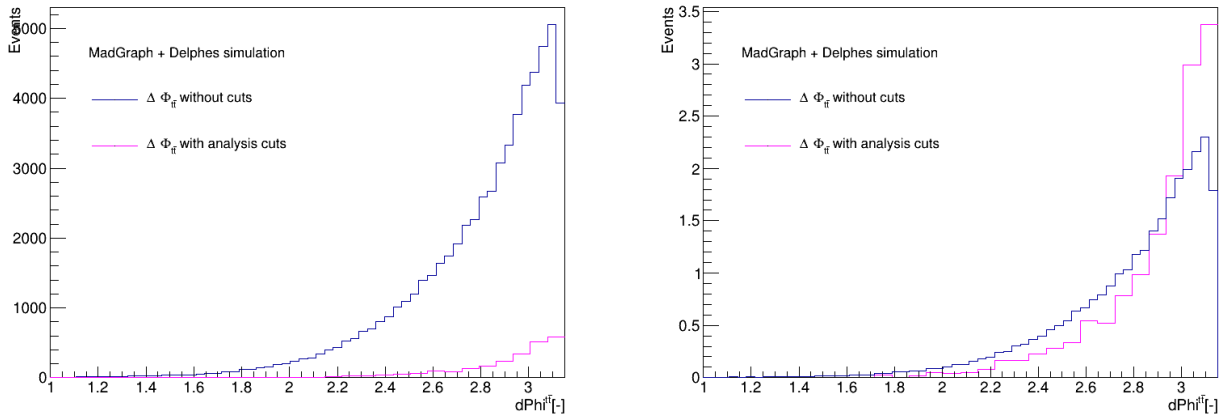

Figure 2.29: The $\Delta\phi$ of the $t\bar{t}$ system $\Delta\phi_{t\bar{t}}$ without analysis cuts (blue) and with analysis cuts (purple) as non-normalized (left) and normalized (right) distributions Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

The $\Delta\phi_{t\bar{t}}$ in Figure 2.29 is the angle between the two jets coming from the top quark and anti-top quark. The requirement of the large p_T cut makes the value of the $\Delta\phi_{t\bar{t}}$ mostly around π which means the two large-R jets to recoil against each other. Non-zero values of $\Delta\phi_{t\bar{t}}$ are due to extra emissions of gluons in the initial of final state. The $y_{t\bar{t}}$ in Figure 2.34 is distributed with Gaussian distribution around zero and is limited only by the detectors positions and gaps between the detectors segments.

Hypothetical new particle

Hypothetically, if there were a new BSM particle Z', we would see the significance in the invariant mass of the top quark pair. As a test the Z' model setup files [19] [18] were imported to MadGraph. The example of the comparison between $t\bar{t}$ signal samples and Z' samples for $m_{Z'} = 1500$ GeV and $m_{Z'} = 2000$ GeV normalized distributions are in Figure 2.30.

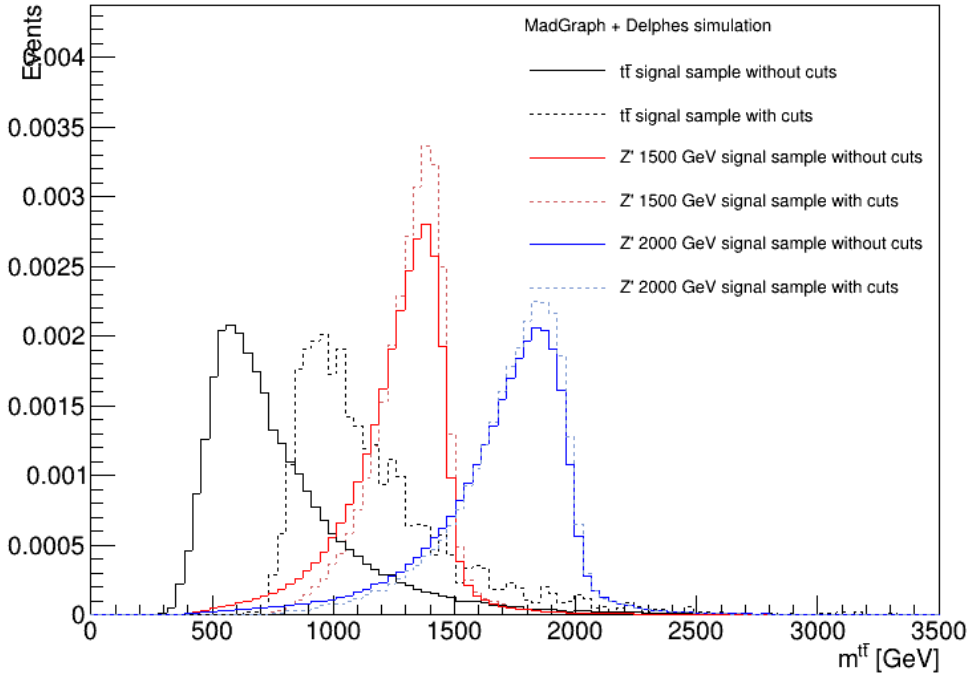

Figure 2.30: The mass of the $t\bar{t}$ system $m_{t\bar{t}}$ for the $t\bar{t}$ signal sample and two Z' samples of $m_{Z'} = 1500$ GeV and $m_{Z'} = 2000$ GeV, normalized distributions Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

After the combination of the $t\bar{t}$ sample and the two ad-hoc normalized Z' samples we can see what would the distribution look like if there were Z' particles of the two masses, in Figure 2.31.

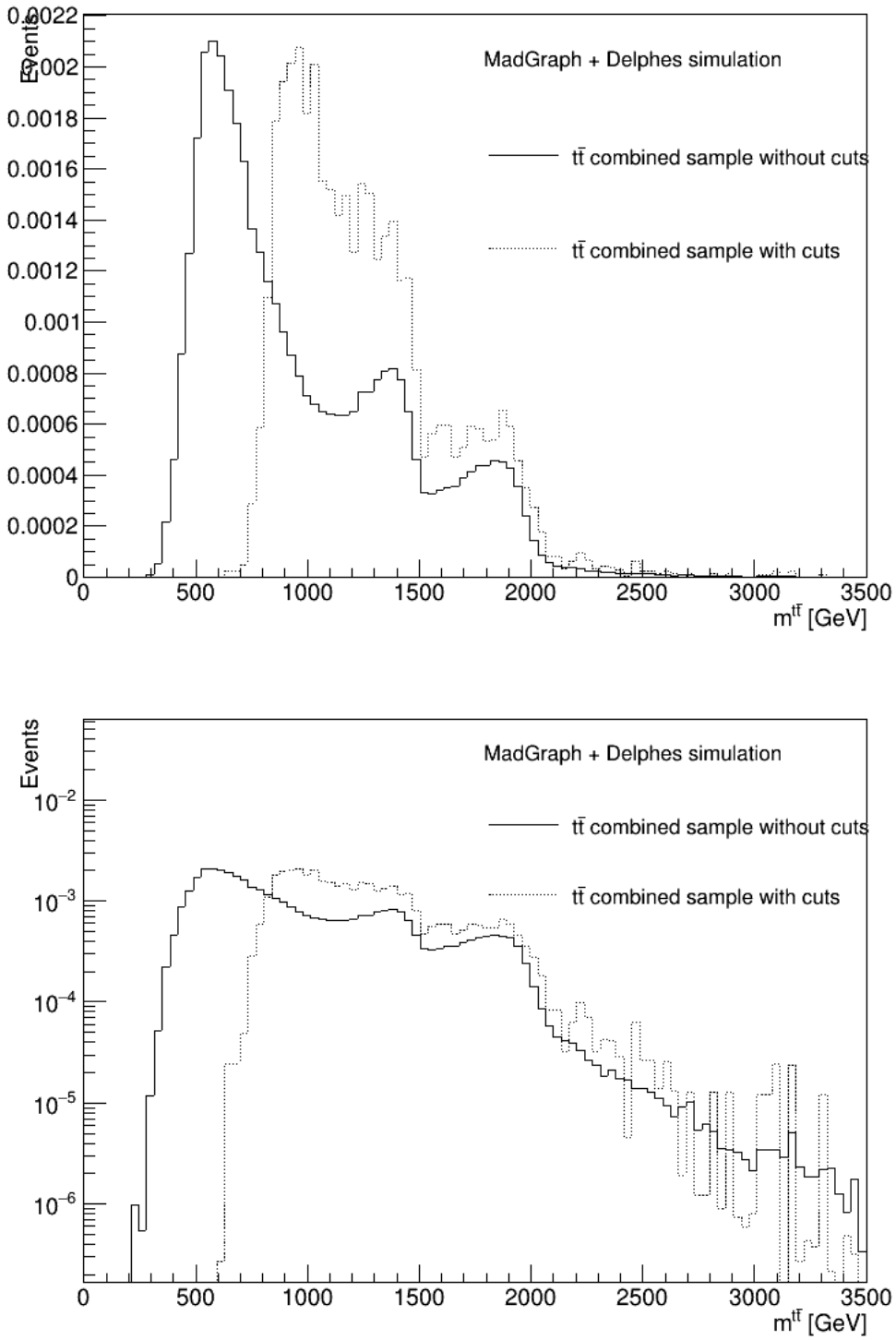

Figure 2.31: The mass of the $t\bar{t}$ system $m_{t\bar{t}}$ for the $t\bar{t}$ signal sample and two Z' samples of $m_{Z'} = 1500$ GeV and $m_{Z'} = 2000$ GeV without log (up) and with log scale (bottom) combined to one sample Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

The $H^{t\bar{t}}$ and the $z^{t\bar{t}}$ variables

The $H^{t\bar{t}}$ and $z^{t\bar{t}}$ variables are p_T based spectra combining the leading top jet p_T and sub-leading top jet p_T. They give us information about the distribution of the total p_T from the collision between these two jets. These spectra are important in the theory and data comparison, specially for the Monte-Carlo tuning. The $H^{t\bar{t}}$ and $z^{t\bar{t}}$ of the $t\bar{t}$ system displayed in Figure 2.32 are both p_T-related variables. Differences are obvious in the Equations 2.8 and 2.9 from [16],

$$H^{t\bar{t}} \equiv p_T^{t,1} + p_T^{t,2} \tag{2.8}$$

$$z^{t\bar{t}} \equiv \frac{p_T^{t,1}}{p_T^{t,2}} \tag{2.9}$$

where the $p_T^{t,1}$ is a leading top jet p_T and $p_T^{t,2}$ sub-leading top jet p_T.

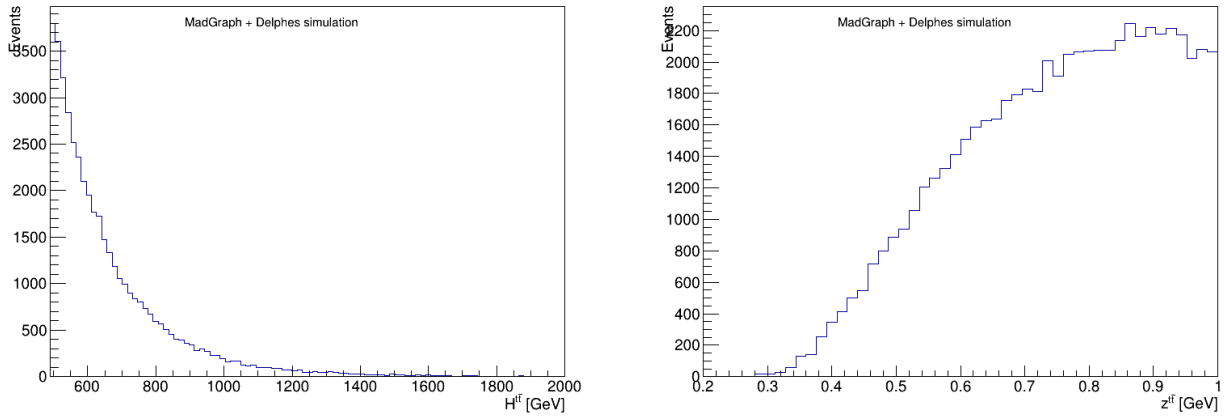

Figure 2.32: The $H^{t\bar{t}}$ (left) and the $z^{t\bar{t}}$ (right) for the signal Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

The χ of the $t\bar{t}$ system

From theoretical study it was obtained that the angular variables are sensitive to a p_T, *i.e.* to the emission of radiation with the top quark pair for the central region of production study is used. The angle between the two jets from the top quark pair is sensitive to hypothetical new particles in the t-channel. The $\chi^{t\bar{t}}$ is defined as [16]

$$\chi^{t\bar{t}} = e^{|y^{t,1} - y^{t,2}|}, \tag{2.10}$$

where the $y^{t,1}$ is the rapidity of the leading top jet and $y^{t,2}$ is the rapidity of the sub-leading top jet, both in the center-of-mass system, the example of the distribution is in Figure 2.33.

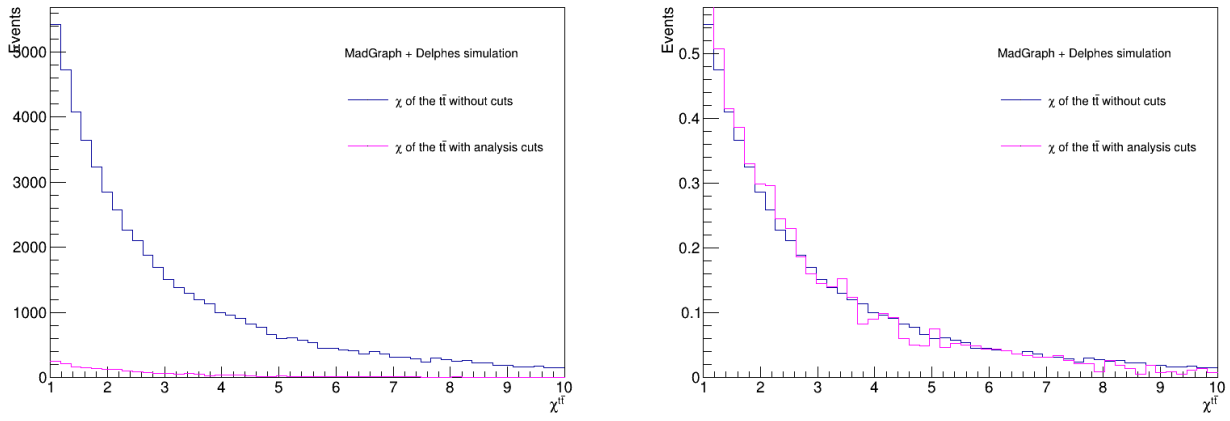

Figure 2.33: The non-normalized distribution (left) and normalized distribution (right) of $\chi^{t\bar{t}}$ without cuts (blue) and with analysis cuts (purple) Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

The number ρ is an assessment value for correlation between kinematic variables, the variables are correlated if the $\rho = 1$ and anti-correlated if the $\rho = -1$. Correlation can be explained as a direct (positive or negative) dependence of the variables on each other. The strong correlation is in Figure 2.37 on the left, the mass of the $t\bar{t}$ system vs. the leading top $p_T^{t,1}$.

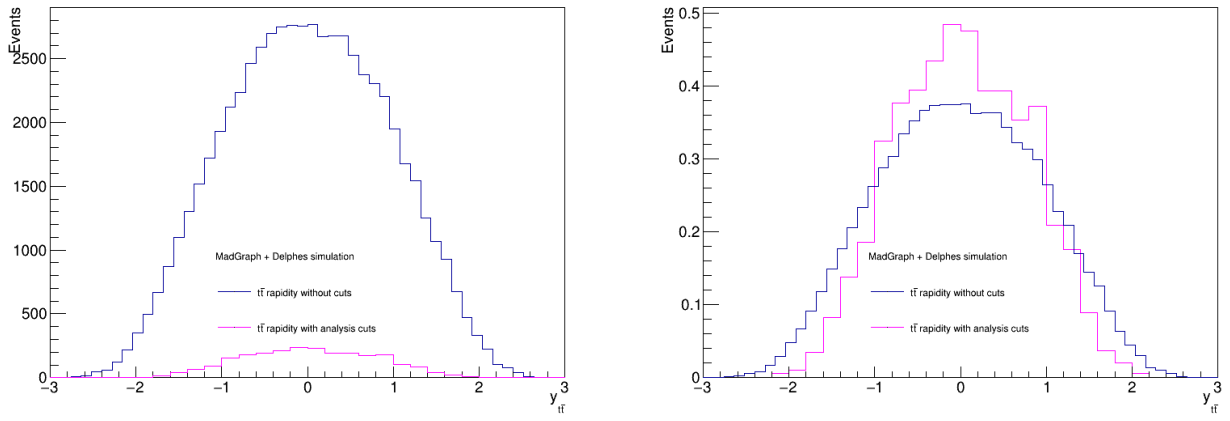

Figure 2.34: The rapidity of the $t\bar{t}$ system $y_{t\bar{t}}$ without analysis cuts (blue) and with analysis cuts (purple) non-normalized (left) and normalized (right) distributions Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

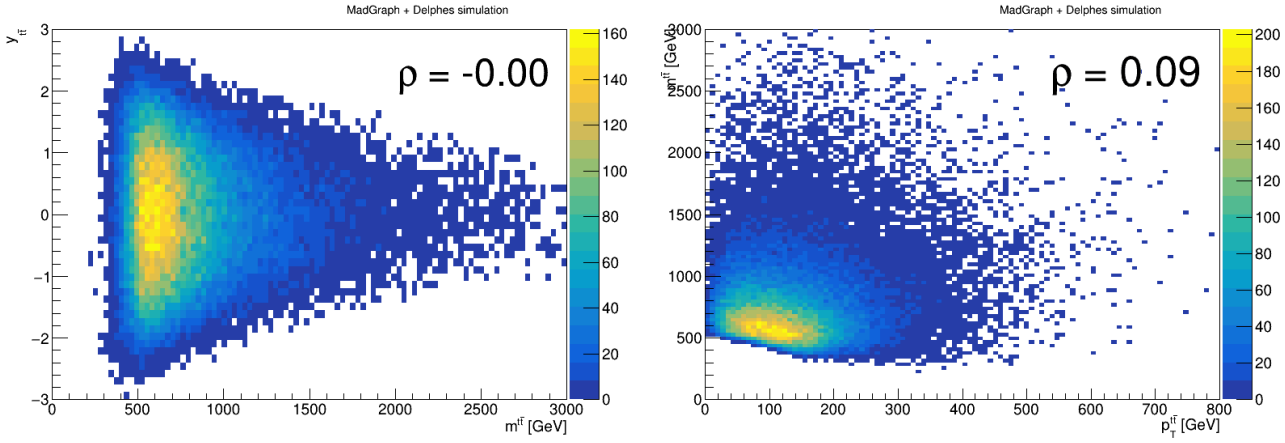

Figure 2.35: The mass of the $t\bar{t}$ system $m_{t\bar{t}}$ vs the rapidity $y_{t\bar{t}}$ (left) and the $p_T^{t\bar{t}}$ vs the mass of the $t\bar{t}$ system $m_{t\bar{t}}$ (right) on the Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

The important part of the analysis is also the correlations histograms between the variables. For example the expected correlation in Figures 2.37 between the leading top p_T and $t\bar{t}$ system p_T and mass, but also interesting anti-correlation in Figure 2.36 on the right for the $p_T^{t\bar{t}}$ vs $\Delta\phi_{t\bar{t}}$.

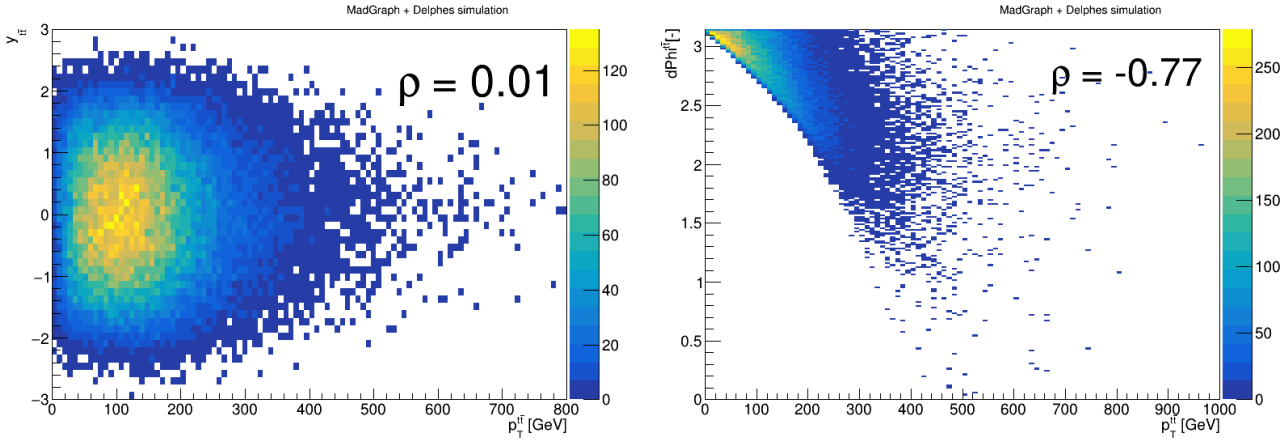

Figure 2.36: The $p_T^{t\bar{t}}$ vs rapidity $y_{t\bar{t}}$ (left) and the $p_T^{t\bar{t}}$ vs $\Delta\phi_{t\bar{t}}$ (right) for the signal Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

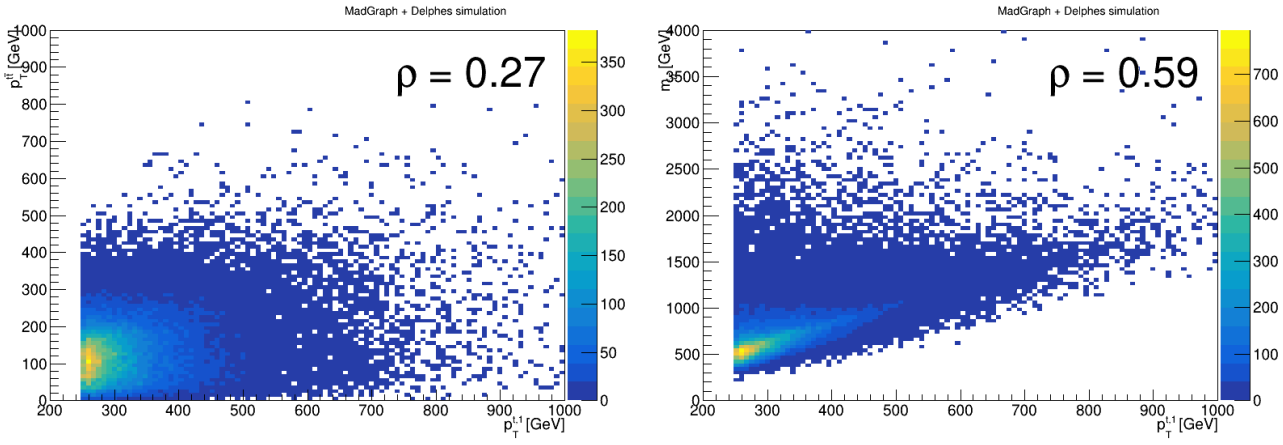

Figure 2.37: The leading top $p_T^{t,1}$ vs $p_T^{t\bar{t}}$ (left) and the leading top $p_T^{t,1}$ vs mass of the $t\bar{t}$ system $m_{t\bar{t}}$ (right) for the signal Monte-Carlo generated with MadGraph, setup in Chapter 2.1.

Chapter 3

ATLAS Analysis

The main part of the presented work within the ATLAS experiment was the cooperation with other physicists, mostly from the Czech Academy of Sciences, to help with the development of the analysis package and measurements of differential cross-sections of top quarks in the all-hadronic channel. The main result of the analysis is the comparison of the corrected data to Monte-Carlo distributions and the calculation of the differential cross-section.

The ATLAS analysis were made on the full data 2015 and 2016 dataset. The biggest contribution to the analysis was adding more physics variables to the package and producing of the final histograms at particle level and parton level. The package was producing spectra of the random top candidate p_T. The work was to modify the package and add the random top candidate rapidity y, the $t\bar{t}$ mass, the $t\bar{t}$ p_T, the $t\bar{t}$ rapidity y, the leading top p_T and rapidity y, the sub-leading top p_T and rapidity y, the inclusive top p_T and rapidity y, p_{out}, $\mathrm{H}_{t\bar{t}}$, $\mathrm{z}_{t\bar{t}}$, the $\Delta\Phi_{t\bar{t}}$ and the $\cos\theta^*$.

The other important study added to the analysis is the correlation among spectra for the theory and measured data comparison with the larger number of cross-sections using a so-called bootstrap method.

3.1 Event selection

The event selection is designed for largest signal statistics with lowest background. Every event must be a reconstructed primary vertex with at least 5 tracks. In the reconstructed event must have zero reconstructed leptons with $p_T > 25$ GeV to eliminate the semi-leptonic events. That requirement is the presence of exactly 2 large-R (anti-k_t $R = 1.0$) jets and for leading top $p_T^{t,1} > 500$ GeV and subleading top $p_T^{t,2} > 350$ GeV and $|\eta| < 2.0$ for both. There must be also two small-R jets (anti-k_t $R = 0.4$) with $p_T > 25$ GeV and $|\eta| < 2.5$.

The signal selection requires the two large-R jets, with the mass within 50 GeV around the top

quark mass $173.34 \pm 0.27(\text{stat}) \pm 0.71(\text{syst})$ GeV [10]. Eventually we require both large-R jets to be top-tagged and b-matched. The top-tagging (using a τ_{32} variable) and b-matching increase the probability that the jet is from the top quark and it is not a background.

3.2 Background

The background are processes that looks similar in the detector as the signal process. In this analysis the signal process is the $t\bar{t}$ production with the allhadronic decay. There are many similar processes but the important in this case is the W+jets and single top quark production depicted in Figure 3.1 or the $t\bar{t}$ + $W/H/Z$ production, because these processes have the largest statistic in the same signal region as our process, but the most important is a QCD background. The background is minimalized by cuts, the most important is a p_T cut and then some sub-structure cuts for example the τ_{32}.

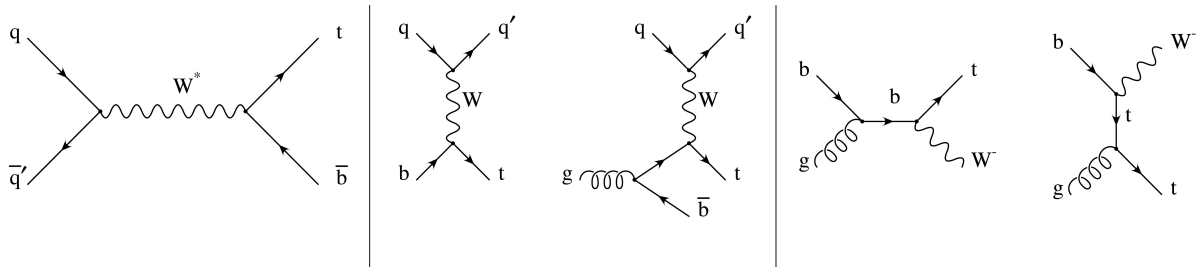

Figure 3.1: The Feynman diagrams of the single top quark production at LO. Edited from [17].

The background coming from the QCD dijet production can be also separated with a data-driven method called ABCD method or the background can be modeled with a Monte-Carlo generator. Than we can remove it but after that the method is not independent of Monte-Carlo, therefore the theoretical prediction. The independence from the theory is necessary for the theory and data comparison.

3.3 TtbarDiffCrossSection package

The analysis flow consists of three main steps. First the AnalysisTop, an official ATLAS analysis package for primary data reduction and object selection, is applied and the data are stored in a ROOT-file. Then the Ttbar Differential Cross-Section package is use for final selection and cuts, for example a p_T cut for leading jet $p_T > 500$ GeV and subleading top jet $p_T > 350$ GeV. After that, there are scripts for performing the unfolding and plotting the data and Monte-Carlo comparison, and the final step is calculating systematical uncertainties. These scripts also produce the final results plots.

3.4 Unfolding

Every measurement is distorted due to detectors resolution effects. The measured data from the detector defines the detector-level spectrum, where some events may have been lost in detector or wrongly reconstructed. The particle-level spectrum is a theoretical spectrum measured with an ideal detector with zero resolution effects and infinite statistics. The unfolding procedure of a given distribution could be summarized with an equation

$$U = \frac{1}{f_{\text{eff}}} \cdot M^{-1} \cdot f_{\text{acc}} \cdot (D - \text{Bg}), \tag{3.1}$$

where D is measured distribution at the detector level, Bg is background modeled with Monte-Carlo generator or calculated from data with some data-driven method which can estimate the background yield from the data; f_{acc} is efficiency of parton/particle cuts at detector selection, called acceptance; f_{eff} is efficiency of detector cuts at parton/particle selection, and M^{-1} is inversion of migration matrix between the detector level and parton/particle level. The difference between the parton and particle level is illustrated in Figure 3.2

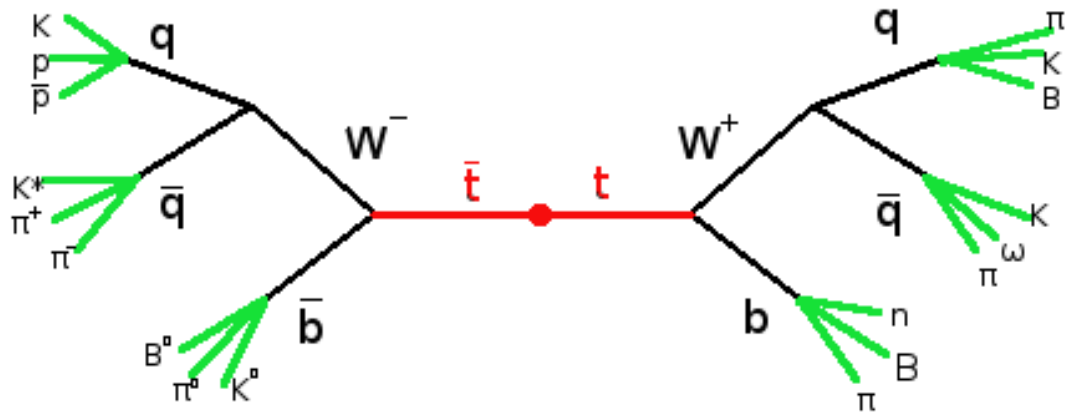

Figure 3.2: The top quark pair production in the all-hadronic decay channel with the parton level highlighted by red lines and the hadronization and particle level by the green lines, the particles types at the end of green lines are only for illustration.

There is a procedure where the spectrum of a chosen variable from data is fluctuated 5000 times in each bin so there are produced 5000 pseudo-experiments. Then the all spectra are analyzed and in each bin the mean value and the width are computed. This width from each bin is the statistical uncertainty.

Bayesian unfolding

Bayesian unfolding [13] is based on the Bayes theorem

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)}, \tag{3.2}$$

where the cause (C) and effect (E) are events and $P(E) \neq 0$. The $P(C)$ and $P(E)$ are the probabilities of observing C or E and observables are independent. The $P(C|E)$ is probability of observing C while the E is true and the $P(E|C)$ is probability of observing E while the C is true.

3.4.1 Particle level

Particle level is defined as a final state formed by stable particles, before entering the detector. Kinematically it is defined close to the detector level, but without the detector limitations (resolution, etc.). It is accessible only in MC simulation. Particle level defines the fiducial phase space. The particle level is a lot easier for unfolding, there are lower modeling uncertainties than when unfolding to the parton level, because this level is much closer to the measured data than the parton level.

3.4.2 Parton level

Unfolding data to the parton level is useful primary for the comparison of experimental data to theoretical predictions from pQCD which come at the full-phase space level. The full-phase space is a production of the top quark pair without any cuts. This level is calculated by theoretical physicists using QCD and QED. However, the results unfolded to the parton level have much larger systematical uncertainties due to large extrapolation and related modeling uncertainties.

3.4.3 Closure test

Closure test is a test of the Unfolding method where there are Monte-Carlo generated spectra at parton/particle level and also Monte-Carlo detector simulation, then the detector simulation called pseudo data is used for unfolding. Ratio of unfolded pseudo data spectrum and generated parton/particle level spectrum must be 1 if the unfolding procedure works correctly. The example from the ATLAS analysis is in Figure 3.3.

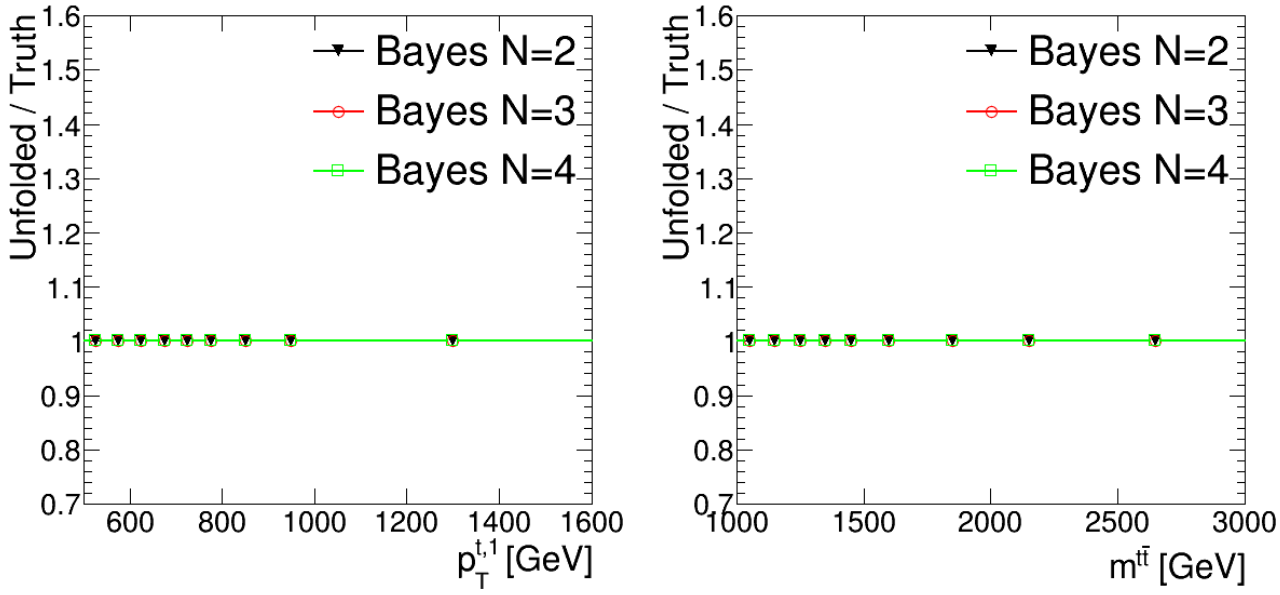

Figure 3.3: The unfolding closure test examples for the $p_T^{t,1}$ and $m_{t\bar{t}}$.

3.4.4 Folding

The other way of the procedure test is through the folding. Folding uses modified Monte-Carlo generated spectra with the migration matrix M and applied it to the particle level distribution Ptc,

$$F = M \cdot \text{Ptc}, \tag{3.3}$$

The folded spectrum should correspond to the MC detector level spectrum, again within errors.

3.4.5 Stress test

A necessary step is to check the unfolding procedure if it will recover the correct spectra therefore independence of the unfolding on the shape of the distribution. This test is made by the reweighting the Monte-Carlo signal sample in two different ways, the difference is in Figure 3.4, which are unfolded and compared to the reweighted sample particle distribution. If the unfolding machinery works correctly the final ratio must be 1 or around within the uncertainties. It is a test that the method is correct and independent shape of the spectra in Monte-Carlo. The stress test results are in Figure 3.5.

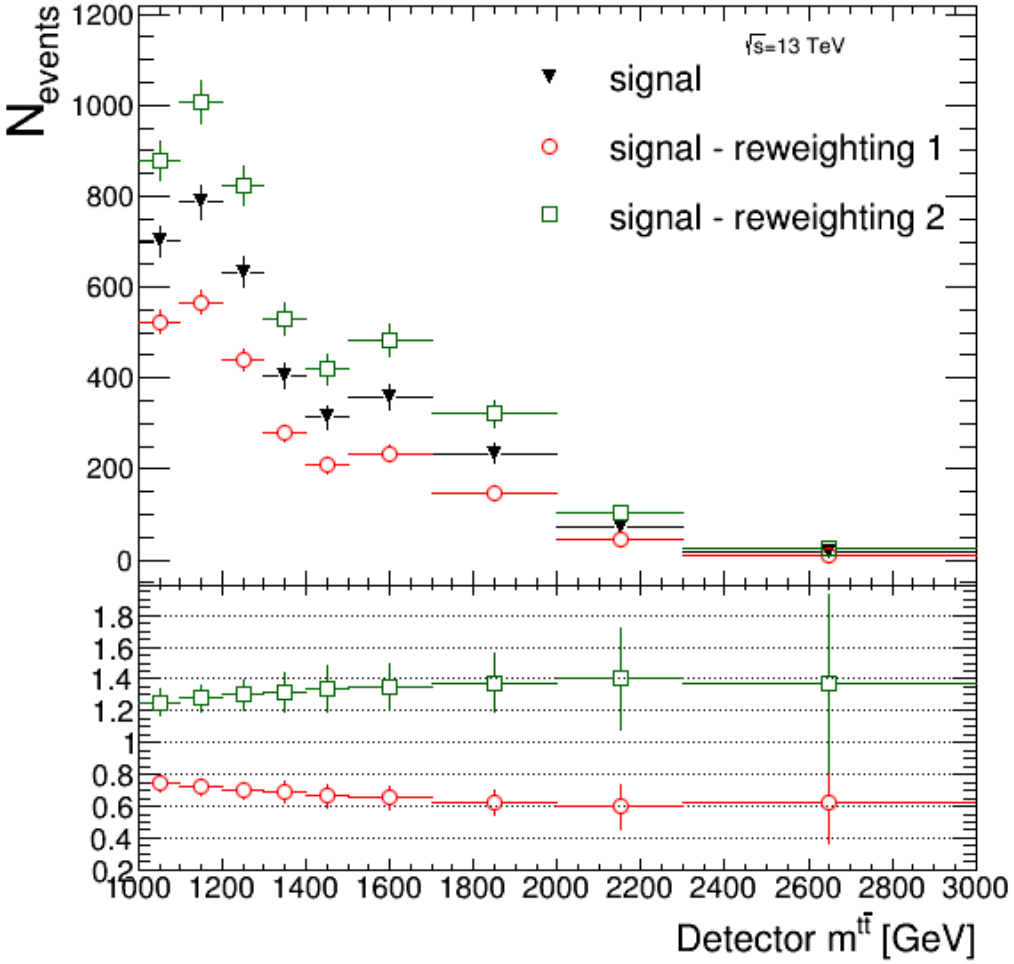

Figure 3.4: The comparison of the reweighted samples and the nominal Monte-Carlo signal sample for the $m_{t\bar{t}}$.

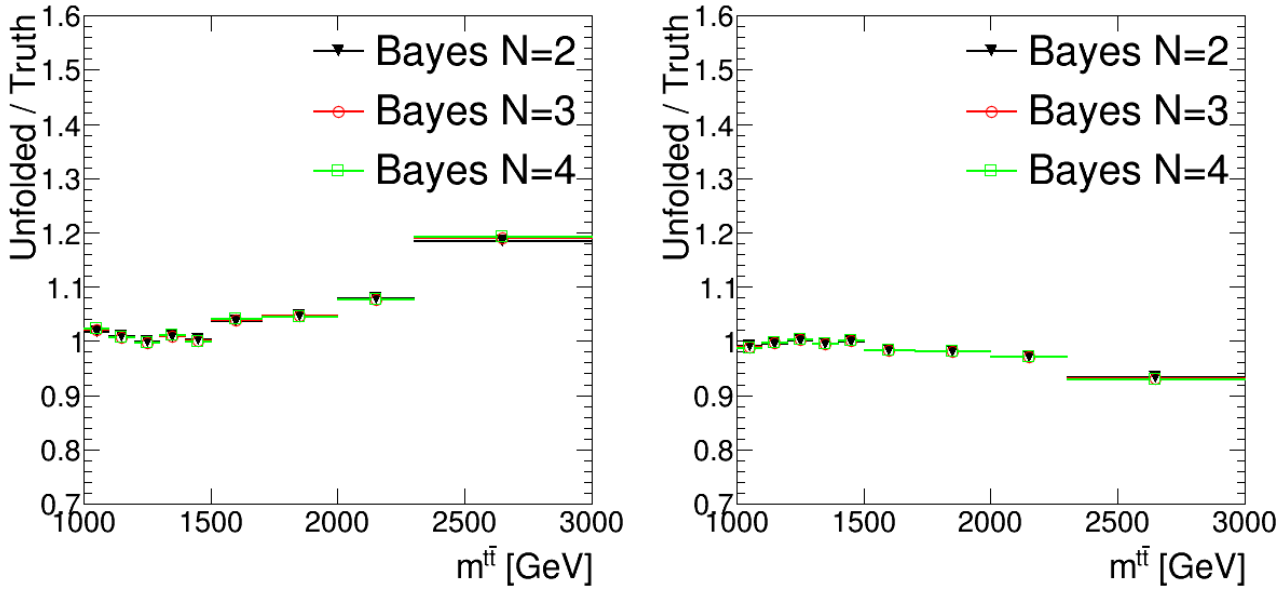

Figure 3.5: The unfolding stress test "reweighted 1" sample ratio (left) and "reweighted 2" sample ratio (right) examples for the $m_{t\bar{t}}$.

The stress test results in Figure 3.5 look reasonable in the context of the total experimental uncertainty except the last bin in both distributions. The larger disagreement is probably due to the lower MC statistic in these bins.

3.5 Results

The shapes of spectra from the real analysis with the ATLAS detector is comparable with the MadGraph analysis results from the Chapter 2.4. Results are presented separately for the particle level and the parton level.

3.5.1 Particle level

In Figures 3.6 –3.12 there are the final results from the analysis at the particle level. All variables have a good agreement of the theory and data within errors. However, the best agreement with the data has the Powheg and Herwig 7 Monte-Carlo generator.

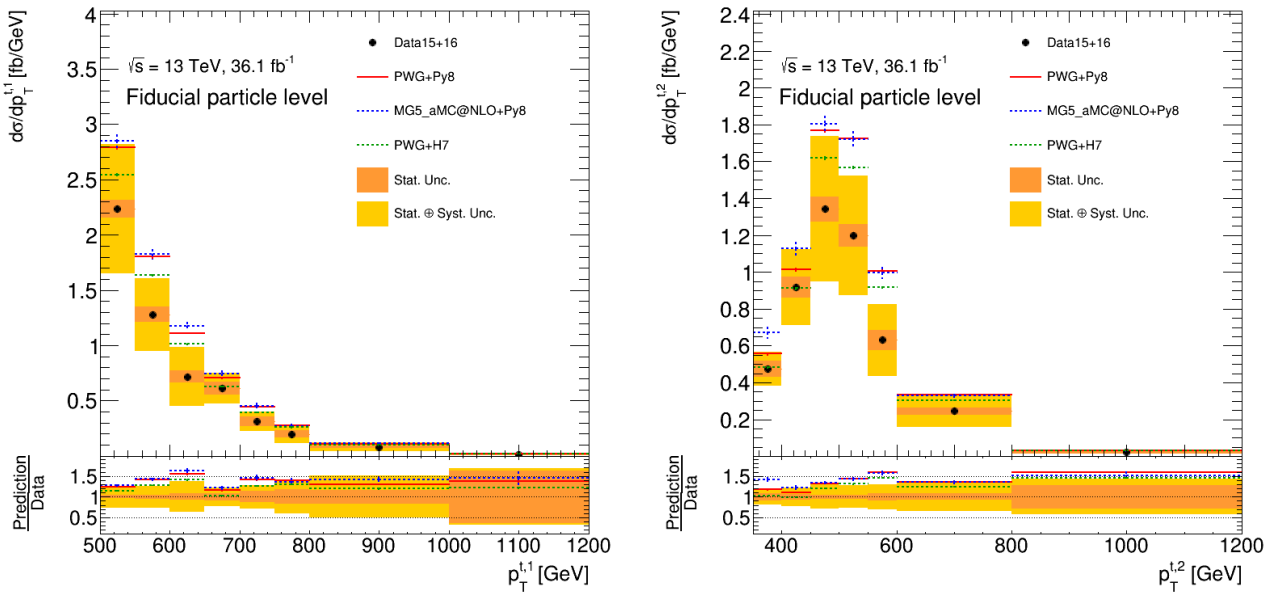

Figure 3.6: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the leading top $p_T^{t,1}$ (left) and the sub-leading top $p_T^{t,2}$ (right) unfolded to the particle level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

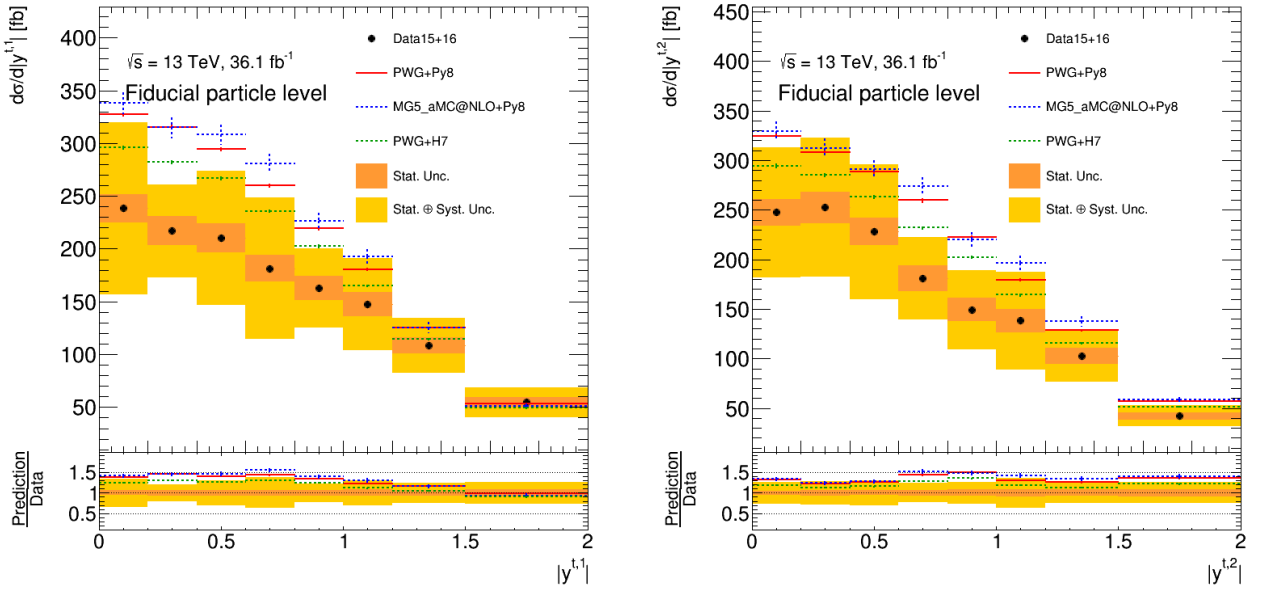

Figure 3.7: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the leading top rapidity $y^{t,1}$ (left) and the sub-leading top rapidity $y^{t,2}$ (right) unfolded to the particle level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

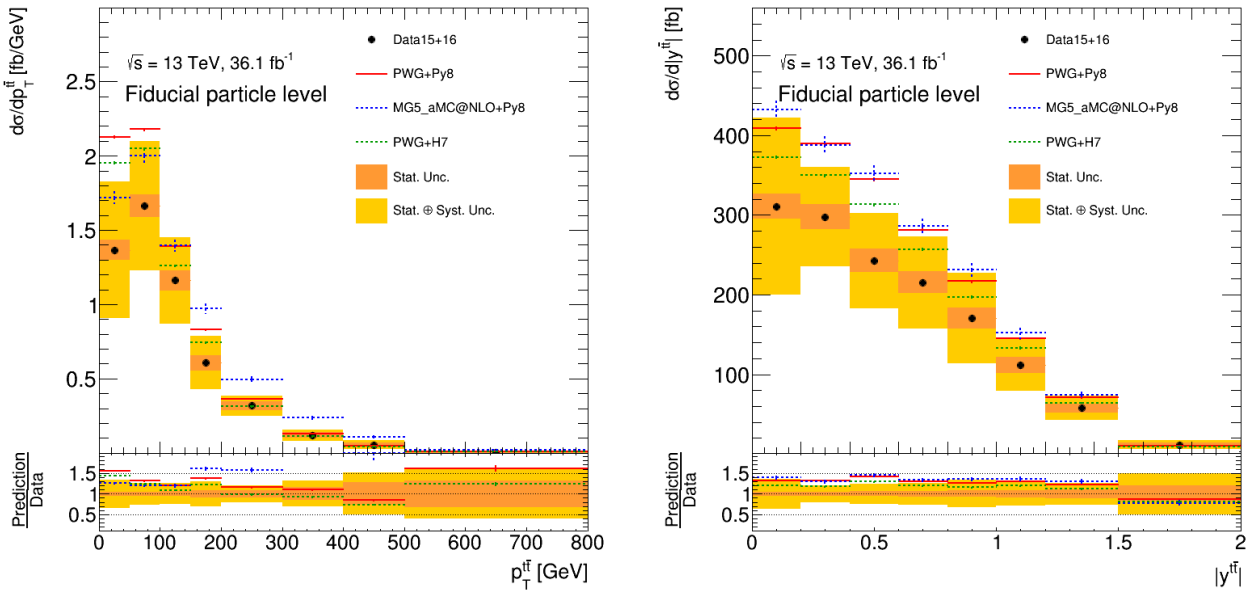

Figure 3.8: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the $p_T^{t\bar{t}}$ (left) and the rapidity $y_{t\bar{t}}$ (right) unfolded to the particle level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

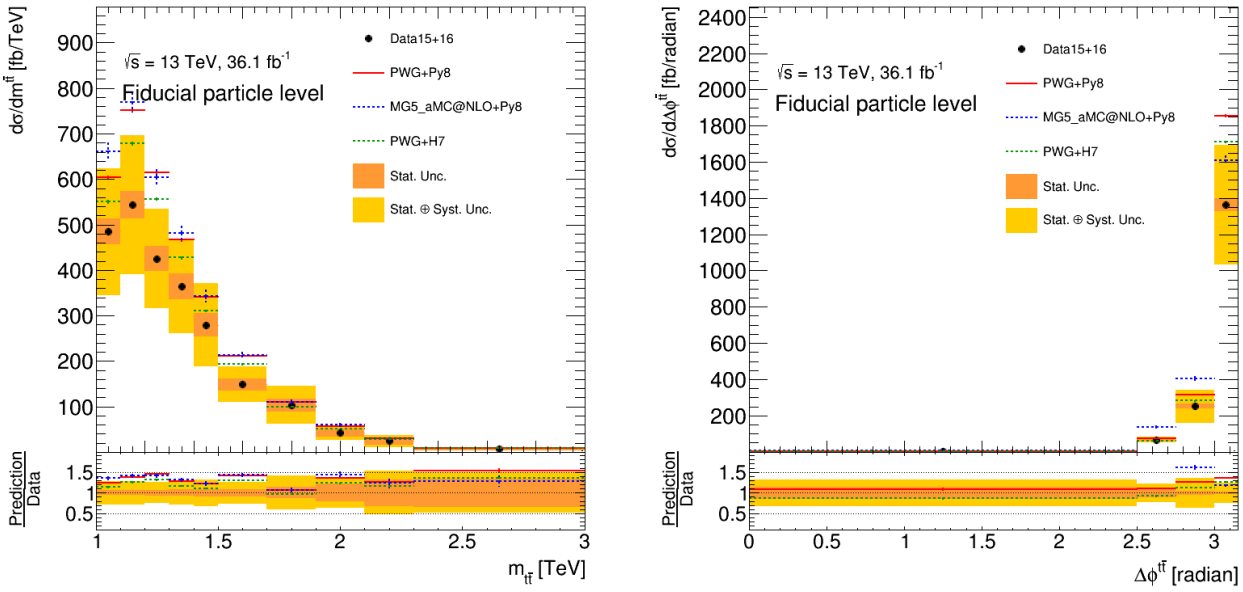

Figure 3.9: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the $m_{t\bar{t}}$ (right) and the $\Delta\Phi_{t\bar{t}}$ (right) unfolded to the particle level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

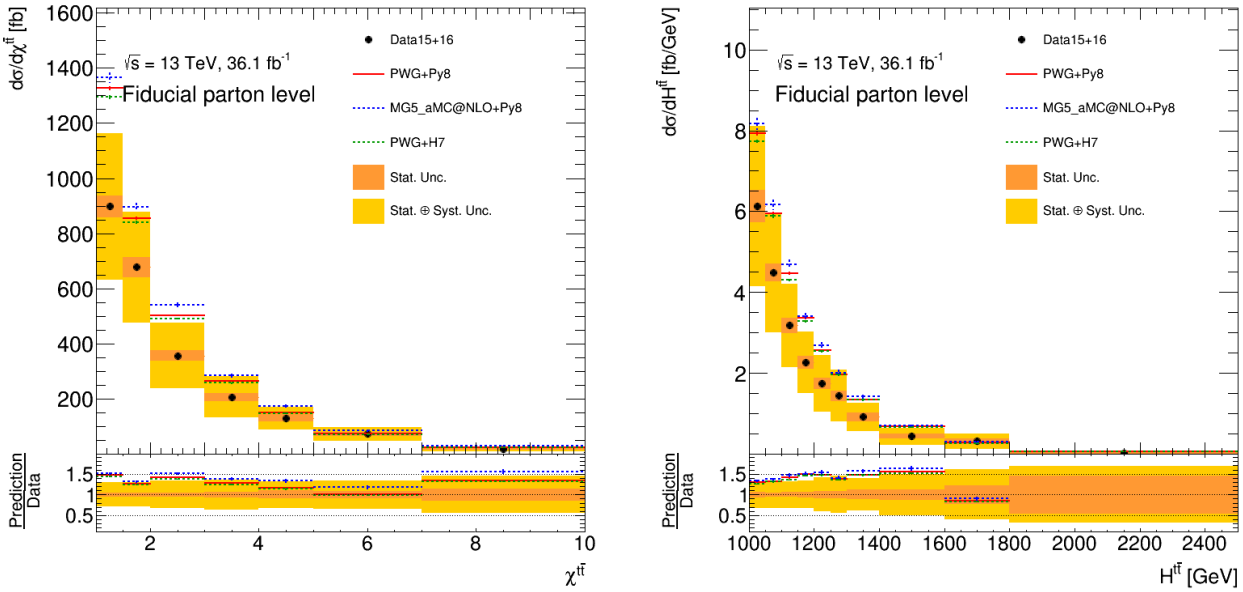

Figure 3.10: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the $t\bar{t}$ system $\chi^{t\bar{t}}$ (left) and the $H^{t\bar{t}}$ of the $t\bar{t}$ system (right) unfolded to the particle level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

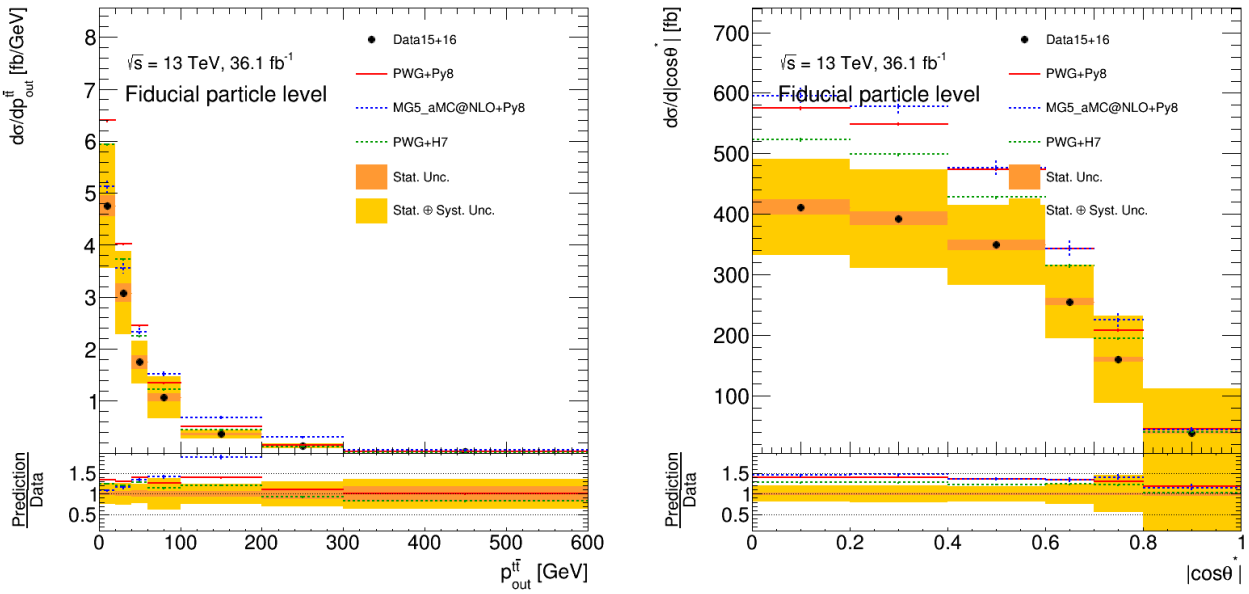

Figure 3.11: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the p_{out} (left) and the $\cos\theta^*$ (right) unfolded to the particle level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

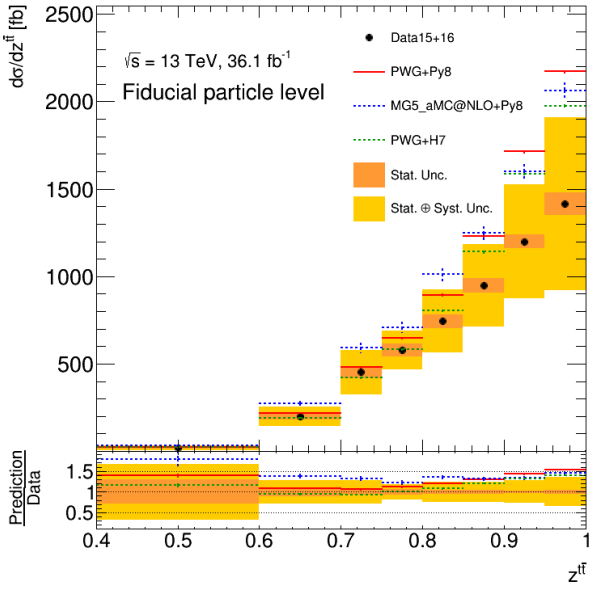

Figure 3.12: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the $z^{t\bar{t}}$ of the $t\bar{t}$ system unfolded to the particle level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

3.5.2 Parton level

In Figures 3.13 –3.19 there are the final results from the analysis at the parton level. All variables have a good agreement between data to theory within errors. However, the agreement is worse than at the particle level and which generator has the best agreement to data is not so obvious as it was at particle level results.

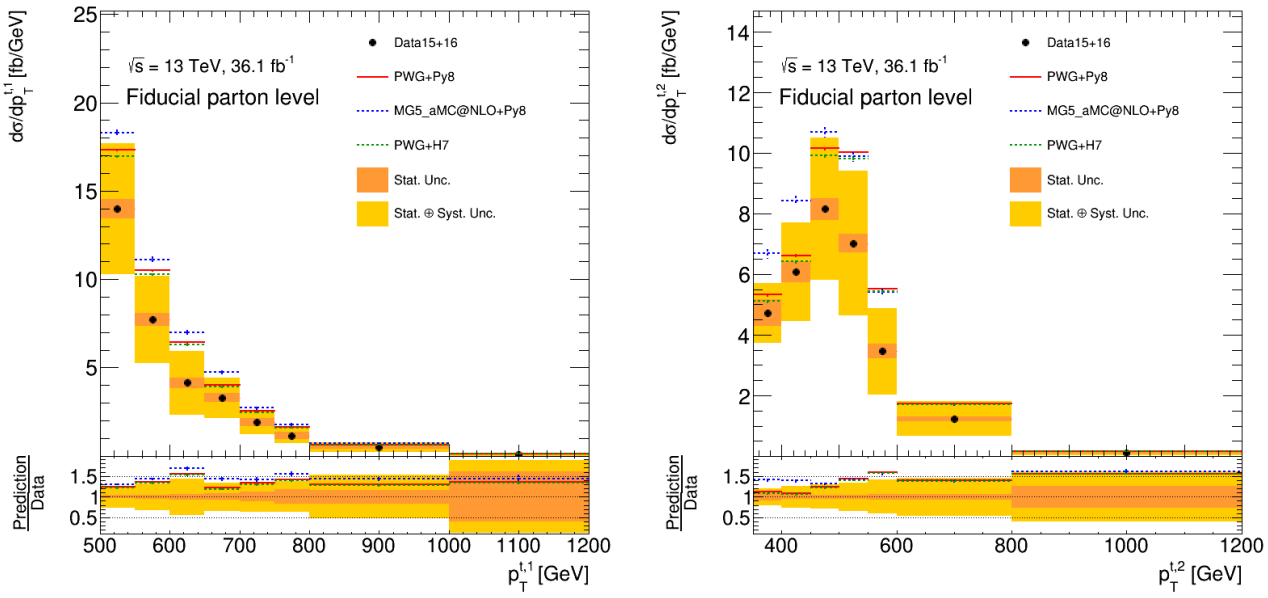

Figure 3.13: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the leading top $p_T^{t,1}$ (left) and the sub-leading top $p_T^{t,2}$ (right) unfolded to the parton level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

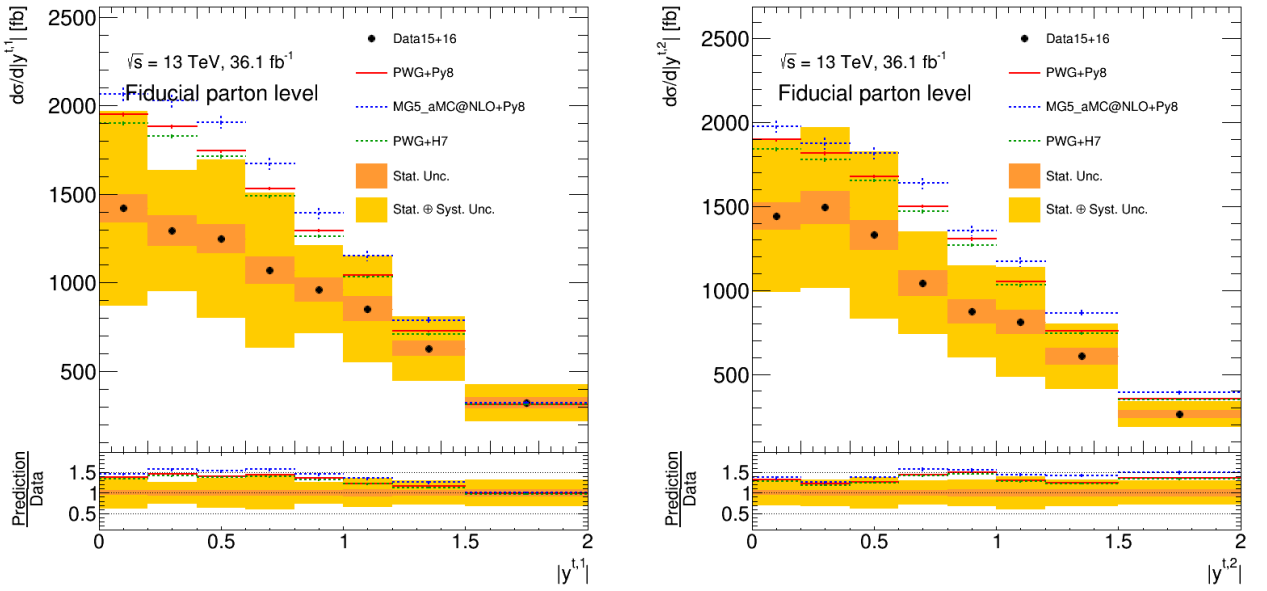

Figure 3.14: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the leading top rapidity $y^{t,1}$ (left) and the sub-leading top rapidity $y^{t,2}$ (right) unfolded to the parton level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

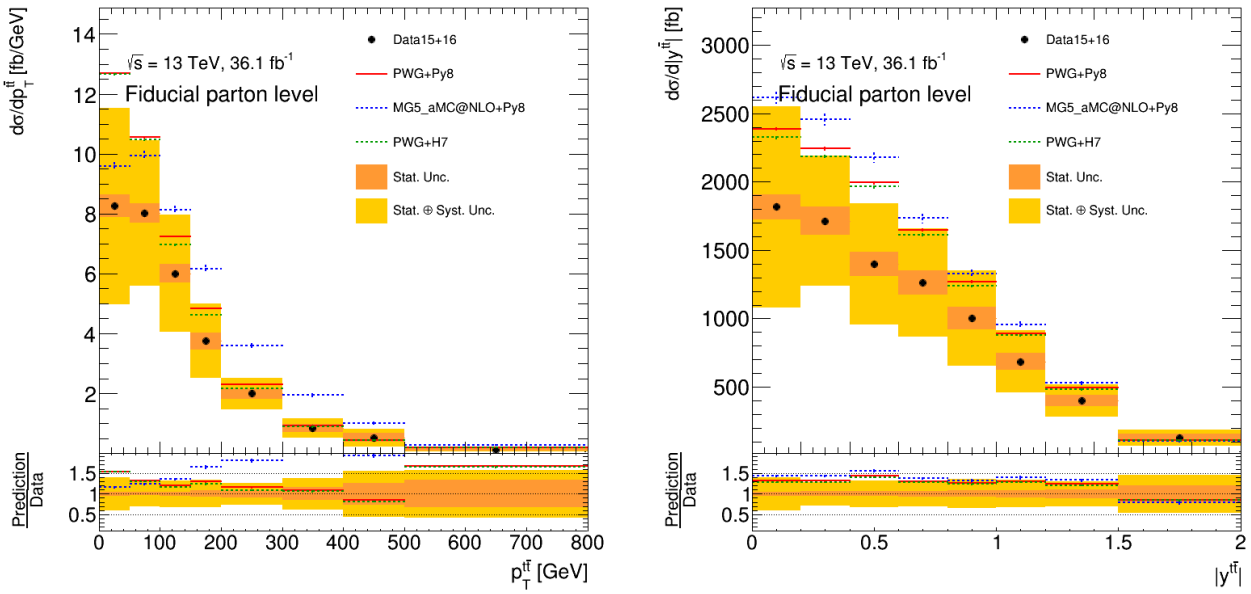

Figure 3.15: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the $p_T^{t\bar{t}}$ (left) and the rapidity $y_{t\bar{t}}$ (right) unfolded to the parton level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

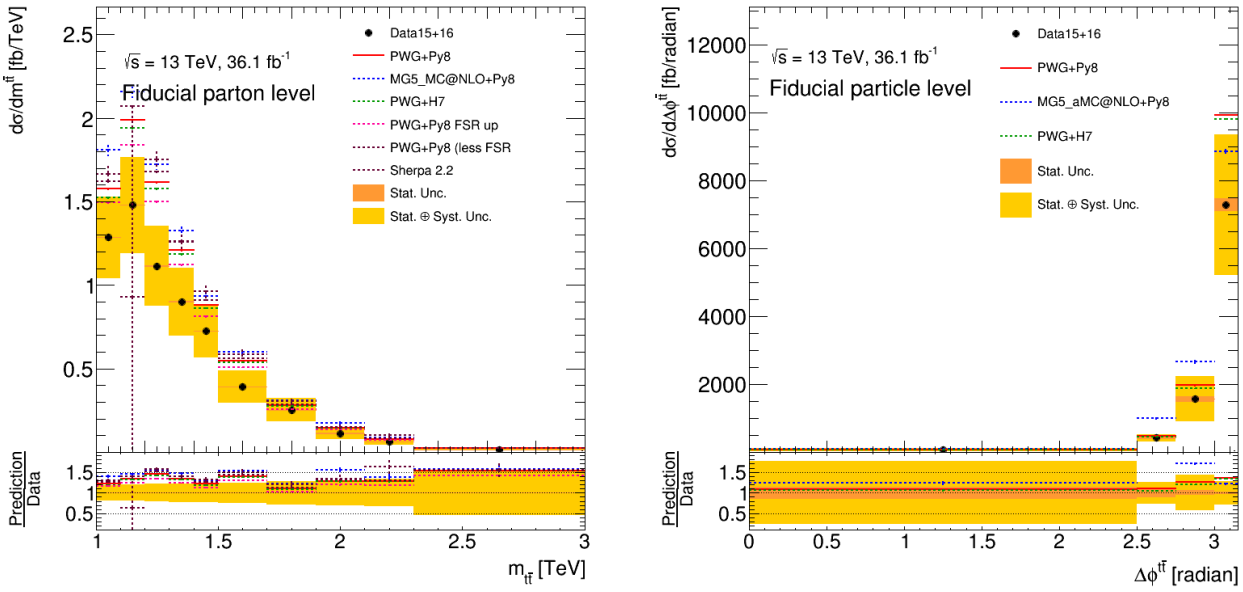

Figure 3.16: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the $m_{t\bar{t}}$ (right) and the $\Delta\Phi_{t\bar{t}}$ (right) unfolded to the parton level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

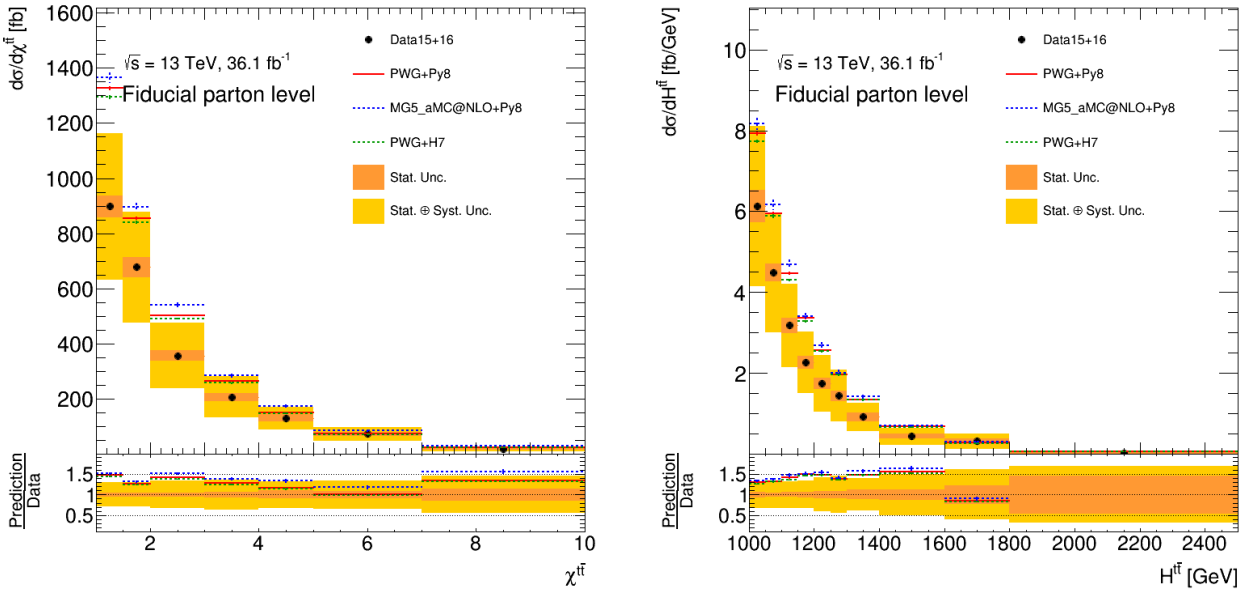

Figure 3.17: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the $t\bar{t}$ system $\chi^{t\bar{t}}$ (left) and the $H^{t\bar{t}}$ of the $t\bar{t}$ system (right) unfolded to the parton level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

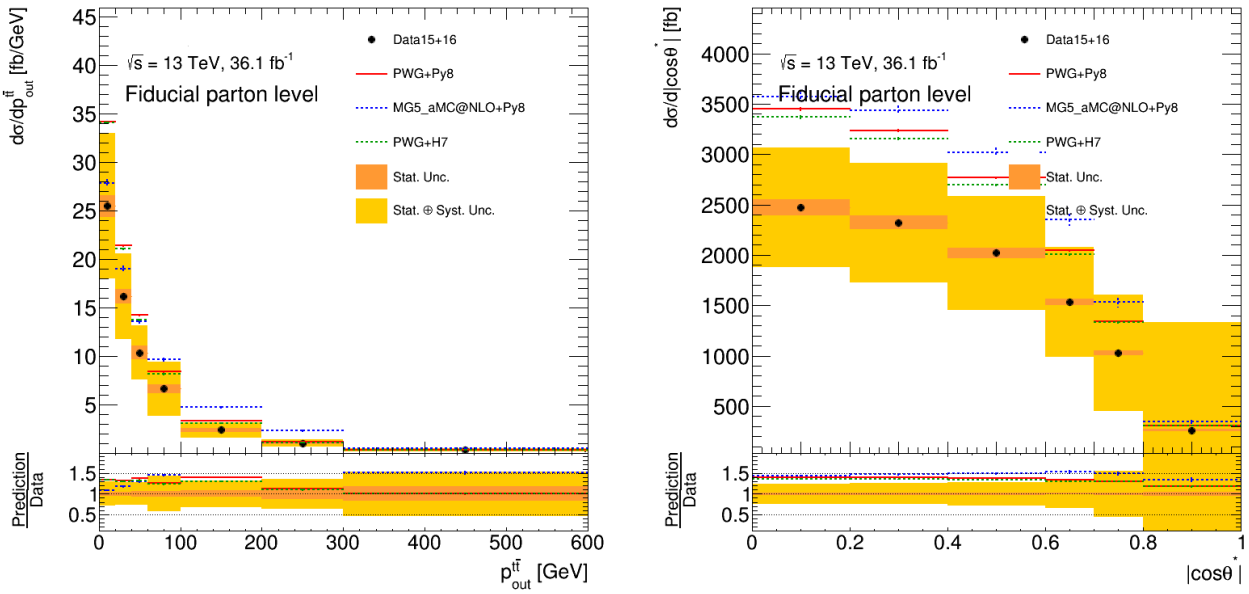

Figure 3.18: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the p_{out} (left) and the $\cos\theta^*$ (right) unfolded to the parton level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

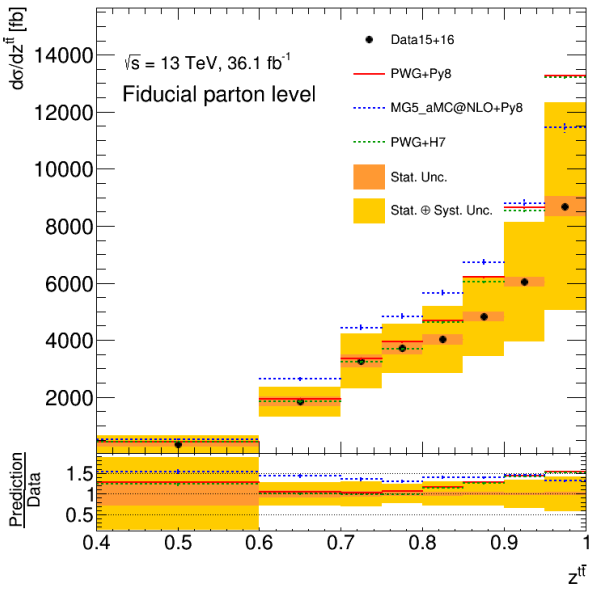

Figure 3.19: Fully corrected data (points) compared to various MC predictions at the NLO order of pQCD (lines) for the $z^{t\bar{t}}$ of the $t\bar{t}$ system unfolded to the parton level. Statistical and total uncertainties are displayed as dark and light yellow band, respectively.

3.6 Correlations between variables

Correlation among the variables can be studied with the Monte-Carlo signal simulation. However, for the comparison of theoretical models and the measured data, we have to unfold a large number of double-differential cross-sections or the single-differential cross-section. But then, statistical correlations between variables have to be measured by unfolding statistically coupled replicas of each spectra. This is made with data and we used a bootstrap method. This method uses replicas of each spectra, derived by fluctuating each event with the Poisson distribution. The correlation matrix can be extracted. The Poisson generator is setup for each event for each replica separately but in a correlated way for all spectra. The result is received by connecting the differential cross-section for each event. The reason for extracting the correlation between variables is to use Cov for a generalized χ^2 [16]

$$\chi^2 = (D - T)^T \mathrm{Cov}^{-1} (D - T), \tag{3.4}$$

where the D is the measured data, T is the theoretical prediction from the Monte-Carlo generator and Cov is the covariance matrix calculated from the correlation matrix and error in each bin. The result from Equation 3.4 use the measured data in a correlated way and the theory and data comparison can be used for more variables simultaneously.

The correlation among unfolded spectra is made by several steps, first step is the event selection,

then the bootstrap framework produces n replicas and each replica is unfolded by usual way. After that the spectra are combined and the correlation matrix is computed.

The correlation is defined by the following equation

$$p_{X,Y} = \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}, \tag{3.5}$$

where the X and Y are the random variables (cross-sections in bins X and Y) with expectation values μ_X and μ_Y with their standard deviations σ_X and σ_Y, E means the expectation value, Cov is covariance, the Corr represents the correlation matrix.

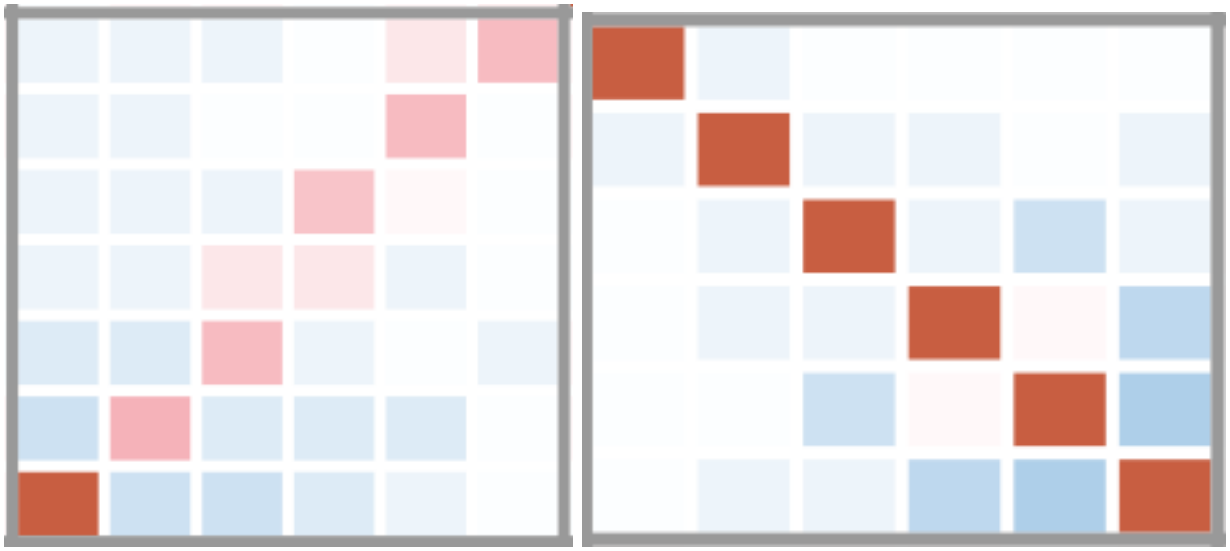

Figure 3.20: An illustration of correlated (left) and anti-correlated (right) spectra.

3.6.1 Spectra correlation

After all the steps of bootstrap correlation, the results of correlations among spectra, for the parton level the absolute cross-section is shown in Figure 3.21 and the relative cross-section in Figure 3.22. There are visible correlation between the $p_{out}^{t\bar{t}}$ and $p_T^{t\bar{t}}$ or between $y_B^{t\bar{t}}$ and $y^{t\bar{t}}$, but also the expected anti-correlation between $\Delta\phi^{t\bar{t}}$ and $p_T^{t\bar{t}}$ because with a increasing p_T is the $t\bar{t}$ boosted in z-axis and decreasing a angle between the two top jets.

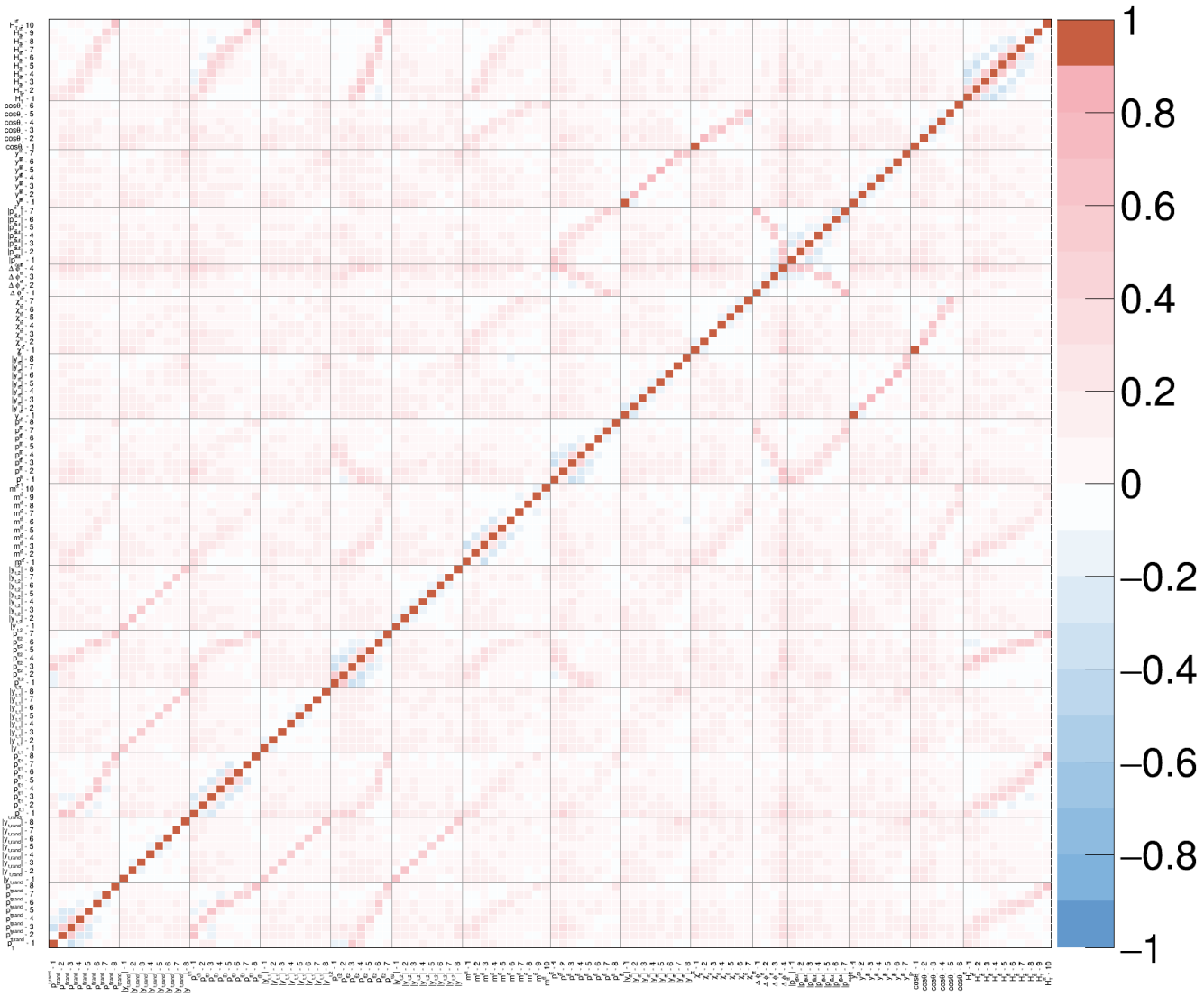

Figure 3.21: The correlation across the absolute spectras at parton level. The variables from left $p_T^{t,1}$, $y^{t,1}$, $p_T^{t,2}$, $y^{t,2}$, $m^{t\bar{t}}$, $p_T^{t\bar{t}}$, $y^{t\bar{t}}$, $\chi^{t\bar{t}}$, $\Delta\phi^{t\bar{t}}$, $p_{out}^{t\bar{t}}$, $y_B^{t\bar{t}}$, $\cos\theta^*$ and $H_T^{t\bar{t}}$.

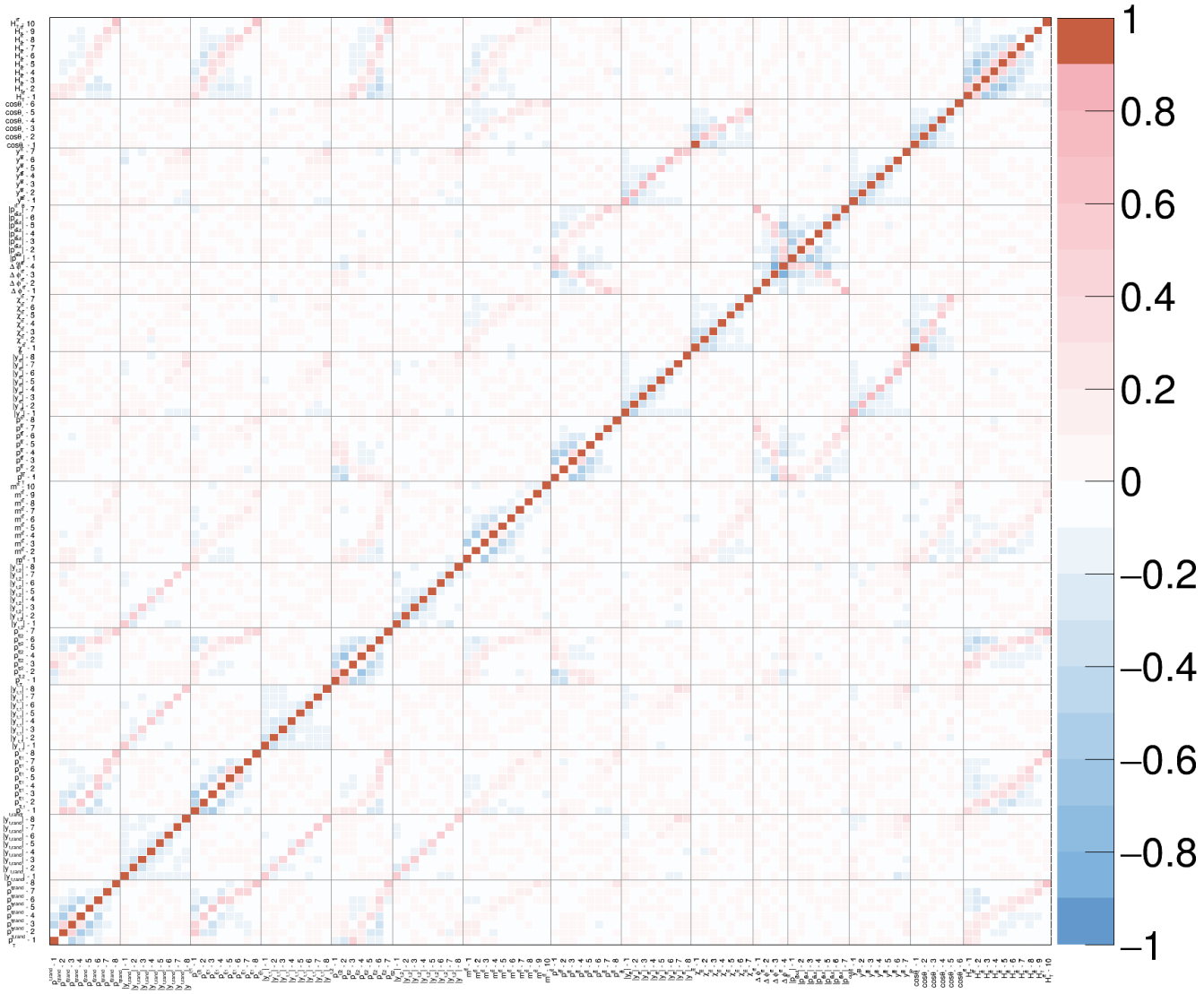

Figure 3.22: The correlation across the relative spectras at parton level. The variables from left $p_T^{t,1}$, $y^{t,1}$, $p_T^{t,2}$, $y^{t,2}$, $m^{t\bar{t}}$, $p_T^{t\bar{t}}$, $y^{t\bar{t}}$, $\chi^{t\bar{t}}$, $\Delta\phi^{t\bar{t}}$, $p_{out}^{t\bar{t}}$, $y_B^{t\bar{t}}$, $\cos\theta^*$ and $H_T^{t\bar{t}}$.

Conclusion

This thesis focused on the analysis of the $t\bar{t}$ production in pp collisions at LHC, using the ATLAS detector. With the MadGraph Monte-Carlo generator samples of the $t\bar{t}$ signal were produced and simple studies of the p_T cuts were performed motivating why we use the high p_T cuts in the boosted region, also there is a study of the τ_{32} and τ_{21} cuts. Together these studies are important for illustrating the event selection described in the ATLAS analysis part of this thesis, which is devoted to a real ATLAS analysis, the work on the analysis package and the main results which is the theory and data comparison and the correlation among spectra. Parts of an ATLAS analysis were a contribution to the ATLAS internal note ATL-COM-PHYS-2016-1696, the title page is shown in Appendix 1. The author would like to thank the project IGA no. PrF/2017/005 for support.

Bibliography

[1] J. Wagner-Kuhr, "Experimental Studies of Top Quark Production," arXiv:1606.02936 [hep-ex].

[2] Emily THOMPSON. What does "boost" mean? In: *Quantum diaries* [online]. Switzerland, 2014 [cit. 2017-01-08]. Available from: http://www.quantumdiaries.org/author/emily-thompson/

[3] J. Erdmann, "Measurement of the inclusive $t\bar{t}\gamma$ cross section at $\sqrt{s} = 7$ TeV with the ATLAS detector," arXiv:1206.5696 [hep-ex].

[4] G. Choudalakis [ATLAS Collaboration], "Unfolding in ATLAS," doi:10.5170/CERN-2011-006.297 arXiv:1104.2962 [hep-ex].

[5] G. Aad *et al.* [ATLAS Collaboration], "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. D **93** (2016) no.3, 032009 doi:10.1103/PhysRevD.93.032009 arXiv:1510.03818 [hep-ex].

[6] ATLAS, Collaboration. Event display of a H - 4mu candidate event. In: *General Photo* [online]. 2012 [cit. 2017-06-05]. ATLAS:1459496. ATLAS-PHO-COLLAB-2012-008. Available from: https://cds.cern.ch/record/1459496

[7] S. Ovyn, X. Rouby and V. Lemaitre, "DELPHES, a framework for fast simulation of a generic collider experiment," arXiv:0903.2225 [hep-ph].

[8] Choudalakis, G., "Fully Bayesian Unfolding", 2012, arXiv:1201.4612

[9] J. Wagner-Kuhr, "Experimental Studies of Top Quark Production," arXiv:1606.02936 [hep-ex].

[10] C. Patrignani *et al.* [Particle Data Group], "Review of Particle Physics," Chin. Phys. C **40** (2016) no.10, 100001. doi:10.1088/1674-1137/40/10/100001

[11] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, "MadGraph 5 : Going Beyond," JHEP **1106** (2011) 128 doi:10.1007/JHEP06(2011)128 [arXiv:1106.0522 [hep-ph]].

[12] T. Sjostrand, S. Mrenna and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," Comput. Phys. Commun. **178** (2008) 852 doi:10.1016/j.cpc.2008.01.036 [arXiv:0710.3820 [hep-ph]].

[13] K. Bierwagen, U. Blumenschein and A. Quadt, "Bayesian Unfolding," doi:10.5170/CERN-2011-006.260

[14] I. Antcheva *et al.*, "ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization," Comput. Phys. Commun. **180**, 2499 (2009) doi:10.1016/j.cpc.2009.08.005 [arXiv:1508.07749 [physics.data-an]].

[15] Maria ALDAYA. NNLO+NNLL top-quark-pair cross sections: Top-quark-pair cross sections at 7, 8, 13, and 14 TeV calculated for specific top-quark-mass values. In: *TWiki CERN* [online]. 2014 [cit. 2017-07-31]. Dostupné z: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO#Top_quark_pair_cross_sections_at

[16] G. Aad *et al.* [ATLAS Collaboration], "Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector," Eur. Phys. J. C **76** (2016) no.10, 538 doi:10.1140/epjc/s10052-016-4366-4 [arXiv:1511.04716 [hep-ex]].

[17] Andrea GIAMMANCO and Jeannine WAGNER-KUHR. Measurement of the t-channel single Top-quark production rates in pp collisions at 7 TeV. In: *CMS Web CERN* [online]. 2011 [cit. 2017-08-02]. Dostupné z: http://cms.web.cern.ch/news/measurement-t-channel-single-top-quark-production-rates-pp-collisions-7-tev

[18] N. D. Christensen and C. Duhr, "FeynRules - Feynman rules made easy," Comput. Phys. Commun. **180** (2009) 1614 doi:10.1016/j.cpc.2009.02.018 [arXiv:0806.4194 [hep-ph]].

[19] J. D. Wells, "How to Find a Hidden World at the Large Hadron Collider," In Kane, Gordon (ed.), Pierce, Aaron (ed.): Perspectives on LHC physics 283-298 [arXiv:0803.1243 [hep-ph]].

[20] Riccardo di Sipio, private communication

Appendix

Appendix 1

ATLAS NOTE

ATL-COM-PHYS-2016-1696

30th July 2017

Draft version 1.3.1 (top wg approval)

Measurements of $t\bar{t}$ differential cross-sections in the all-hadronic channel with the ATLAS detector using highly boosted top quarks in pp collisions at $\sqrt{s} = 13$ TeV with the full 2015+2016 dataset

Peter Berta[a], Ye Chen[c], Kyle Cormier[b], Riccardo Di Sipio[b], Robin Hayes[b], Petr Jačka[f], Roman Lysák[f], Jiří Kvita[d], Rupert Leitner[a], Cassandra Miller[b], Shima Shimizu Noda[c], Jan Palicka[d], Marino Romano[e], Pekka Sinervo[b], Francesco Spanó[g], Federica Fabbri[e], Yuji Yamazaki[c]

[a]*Charles University in Prague,* [b]*University of Toronto,* [c]*Kobe University,* [d]*Palacky University,* [e]*University of Bologna,* [f]*Acad. of Sciences of the Czech Rep.,* [g]*Royal Holloway, University of London*

Abstract

This note documents the measurement by the ATLAS experiment of the top quark pair production differential cross-sections as a function of the top quark and $t\bar{t}$ system kinematic observables. The study was performed using the full dataset from pp collisions at $\sqrt{s} = 13$ TeV collected by the ATLAS detector in 2015 and 2016, corresponding to an integrated luminosity of 36.1 fb^{-1}. The $t\bar{t}$ events were selected in the all-hadronic channel using highly-boosted top quarks. A measurement is made of the fiducial phase-space total cross-section and differential cross-sections. Measurements of full phase-space differential cross-sections are presented. Due to the large $t\bar{t}$ cross-section at the LHC, such measurements allow a detailed study of the properties of the top-quark production and decay, enabling precision tests of perturbative QCD.