Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Corso di Laurea Magistrale in Fisica

Measurement of $t\bar{t}$ production differential cross section at $\sqrt{s} = 8$ TeV at ATLAS

Relatore:
Prof. Nicola Semprini Cesari

Presentata da:
Federica Fabbri

Correlatori:
Dott. Lorenzo Bellagamba
Dott. Matteo Franchini

Sessione II
Anno Accademico 2013/2014

Abstract

Il top quark gioca un ruolo fondamentale nel programma scientifico di LHC. L'enegia e la luminosità raggiunte permettono l'acquisizione di una grande mole di dati, specialmente in regioni cinematiche mai studiate precedentemente.

In questa tesi è presentata la misura della sezione d'urto differenziale di produzione di coppie $t\bar{t}$ effettuata sui dati raccolti da ATLAS nel 2012 in collisioni protone protone a $\sqrt{s} = 8$ TeV, corrispondenti ad una luminosità integrata di 20.3 fb^{-1}. La misura è effettuata selezionando eventi $t\bar{t}$ nel canale semileptonico il cui quark top, che decade adronicamente, ha un impulso trasverso superiore a 300 GeV. Il decadimento adronico del quark top è ricostruito come un unico jet di ampio raggio e identificato studiando la struttura interna del jet stesso. La sezione d'urto differenziale ottenuta è stata confrontata con le principali distribuzioni teoriche evidenziando discrepanze anche di circa il 25% (dipendente dal generatore Monte Carlo adottato). Dato che le distribuzioni cinematiche del processo di produzione $t\bar{t}$ dipendono sensibilmente dalla scelta del set di funzioni di distribuzione partonica utilizzato nelle simulazioni, si possono ottenere informazioni rilevanti sulle PDF dei gluoni. In particolare, in questa tesi è stato effettuato uno studio sistematico utilizzando diversi insiemi di PDF di protoni con lo scopo di determinare quello che meglio descrive le distribuzioni sperimentali. Le tecniche qui utilizzate nelle analisi di eventi "boosted" saranno fondamentali nella presa dati a $\sqrt{s} = 13$ TeV quando la maggior parte degli eventi di interesse sarà costituita da particelle di elevato impulso.

Abstract

Top quark studies play an important role in the physics program of the Large Hadron Collider (LHC). The energy and luminosity reached allow the acquisition of a large amount of data especially in kinematic regions never studied before. In this thesis is presented the measurement of the $t\bar{t}$ production differential cross section on data collected by ATLAS in 2012 in proton proton collisions at $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity of 20.3 fb^{-1}. The measurement is performed for $t\bar{t}$ events in the semileptonic channel where the hadronically decaying top quark has a transverse momentum above 300 GeV. The hadronic top quark decay is reconstructed as a single large radius jet and identified using jet substructure properties. The final differential cross section result has been compared with several theoretical distributions obtaining a discrepancy of about the 25% between data and predictions, depending on the MC generator. Furthermore the kinematic distributions of the $t\bar{t}$ production process are very sensitive to the choice of the parton distribution function (PDF) set used in the simulations and could provide constraints on gluons PDF. In particular in this thesis is performed a systematic study on the PDF of the protons, varying several PDF sets and checking which one better describes the experimental distributions. The boosted techniques applied in this measurement will be fundamental in the next data taking at $\sqrt{s}$=13 TeV when will be produced a large amount of heavy particles with high momentum.

Contents

Introduction

Since its discovery at Fermilab in 1995 the study of the top quark has represented one of the most interesting fields in particle physics because of its very peculiar properties, as the short life-time and the high mass, that makes the top quark the heaviest known fundamental particle.

Top quark studies play an important role in the physics program of the Large Hadron Collider (LHC) and in particularly for the ATLAS experiment. Thanks to the large center of mass energy of $\sqrt{s} = 7 \div 8$ TeV at LHC the cross section for $t\bar{t}$ production is a factor 20 higher than the one at the Tevatron and the statistic is furthermore increased by the large luminosity. The energy and luminosity reached by the LHC, allow the acquisition of a large amount of data especially in kinematic regions never studied before.

In this thesis the measurement of the $t\bar{t}$ production differential cross section at very high transverse momentum (>300 GeV) is performed as a function of the top quark p_T. In this analysis only events with top-antitop pairs decaying semileptonically are selected, where one top quark decays hadronically and the other one leptonically. Such measurement is performed using new reconstruction techniques specifically designed to deal with the collimated decay topology of high p_T quarks; in particular the hadronic top quark decay is reconstructed as a single large radius (large-R) jet. Specifically algorithms have been developed to treat boosted objects using techniques of background subtraction or identification based on large-R jet substructure variables. The selection and reconstruction algorithms have been optimized to increase the $t\bar{t}$ selection efficiency in a very high kinematic region reaching the TeV range. This new methods will be used and further more optimized also in the next data taking at $\sqrt{s} = 13$ TeV, where top pairs will be produced at even higher momentum and the boosted regime will became predominantly. In the new scenario opened by the larger energy reached by LHC the algorithms optimized in resolved regime will lead to an inadequate reconstruction efficiency and will be fundamental apply the boosted techniques, both in study of top than other heavy

particle.

The $t\bar{t}$ production differential cross section is measured as a function of the top transverse momentum using the full 2012 data taking corresponding to an integrated luminosity of 20.3 fb^{-1}. The measured spectra obtained separately in the electron and muon channel are corrected for the distortions introduced from the detector and successively combined; the results have been compared to theoretical predictions in a fiducial phase space region. The differential cross section final result shows a good shape accordion with the theoretical expectations but a discrepancy in normalization of about 20 % is observed in all the top p_T spectrum. I collaborated in all the analysis steps developing part of the code for the event selection, object reconstruction, histogram production, data MC comparison and performing some study on reconstruction efficiency.

The study of differential distributions furthermore allows to perform convincing tests on theoretical predictions of QCD perturbative calculations. In particular in this thesis I have performed a systematic study on the parton distribution function (PDF) of the protons, varying several PDF sets and checking which one better describes the experimental distributions. This analysis has an high relevance because the kinematic distributions of the $t\bar{t}$ production process are very sensitive to the choice of the PDF set used in the simulations. The $t\bar{t}$ differential cross section measurements could furthermore provide a constraint on the gluons PDFs and provide input to the PDF fits.

The structure of the thesis is the following. In Chapter 1 an overview on the Standard Model is presented, with particular attention on the theoretical aspects of the top quark and its cross section measurements. The description of the ATLAS detector is reported in Chapter 2; here all the detector resolutions and acceptance's regions are also presented. The full analysis chain is described in Chapter 3, here the reconstruction of the physical objects used during the analysis is presented beside the treatment of high transverse momentum objects topology. In the same chapter a description of the collected data samples and the Monte Carlo simulations for both the signal and the different considered backgrounds are reported; also the event selection, the unfolding procedure and the analysis framework are presented here. A study on PDF systematics and different PDF sets performed on data collected in 2011 with $\sqrt{s} = 7$ TeV is presented in Chapter 4. The evaluation of the measure uncertainties and the presentation of the final differential cross section results are accomplished in Chapter 5.

Chapter 1

Introduction to Top quark physics

The study of elementary particles physics may be said to began in 1897 with the discovery of electron made by Joseph J. Thomson. In the following century many particles have been discovered, initially using the cosmic rays, the only source of high energy available at that time, and subsequently, the high energy accelerators. The improvement in the quality and intensity of the beams available with higher and higher energies has been the keystone of a long process of discoveries and acquisition now consolidated into a large experimental-theoretical framework known as Standard Model.

1.1 The Standard Model

The first particles discovered have been electrons and protons, that together with the neutrons are the fundamental components of ordinary matter. But soon, approximately in the period between 1935 and 1950, new particles without any role in the constitution of ordinary matter were discovered (muons, pions and kaons) giving rise to the modern elementary particle physics. In the 50s, the unexpected discovery of more than a hundred new strongly interacting particles (hadrons proliferation) led to the formulation of the quark model by Gell-Mann and Zwei (1964), one of the cornerstones of the standard model.

According to this model, all matter is made up of 12 particles of spin $\frac{1}{2}$ (fermions) divided in two families of six elements called leptons and quarks (fig. 1.1).

In the SM beside the 12 fundamental particles are present 12 antiparticles with the same mass and lifetime but opposite charge and magnetic moment, predicted for the first time by Dirac in 1931.

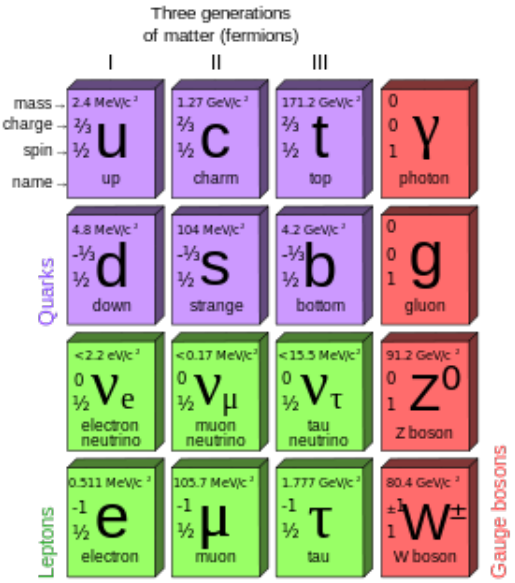

Figure 1.1: Particles of the standard model, here are reported both fermions than interactions mediators. For each particle are indicated mass, spin and charge.

On the basis of weak interaction properties the leptons are divided in three generations each one containing an electron-like particle that carries unitary negative charge (e, μ and τ) and a neutral lepton called neutrino (ν). A different flavour of neutrino is coupled with each flavour of charged lepton. This property described by a conserved quantum number called the *leptonic number*, (if a charged particle decays to an electron, electronic lepton number (L_e) equal to 1, the decay must contain also a neutrino of the same flavour that has L_e equal to -1). The first neutrino (ν_e) was initially postulated by Pauli in 1930, in order to account for the energy and momentum missing in the β-decay. The charged muon and tauon are both unstable, and decay spontaneously to electrons, neutrinos and other particles; their mean lifetime [1] are respectively $(2.1969811 \pm 0.0000022) \cdot 10^{-6} s$ and $(290.6 \pm 1.0) \cdot 10^{-15} s$.

The SM explains all the strong interacting particles as formed by bounded state of three quarks (*baryons*) or of a quark and an antiquark pair (*mesons*), called both hadrons. As in the case of leptons, there are six flavours of quarks organized in three families carrying each one a fractional electric charge: $\frac{2}{3}e$ for up(u), charm (c) and top (t) and $-\frac{1}{3}$ for down (d), strange (s) and beauty (b). While leptons exist as free particles is not so far been possible to isolate a quark from the respective hadron, a property of strong interaction called quark confinement. The mass of the quarks vary in a very broad interval: from few Mev till almost 200 GeV, as is

Strong	Electromagnetic	Week	Gravity
1	10^{-2}	10^{-7}	10^{-39}

Table 1.1: Relative strenght of different interactions

shown in figure 1.1

The stable matter in our universe is made up of electrons, up and down quarks (protons and neutrons) i.e. only three of the 24 Standard Model particles. The combinations of heavier quarks s, c, b are unstable decaying rapidly (usually 10^{-13} s) to u, d ones and have played a role in the cosmos in the previous stages of development. The only quark that decays before hadronization is the top quark that has a mean lifetime smaller than the mean time needed to hadronize, so the study of this particle, as will be discussed later, is really important because it is the only quark that can be studied as a free quark.

The quarks carry also a further charge, called colour, that can assume three different values (blue, green and red), consequently the number of different quark interacting to form hadrons is 18, 6 flavour for 3 color. This name as been given only to symbolize that the combination of three colour, as the combination of a colour and an anticolour, is colourless, "white".

The SM is a renormalisable field theory that describe all the particles as quantum of an associated field, it comprehends also interactions between these particles that are described in terms of the exchange of bosons (see figure 1.1) between the fermions constituents.

Up to now four different types of fundamental interactions are known each one described by a family of fields whose quanta are listed in figure 1.1:

- *Electromagnetic interaction:* is due to the electric charge which couples material and electromagnetic field (whose quanta are called photons). It is responsible for almost all phenomena in extra nuclear physics.

- *Weak interaction:* is due to weak isospin, it is the responsible for hadron instability and the only one that can involve neutrinos and that can change the flavour of a particle. The mediators of weak interactions are the $W^{\pm}$ and Z^0 bosons.

- *Strong interaction:* is due to the color charge, it is the interaction responsible for binding the quarks in the hadrons. It is mediated by 8 massless particles, the gluons.

- *Gravitational interaction:* interaction that act between all types of particles. It is the weakest at the scale of the particle physics , as can be seen in table 1.1, although it is dominant on the scale of the universe. At the moment isn't inserted in the SM but it is supposedly mediated by a spin 2 boson, the graviton.

The electromagnetic and weak force had been unified in 70^{th}, thanks to the work of Glashow, Weinberg, Salam, t'Hooft and others. The main problem was the origin of mass of mediator of weak field that for that for the gauge field theory should be massless and with an infinite interaction range . This is in contrast with the experimental observations that found a short range for the weak interactions. A solution was proposed in 1964 by Anderson, Higgs [2], Englert and Brout [3], they added to the SM Lagrangian a scalar field that permeated the vacuum and associated both to the material fermion fields and to weak field. Through the spontaneous symmetry breaking of the Lagrangian this field could give mass to W and Z bosons (BEH mechanism) and interacting with material fields also to all the fermions. This scheme furthermore requires a single coupling for the weak and electromagnetic interaction leaving the photon massless. The theory include also a massive, chargeless boson (Higgs boson) that is the quantum of the vacuum field introduced.

The discovery of a new boson was announced by ATLAS [4] and CMS [5] experiments on 4^{th} of July 2012 at CERN. Experiments conducted in last two years seems to confirm that this new particle of mass (125.9 ± 0.4) GeV has all the characteristics to be the Higgs boson.

1.1.1 Electromagnetic interaction

The Electromagnetic interaction, described by the quantum electrodynamics (QED), is the most known and well described interaction in the standard model.

The electromagnetic potential $V(r)$ has infinite range (r) and in the simplest case assumes the form:

$$V(r) \propto \frac{\alpha}{r} \tag{1.1}$$

where α is the coupling constant:

$$\alpha = \frac{e^2}{4\pi\hbar c} = \frac{1}{137.0360} \tag{1.2}$$

The figure 1.2 shows the fundamental vertex of this interaction, the vertexes of

boson auto-interaction are forbidden.

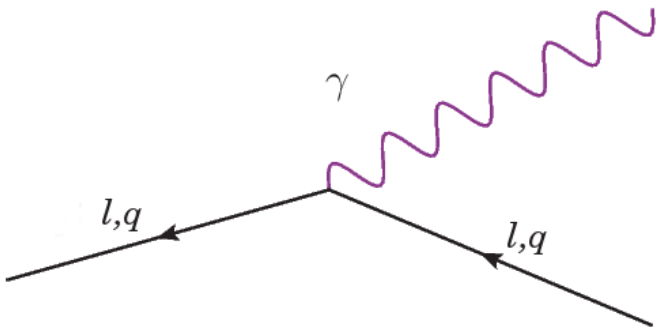

Figure 1.2: Fundamental vertex of QED.

1.1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the gauge field theory that describes the strong interaction of quarks and gluons. The strong coupling is based on colour, a new charge introduced to explain some theoretical problems of quark model [6]:

- The quark scheme predicts for the structure of the baryon Δ^{++} ($J = \frac{3}{2}$) a combination of three identical fermions u with $J = \frac{1}{2}$, in a complete symmetric ground state. Such a state is forbidden for 3 fermions by Fermi statistics;

- Some quarks combinations predicted from the theory hadn't been observed, such as: $\bar{q}\bar{q}$ or qq or a single free quark.

Strong interaction bind quarks and gluon inside hadrons but are the gravitational and the electromagnetic forces which play the dominating role in the macroscopic world. The reason is that the strong force, as well as the weak interaction, only acts at short range (subatomic distances). Where the strong force not only determines the binding of quarks and gluons inside hadrons but also of protons and neutrons inside atomic nuclei.

The restriction of the strong force to subatomic distances is a consequence of two characteristic features: "confinement" and "asymptotic freedom". Both these features may be linked to the behavior of the strong coupling constant α_S [7]:

$$\alpha_s(q^2) = \frac{\alpha_s(\mu^2)}{1 + \frac{7}{4\pi}\alpha_s(\mu^2)\ln\left(\frac{q^2}{\mu^2}\right)} \tag{1.3}$$

where q^2 indicates the transferred momentum of the interaction and μ^2 is a fixed value of q^2.

To express formula 1.3 in an easier way is frequently used the parameter $\Lambda_{QCD} = \mu^2 e^{-\frac{4\pi}{7\alpha_s(\mu^2)}}$,

$$\alpha_s(q^2) = \frac{1}{\frac{7}{4\pi}\ln\left(\frac{q^2}{\Lambda^2_{QCD}}\right)}. \tag{1.4}$$

At asymptotically large q^2 the coupling $\alpha_s(q^2)$ tends to zero and the quarks behave as free (asymptotic freedom), on the other hand confinement arise from the behavior of the coupling parameter at low $q^2(q^2 \to 0 \implies \alpha_S \to \infty)$, the trend of α_{QCD} in dependance on energy [8] is showed in figure 1.3. Confinement meaning that at large distances, or at low momentum transfer, the strong force prevents the existence of free quarks. Also in a high-energy scattering process the quarks emitted out of a hadron were never observed as free particles, but due to the hadronization process, as bundles of hadrons (jets). The confinement reflects the properties that only color singlet could be observed, so also the quarks inside the hadrons must combine to obtain a global null charge for the hadrons.

The term asymptotic freedom is used to describe the behavior of quarks at high-energy or momentum transfers, the only condition in which QCD can be described perturbatively. This feature is based on experimental observations: at high-energy scattering processes the dynamics reveal that scattering occurs between point-like and massless constituents rather than between homogeneous objects with the size of a proton.

The behavior of the strong coupling constant is due to the presence of vertex of autointeraction among the gluons. This difference with the QED is due to the charge of the mediators: in the QCD each gluon carries a combination of color and anticolor instead in the QED the photons are chargeless.

1.1.3 Weak interaction

The lightest hadrons are the three pions, $\pi^{\pm}$ and π^0, all composed by u and d quarks. The decay of the neutral pion differs from the decays of the charged ones, the π^0 is the only that can decay in two photons with a lifetime [1] of $(8.4 \pm 0.6) \cdot 10^{-17}$ while the charged pions decay to a charged and a neutral lepton, $\pi^{\pm} \to \mu^{\pm}\nu$, with a lifetime of $(2,6003 \pm 0,0005) \cdot 10^{-8}s$. This much longer lifetime has been one of the hints for the existence of a new fundamental force, weaker than the electromagnetic interaction, called indeed the weak force. This interaction is, in

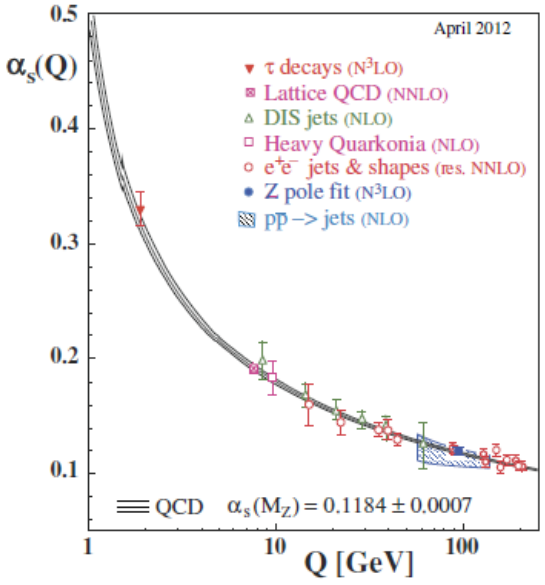

Figure 1.3: α_S versus energy scale Q. Measurements obtained from several sources.

fact, usually swamped by the strong and electromagnetic force unless these are forbidden by conservation rules, for example the processes involving neutrinos or quarks with a flavour changing can be ascribed only to the weak interactions. The flavour changing is mainly diagonal but it follows complex rules producing linear combination of mass eigenstate. This happens because the eigenstate of the weak interaction are a mixing of the eigenstate of mass so a quark u will decays in a quark d', mixing of the mass eigenstate d,b,s quarks. The mixing is represented using the Cabibbo Kobaiashi Maskawa (CKM) matrix, usually reported as:

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

The observable effect of the mixing is a different probability of decay of the quark inside the same generation or among different generations, this probability is proportional to the square of the matrix elements.

The elements of CKM have been calculated experimentally, the latest results [1] are reported in table 1.5.

$$M_{CKM} = \begin{pmatrix} 973.83^{+0.24}_{-0.23} & 227.2 \pm 1.0 & 3.96 \pm 0.09 \\ 227.1 \pm 1.0 & 972.96 \pm 0.24 & 42.21^{+0.10}_{-0.8} \\ 8.14^{+0.32}_{-0.64} & 41.61^{+0.12}_{-0.78} & 999.100^{+0.034}_{-0.004} \end{pmatrix} \cdot 10^{-3} \qquad (1.5)$$

The matrix is unitary and approximately diagonal, so the decays inter generation are dominant with respect to decays over different generation. Horizontally decays are not allowed for the Glashow IIliopoulos Maiani (GIM) [9] mechanism that assures a natural suppression of flavour changing neutral current (FCNC) at the tree level. FCNC, consequently, are rare processes and difficult to be observed since they could be realized only in higher order diagrams, the leading contributions result from the one-loop diagrams in figure 1.4.

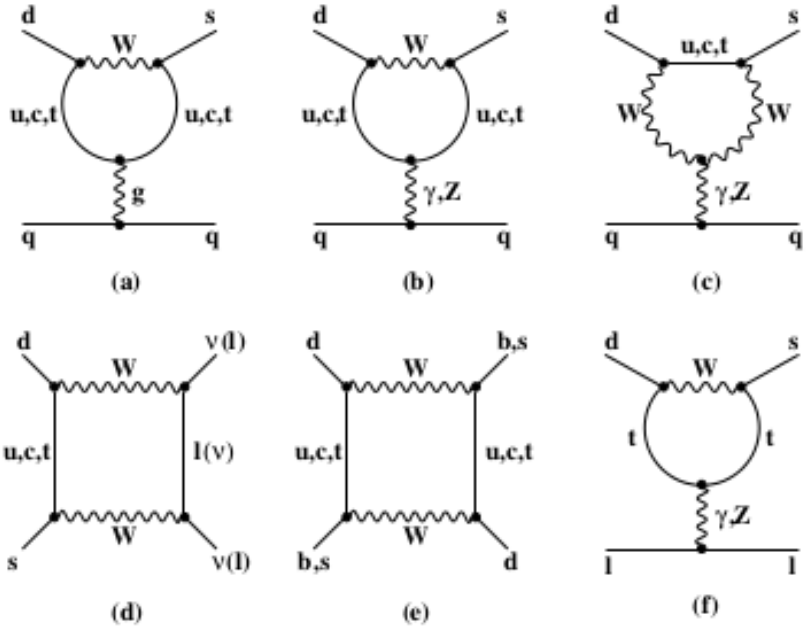

Figure 1.4: Penguin and box diagrams, leading process in FCNC.

1.2 Top quark properties

The existence of the top quark was postulated many years before its experimental evidence, in particular after the discovery the b-quark, in 1977. In the following years, indirect evidence on the top quark existence was obtained from limits on FCNC and from indications that the b-quark must be the member of an isospin double; its weak isospin T3 = $-\frac{1}{2}$ was determined from measurements at LEP and SLC (Stanford Linear Collider), leading to the conclusion that the postulated partner of the b-quark should exist with T3 = $\frac{1}{2}$.

The discovery was announced in 1995 by the CDF and D0 [10] [11] experiments at the proton antiproton collider Tevatron (Fermilab, Chicago) and from that moment the top quark has remained in the focus of particle physics research. There are several peculiar characteristics of the top quark that conduct to consider it one of the most interesting field of research: first of all its mass, $m_t = 173.07 \pm 0.52 \pm 0.72$ GeV, that is the heaviest of all known quarks and particles. The top mass is around 40 times heaver than the b-quark and is comparable to the one of a Rhenium atom (atomic number Z = 75); visual comparison between the masses of SM particles is shown in figure 1.5. Because of its mass the top quark has a large coupling with

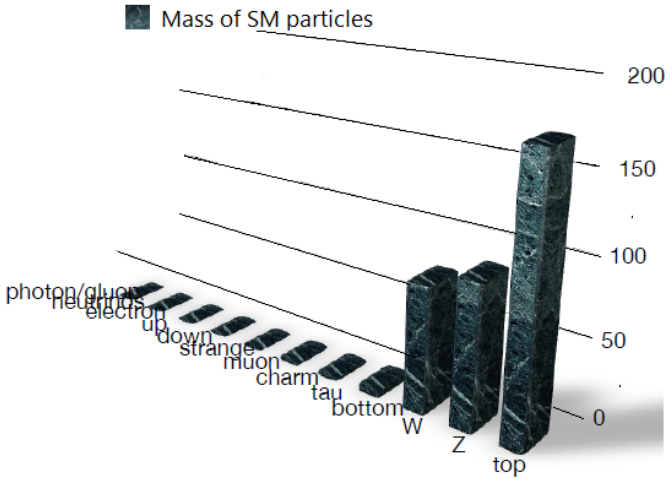

Figure 1.5: Comparison between mass of particles belonging to the Standard Model. The Higgs mass is not showed in this figure.

the Higgs: its Yukawa coupling, $y_t = m_t/v$ (where $v \simeq 246$ GeV is the vacuum expectation value) its near to one. This observation could imply that the top quark may play an important role in electroweak symmetry breaking. Furthermore the top quark appears in higher order loop diagrams of the electroweak theory, so the measurement of its mass is crucial in this theory.

The top quark plays an important role in many scenarios for new physics beyond the SM (BSM)[12]. Many BSM theories propose that the SM could be a low energy theory resulting from the breaking of extended gauge groups, formulated following the example of the electroweak spontaneous symmetry breaking of the group SU(2)xU(1) that leads to the generation of massive bosons. These models lead to additional scalar Higgs bosons and other massive gauge bosons, particle called Z', that should interact strongly with the top quark and could be seen as a

peak in the invariant mass spectrum in the $t\bar{t}$ pair. To introduce gravity in the SM other BSM have predicted the existence of extra dimensions to space time as for example the Randall-Sundrum, the Kaluza-Klein gravitons (the quanta of gravity) can propagate in these dimensions. Also these particles could be seen in the $\bar{t}t$ invariant mass spectrum as a series of wide resonances.

Besides models that predict the existence of new particles which decay predominantly to quark pairs, new particles may also be produced in top quark decays as, for example, a charged Higgs boson in the $t \to H^+ b$ channel. In addition differential measurements of the properties of the top quark and its interactions may reveal effects from new physics.

1.2.1 Measurement of top quark mass

The top quark mass is a fundamental parameter of the SM so from its discovery many measurements both direct than indirect have been done to evaluate its value.

Indirect constraints on m_t can be obtained from precision measurements of the parameters of the electroweak theory. It is possible to express the W boson mass as a function of the electromagnetic coupling $\alpha(M_Z^2)$ and two parameter of electroweak theory: the Fermi constant G_F and the mixing angle θ_W.

$$m_W^2 = \frac{\pi\alpha(M_Z^2)/\sqrt{2}G_F}{\sin^2\theta_W(1-\delta r)} \tag{1.6}$$

where δr contains contributions from higher order electroweak loop diagrams involving the top quark which depend quadratically on m_t. The most recent indirect constraint on m_t based on electroweak precision measurements is $m_t = 179.7^{+11.7}_{-8.7}$ GeV [13].

Direct measurements have been performed both at Tevatron and at LHC, the main results from this collaborations are reported in table 1.2 [12].

All the direct measurements are in good agreement and are in the expected range with respect the indirect values.

The world average top mass value, reported on particle data group and calculated using both LHC than Tevatron measurements is [1]:

$$m_t = 173.07 \pm 0.52 \pm 0.72 GeV. \tag{1.7}$$

This measurement has a precision of $\simeq 0.6\%$, this makes the mass of the top quark not only the highest but also the one known with better precision among all the

quarks.

$m_t(GeV/c^2)$	Source	$\int Ldt$	Channel
$174.94 \pm 1.14 \pm 0.96$	D0 Run II	3.6	l+jets
$172.85 \pm 0.71 \pm 0.85$	CDF Run II	8.7	l+jets
$173.93 \pm 1.64 \pm 0.87$	CDF Run II	8.7	Missing E_T+jets
$172.5 \pm 1.4 \pm 1.5$	CDF Run II	5.8	All jets
$172.31 \pm 0.75 \pm 1.35$	ATLAS	4.7	l+jets
$173.09 \pm 0.64 \pm 1.50$	ATLAS	4.7	ll
$174.9 \pm 2.1 \pm 3.8$	ATLAS	2.04	All jets
$173.49 \pm 0.43 \pm 0.98$	CMS	5.0	l+jets
$172.5 \pm 0.4 \pm 1.5$	CMS	5.0	ll
$173.49 \pm 0.69 \pm 1.21$	CMS	3.54	All jets
$173.20 \pm 0.51 \pm 0.71$	CDF,D0 (I+II)	< 8.7	publ. or prelim. res.
$173.29 \pm 0.23 \pm 0.92$	ATLAS,CMS	< 4.9	publ. or prelim. res.

Table 1.2: Measurement of top mass at LHC and Tevatron. The Tevatron average is a combination of published Run I and preliminary or pub. Run-II meas. The LHC average includes both published and preliminary results.

1.2.2 Decays

From the values present in the CKM matrix it is evident that the top decays $t \to Wb$ has almost the 100 % branching fraction. The diagrams of this decay is shown in figure 1.6.

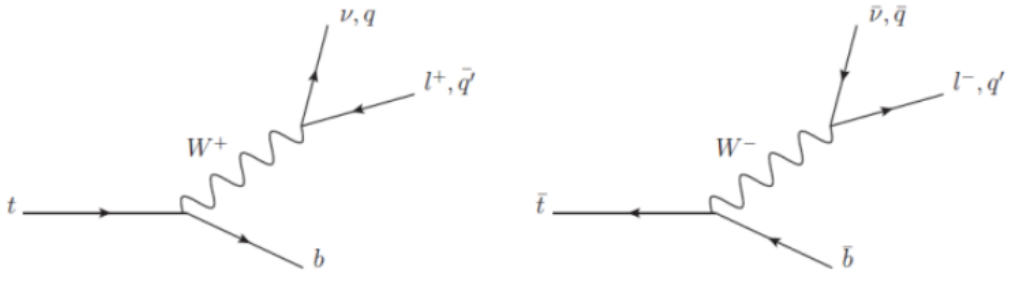

Figure 1.6: Feynman diagrams illustrating the leading decays of top and antitop.

The lifetime of the top quark is very short, the last estimation obtained as medium of several measurements [1] is $\Gamma_t = 2.00 \pm 0.5$ GeV that leads to a lifetime of $\simeq 10^{-25} s$, in good agreement with SM expectations. Due to this very short

lifetime the top quark is the only one that can decay before to hadronize, because the typical time for hadron formation is $\sim 3 \cdot 10^{-24} s$, an order of magnitude larger than the top lifetime. A consequence of this short lifetime is that *toponium*, the bounded state of top antitop pair, cannot exist. The top quark, consequently, is the only quark that can be studied as *"bare"* and many studies on it may be derived by studying its decay products.

After its formation the b quark will hadronize to a B meson or baryon (and other particles), that will decay with process like $b \to Xc$ and $b \to Xu$. Both these processes undergo a suppression factor, as can be seen from CKM, consequently the lifetime of the B hadrons is quite long ($\sim 10^{-12}$s). This characteristic of the B hadrons is exploited by the reconstruction algorithms used to identify them looking for a second decay vertex. The shower of hadrons and decay products of the b quark is seen in the detector as a deposit of energy in the calorimeter, called jet.

The W boson can decay in two different ways, in two leptons $W \to l\nu_l$ or in two quarks $W \to qq'$: in the first case the final state contains a lepton and a neutrino, seen as missing energy by the detector, instead for the second one with the presence of 2 jets.

The decays of the top quark pair, studied in this thesis, are classified according to the decay of the W bosons.

- **Di-lepton channel**: both the W bosons decay to leptons. This channel has a cleaned final state but with many missing energy in the detector. The branching ratio (BR) of this channel is only 6.45%, and these events consists in a sources of background for the analysis presented in this thesis.

- **Hadronic channel**: both the W bosons decay in hadrons. This channel has the highest BR$(46, 2\%)$ but is really difficult to study due to the six jet final state. These processes are a high sources of background.

- **Semileptonic channel**: one of the W decays to leptons and the other to quarks. The branching fraction is $43, 5\%$, approximately evenly split for W-boson decays to electron, muon and tau. The presence of an high p_T lepton makes this final state easier to identify from the trigger system with respect to the fully hadronic channel. This characteristic and the high BR made of this decay mode the *golden-channel* and the one used for the study described in this thesis.

In figure 1.7 are summarized the different branching ratio for all channels of $t\bar{t}$ decay.

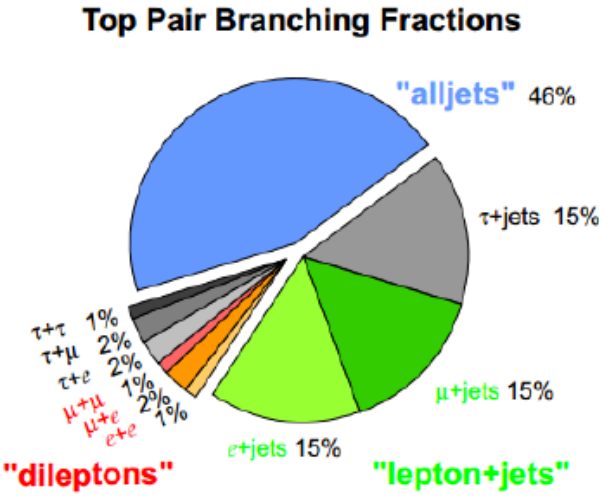

Figure 1.7: The pie chart illustrates the BR of the three decay channels of the $t\bar{t}$. The contribution of each lepton generation inside the channel is reported separately.

1.3 Top quark production

In a high energy proton-proton collider the top quark dominant production mechanism is $t\bar{t}$ pair, faced in this thesis, even if, with low probability, it can be produced also as single top. The two production mechanism are quite different since they involve two distinct interactions.

1.3.1 Top quark pair production

The SM predicts that the dominant production mechanism for top quark pair is mediated by the strong interaction. At the LHC center of mass energy $\sqrt{s} = 8$ TeV, around 80% of the initial state interaction is due to gg contribution [14], while the remainder is mostly due to $q\bar{q}$; this situation is exactly the opposite of the Tevatron collider because of the lower center of mass energy ($\sqrt{s} = 1.98$ TeV)and the different initial state (proton and antiproton). The leading-order (LO) diagrams of the $t\bar{t}$ prodution, shown in figure 1.8, involve only gg and $q\bar{q}$ in the initial states while processes that involve a mixed quark-gluon or antiquark-gluon appears only at next to leading order (NLO), shown in figure 1.9.

The $t\bar{t}$ pairs production cross section is obtained by the relation:

$$\sigma = \frac{N_{obs}}{\int Ldt \cdot \epsilon} \tag{1.8}$$

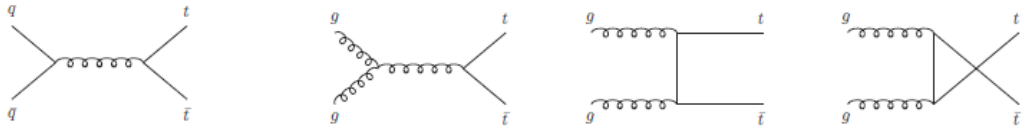

Figure 1.8: Leading order process in top pair production. The first diagram show the most common production process at Tevatron while the other three diagrams illustrates leading process at LHC.

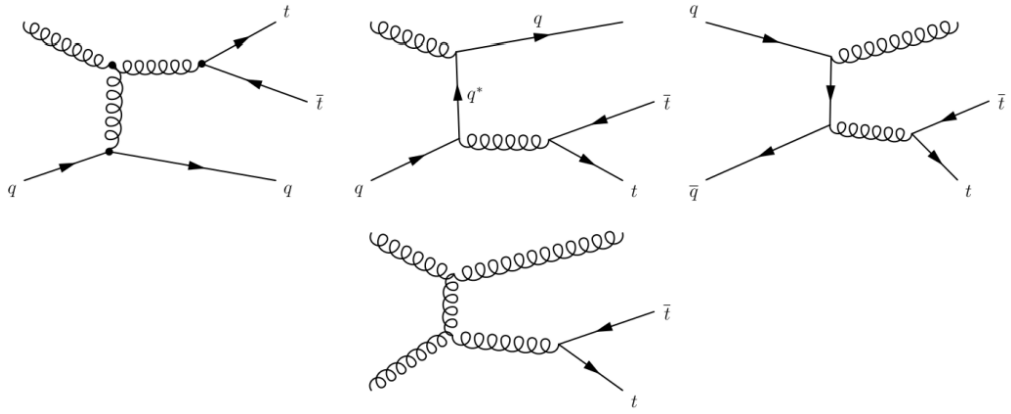

Figure 1.9: NLO process in top pair production.

where N_{obs} is the number of events observed, L is the luminosity and ϵ is the efficiency of the detector. In equation 1.8 the luminosity is integrated overall the period of the data acquisition and it is called integrated luminosity. The instantaneous luminosity depends on the features of the collider, and may be calculated as:

$$L = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{1.9}$$

where n_1 and n_2 are the numbers of particles contained in each bunch (at LHC there are about 10^{11} protons per bunch), f is the frequency of round ($f_{LHC} \simeq 11kHz$), σ_x and σ_y are the beam particle distribution along two orthogonal axis with respect to the beam direction ($\sigma_{LHC} \simeq 50\mu m$). The cross section has the dimension of an area and is usually measured in barn (b) ($1b = 10^{-24}cm^2$). The total inclusive cross section is the sum of all the possible elastic and inelastic scattering processes; the exclusive one is, instead, related to a particular process such as ($pp \to t\bar{t}$) and is related to the probability for that process to be produced in a pp collision. From a theoretical point of view the differential cross section is given by:

$$d\sigma = \frac{|M|^2}{F}dQ \tag{1.10}$$

where M is the invariant amplitude of a process, F is the particles flux and dQ is the element of phase space.

Because the parton scattering process is fairly well separated in time from the hadronization one, it is possible to factorize these two contributions in the evaluation of the production cross section (this factorization is possible only at high energy).

The uncertainty principle allow to calculate the time interval within which parton scattering occurs as $\tau \sim \frac{\hbar}{p_T c}$. The time of hadronization can be calculated supposing a characteristic distance R for the confinement (R $\sim$ 0.5 fm is a typical hadron 'size') as $T \sim \frac{R p_T}{mc^2}$ (where $\frac{p_T}{mc}$ being the Lorentz dilation in transforming from the rest frame of the parton to the centre-of-mass frame of the system). Thus as p_T increases the processes of parton scattering and hadron production become separated by longer times. The σ, consequently, could be factorized for the hard scattering and the hadronization.

The inclusive production cross section for the process $pp \to t\bar{t}$, as a function of m_t and the center of mass energy squared of the collider $s = 4E^2_{beam}$, can be expressed as a convolution of PDF and, a partonic cross sections $\hat{\sigma}$ [15] in the following way.

$$\sigma_{pp\to t\bar{t}}(s, m_t) = \sum_{i,j=q,\bar{q},g} \int dx_i dx_j f_i(x_i, \mu_f^2) f_j(x_j, \mu_f^2) \cdot \hat{\sigma}_{ij\to t\bar{t}}(\hat{s}, m_t, \mu_f, \mu_r, \alpha_s) \quad (1.11)$$

The sum runs over all quarks, antiquarks and gluons contributing, x_i are the parton momentum fractions with respect to the proton momentum (calling p_a the parton momentum and p_A the hadron momentum $x_a = p_a/p_A$), $f_i(x_i; \mu_f^2)$ are the probability density function of the partons momentum inside the proton, μ_r and μ_f are the factorization and renormalization scales, α_s is the strong coupling and $\hat{s}$ is the partonic center of mass energy given by:

$$\hat{s} = (p_a + p_b)^2 \simeq 2p_a \cdot p_b = 2x_a x_b p_A \cdot p_B \simeq x_a x_b s. \quad (1.12)$$

The scattering process at high momentum transfer can be successfully described by the quantum chromodynamics in a perturbative regime since the top quark mass m_t is much larger than the Λ_{QCD} and so can be used a perturbative approach.

In figure 1.10 it is reported the behaviour of the $t\bar{t}$ production cross section versus energy.

This thesis presents the measurement of the differential cross section in $t\bar{t}$ pro-

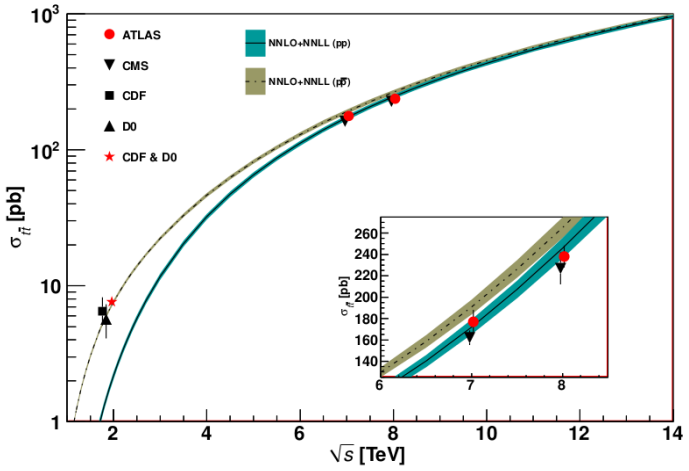

Figure 1.10: Measured and predicted $t\bar{t}$ production cross sections from Tevatron energies in $p\bar{p}$ collisions to LHC energies in pp collisions.

duction with respect to p_T^t of the hadronic top. The differential cross section may be expressed as a function of ϵ and of the events observed in certain interval i of the variable studied:

$$\frac{d\sigma}{dp_T^{t\bar{t}}} = \frac{N^i}{\epsilon^i \Delta^i_{dp_T^t} \int Ldt} \tag{1.13}$$

the superscript i indicates a bin of the p_T^t distribution and the Δ^i the bin width. A review of latest results in experimental measurements of total and differential cross section in $t\bar{t}$ production will be presented in section 1.3.4.

1.3.2 Single top production

The single top production is based on electroweak interaction and usually involves Wtb vertex, that is almost exclusively since $|V_{tb}| >> |V_{td}|, |V_{ts}|$ as can be seen in the CKM matrix. The diagrams for single top production are reported in figure 1.11.

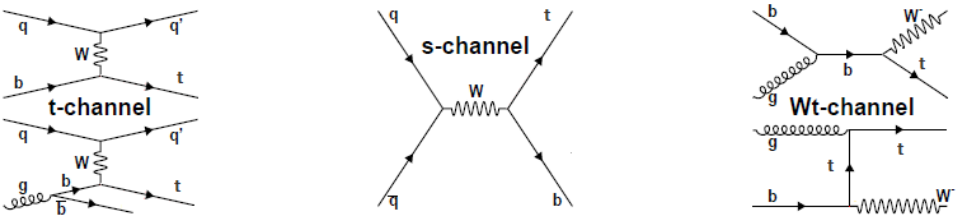

Figure 1.11: The diagram on the left shows t-channel single-top production, the diagram in the middle shows s-channel single-top production, and the diagram on the right shows Wt-channel single-top production.

Channel	8 TeV (pb)
t-channel	$87.8^{3.4}_{-1.9}$
s-channel	5.6 ± 0.2
Wt	22.4 ± 1.5

Table 1.3: Expected single top quark production cross sections in different channels at a center-of-mass energy of 8 TeV, given by approximate NNLO [16] [17] [18]

Both at Tevatron and LHC the dominant channel is the t-channel while the s-channel gives the smallest contribution to the single top production. The associated production of a top quark with a W boson, figure 1.11c, has a very small cross section at the Tevatron, but is significant at the LHC. At $\sqrt{s} = 7$ TeV, the ratio of the t-channel single top quark (and anti-quark) production cross section to the $t\bar{t}$ cross section $\sigma_{t+\bar{t}}(t - ch :)/\sigma_{t\bar{t}}$ is around 40%.

Single top quark production could be interesting for various reasons, non-standard couplings would indicate the presence of contributions from new physics and it could allow to directly measure the CKM matrix elements, however the single top represents one of the main background sources for the analysis of $t\bar{t}$.

The expected cross section for the three channels at 8 TeV are reported in table 1.3.

In figure 1.12 is plotted the t-channel combined cross section (single top + single antitop) versus energy for CDF, D0, ATLAS and CMS experiments. Several measurements of the single top quark production cross section in all three channel were performed by ATLAS using various techniques and data samples. At the LHC, the t-channel mode of single top quark production has the cleanest signature with a light quark jet recoiling against the top quark, the latest results obtained at $\sqrt{s} = 8$ TeV by ATLAS with L=5.8 fb^{-1} of data taken in 2012 [19] and by CMS based on a data sample corresponding to an integrated luminosity of L=19.7 fb^{-1} [20] are:

$$\sigma_t^{ATLAS}(t - channel) = 95 \pm 2(stat) \pm 18(syst) \quad pb.$$

$$\sigma_t^{CMS}(t - channel) = 83.6 \pm 2.3(stat.) \pm 7.4(syst) \quad pb.$$

The tW -channel interferes at NLO QCD with top quark pair production and several methods have been implemented in current MC generators to allow an unambiguous signal definition. Both ATLAS, at $\sqrt{s} = 7$ TeV, than CMS, at $\sqrt{s} = 8$ TeV, have

an evidence of production and measured a cross section [21]:

$$\sigma_t^{ATLAS}(Wt-channel) = 16.8 \pm 2.9(stat) \pm 4.9(syst) \quad pb.$$

$$\sigma_t^{CMS}(Wt-channel) = 23.4 + 5.5 - 5.4 \quad pb.$$

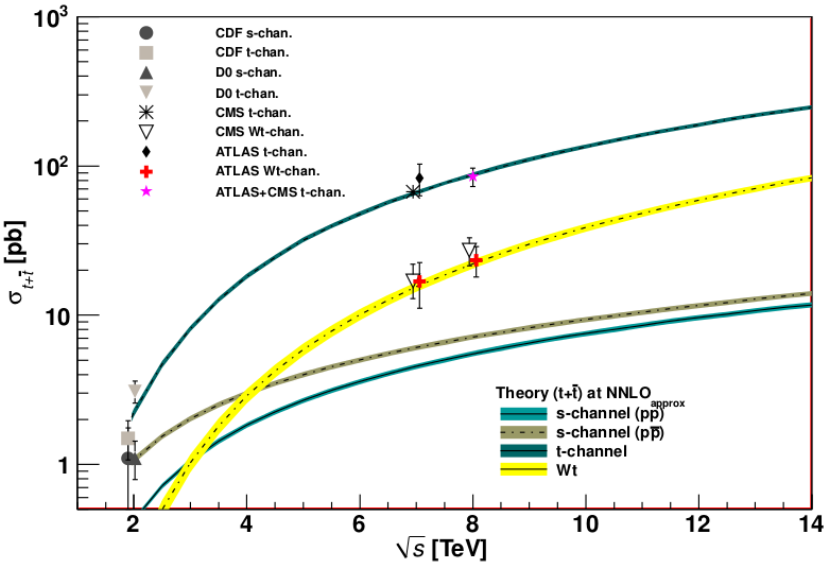

Figure 1.12: Measured and predicted single top production cross sections from Tevatron energies in pp collisions $\sqrt{s} = 1.96$ TeV to LHC energies $\sqrt{s} = 7$ TeV in pp collisions.

1.3.3 Parton distribution function

An important contribution to the cross section of a particular process is given by the PDF, as can be seen from equation 1.11.

The PDF represents the probability that a parton i with a fraction of momentum x_i is present in the proton. At LHC the energy of the hard scattering at parton level is not 8 TeV but it depends on the energy carried from each parton, as can be seen in equation 1.12, consequently a deep knowledge of the PDFs is important in all data and background simulations. Also the difference between the leading process in $t\bar{t}$ production at LHC and Tevatron can be ascribed to the PDF. The typical value of momentum fraction for the $t\bar{t}$ production is $x = 2mt/\sqrt{s}$, at LHC (if $\sqrt{s} = 7$) this mean $x \sim 0.05$ otherwise at Tevatron the medium value for x was ~ 0.2. The figure 1.13 shows how this difference reflects in different amount of gluon and quark population inside the hadrons; at low x the probability to have gluons

is above the one of quarks. The (i.e., process independent) proton PDF $f_i(x_i, \mu_f^2)$

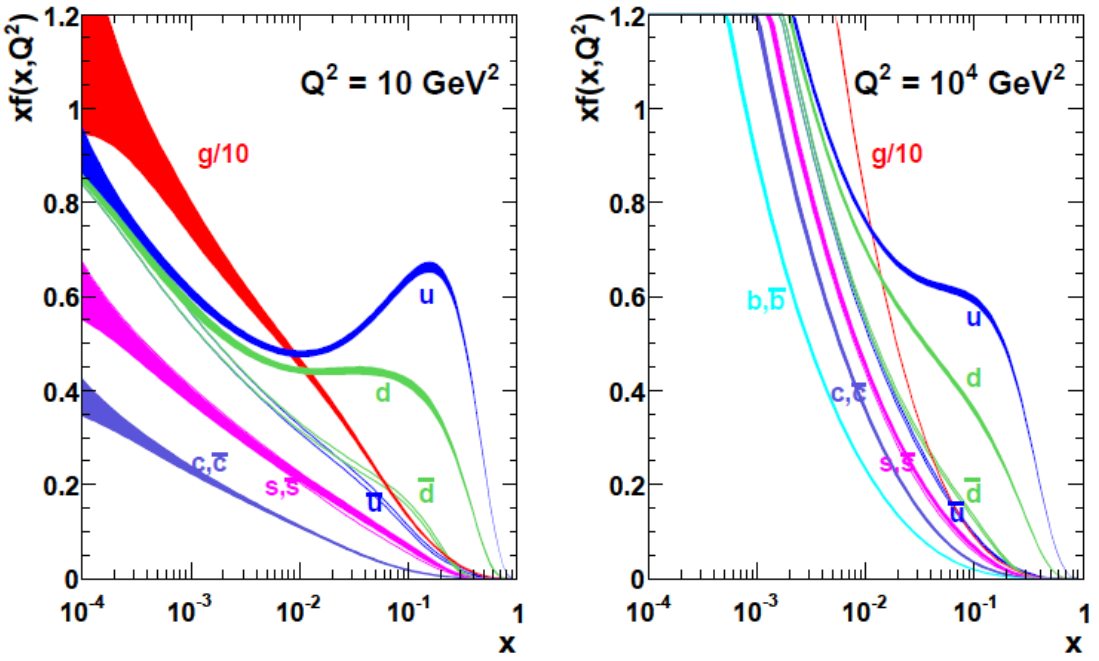

Figure 1.13: MSTW2008NLO PDF set prediction for quarks and gluons momentum distribution inside the proton [22] at different Q^2 values.

are determined by several groups [22] [23] [24] [25] from global fits to experimental data on deep-inelastic scattering but also on jet and heavy quark production at hadron colliders. There are several PDF sets that may be used in the analysis and, for a given order in perturbation theory, the main differences between these PDF sets arise from the choice of the included data, the treatment of systematic uncertainties, the parametrization at the starting scale, the chosen heavy-quark scheme, the values of the quark masses and the one of $\alpha_S(m_Z)$. At present, all PDF sets exhibit a significant uncertainty on the gluon density at medium-high parton momentum fractions.

As can be seen in equation 1.11, PDF and cross section calculation depend on the μ_r and μ_f coefficients: the dependence on μ_r arises in particular from the definition of the renormalized coupling α_S and the dependence of the partonic cross section and the PDF on μ_f arise from absorbing uncanceled collinear initial state singularities into the PDF. The renormalization and factorization scales are typically set at the hard scale of the process which for the $t\bar{t}$ production is usually set to the top mass. Other possible choices are $\mu_r = \mu_f = m_t/2$ and $\mu_r = \mu_f = 2m_t$, used to calculate systematic errors introduced from the choose of normalization and

factorization scales. Also the choose of the PDF set for the analysis is a source of systematic error. Some studies done using different PDF sets will be illustrated in Chapter 4.

1.3.4 $t\bar{t}$ cross section measurements

The total $t\bar{t}$ cross section was measured for the first time in $p\bar{p}$ collisions at the Tevatron collider ($\sqrt{s} = 1.96$ TeV) [26] [27]:

$$\sigma_{t\bar{t}}^{D0} = 7.56^{+0.63}_{-0.56} \quad pb$$

$$\sigma_{t\bar{t}}^{CDF} = 7.50 \pm 0.48 \quad pb$$

obtained as the combination of the various measurements in all the decay channel; both are in good agreement with the SM predictions:

$$\sigma_{t\bar{t}}^{theor} = 7.16^{+0.20}_{-0.23}\ pb \tag{1.14}$$

CMS and ATLAS performed the of cross section measurements both at $\sqrt{s} = 7$ TeV than at $\sqrt{s} = 8$ TeV. The $\sigma_{t\bar{t}}$ is strongly dependent on the center of mass energy so the cross section expected and measured at LHC is higher than at Tevatron, as can be seen in figure 1.10.

The values obtained from ATLAS ($L_{int} = 0.70 fb^{-1}$) [28] and CMS ($L_{int} = 36 pb^{-1}$) [29] at $\sqrt{s} = 7$ TeV are:

$$\sigma_{t\bar{t}}^{ATLAS} = 177 \pm 6(stat.)^{+17}_{-14}(syst)^{+8}_{-7}(lum.) \quad pb$$

$$\sigma_{t\bar{t}}^{CMS} = 168 \pm 18(stat.) \pm 14(syst) \pm 7(lumi.) \quad pb$$

These measurements are obtained combining all three $t\bar{t}$ channels decay and resulted compatible with the theoretical expectation:

$$\sigma_{t\bar{t}}^{theor} = 165^{+11}_{-16}\ pb \tag{1.15}$$

In figure 1.14 measurements are compared with theoretical expectation at NL0 and NNLO, where the last one resulted the best agreement. A summary of values of $\sigma_{t\bar{t}}$ obtained at $\sqrt{s} = 7$ TeV in all channels is reported in figure 1.15. In figure 1.14 b ATLAS and CMS measurements are compared with NNLO, the error to theoretical

expectation is calculated from PDF and varying normalization and factorization scale.

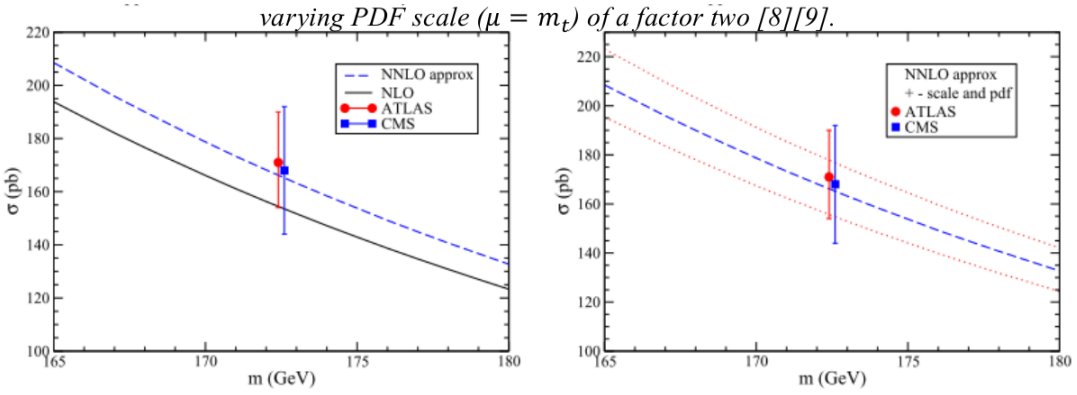

Figure 1.14: LHC $\sigma_{t\bar{t}}$ cross section measurements compared, on the left plot, with NLO and NNLO prediction and, on the right, with NNLO scale variation uncertainties estimated varying scale ($\mu_r = \mu_f = m_t$) of a factor two.

Preliminary measurements at $\sqrt{s} = 8$ TeV in the semi leptonic channel have also been performed in ATLAS ($L = 5.8 fb^{-1}$) [30] and CMS ($L = 2.8 fb^{-1}$) [31]:

$$\sigma_{t\bar{t}}^{ATLAS} = 241 \pm 2(stat.) \pm 31(syst) \pm 9(lum.) \quad pb$$

$$\sigma_{t\bar{t}}^{CMS} = 228.4 \pm 9.0(stat.)^{+29.0}_{26.0}(syst) \pm 10.0(lumi.) \quad pb$$

also these in agreement with the theoretical expectations:

$$\sigma_{t\bar{t}}^{theor} = 238^{+22}_{-24}\ pb \tag{1.16}$$

The measurements obtained at $\sqrt{s} = 8$ TeV may be improved in precision using the total luminosity ($\sim 20 fb^{-1}$).

1.3.5 $t\bar{t}$ differential cross section

The large abundance of top quark pair production at LHC allows to measure not only the total cross section $\sigma_{t\bar{t}}$, but also differential cross sections $d\sigma_{t\bar{t}}/dX$ where X are relevant kinematic variables as the mass or the transverse momentum of the top (or the top pair). These distributions may validate MC models as well as explicit higher order QCD calculations of top quark production; it is moreover possible to evaluate deviations in the distribution that could be signal of new physics.

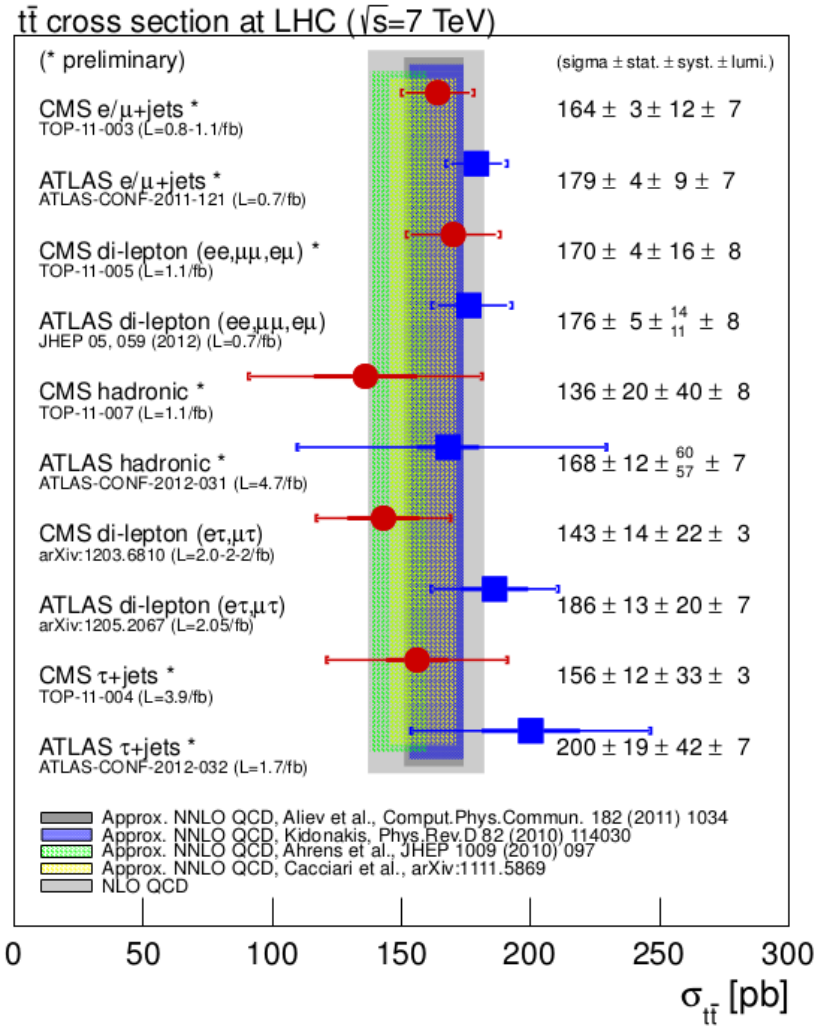

Figure 1.15: Summary of the most precise measurements of $\sigma t\bar{t}$ per decay mode and experiment, compared with several theory predictions at NLO and approximate NNLO QCD.

Cross sections may be quoted either after extrapolation to the full phase space (as done in the case of the total cross section), or only within the kinematic range where the decay products are measured within the detector (so-called visible phase space); this is the case of the analysis presented in this thesis where is analyzed a very energetic phase space region where the standard model predictions are mostly extrapolated.

In order to facilitate comparisons with theoretical models several corrections need to be applied to the observed spectra: efficiency, acceptance, and the unfolding procedure that corrects for migration effects due to the detector. The cross section could be defined at parton level if the analysis considers partons before their hadronization otherwise at particle level if the analysis involves only objects visible

by the detector within a reduced, or fiducial, phase space. In the second case only the detector response correction is used, so the results on the cross section have a low model dependence. The particle level definition is closer to what is measured experimentally and can be easier compared with MC simulations, instead of the parton level definition that may be needed in order to compare cross section with fixed order of QCD calculations.

The first measurements of the differential cross section were performed at Tevatron both by D0, with an integrated luminosity of 9.7 fb^{-1}, and by CDF, with an integrated luminosity of 2.7 fb^{-1}, reported in figure 1.16 and 1.17 [32].

Several measurements with increasing statistic have been performed by ATLAS and CMS on different decay channels; their results are expected to have an higher precision and reliability because of the the high cross section in $t\bar{t}$ production at the LHC energy. In figure 1.18 are presented the distributions obtained by ATLAS [33] with data collected at a center of mass energy of $\sqrt{s} = 7$ TeV in 2011 with a statistics corresponding to an integrated luminosity of $L = 4.7fb^{-1}$ in the lepton+jets channel. In figure 1.19 are presented results obtained by CMS [34] from data collected in 2011 corresponding to an integrated luminosity of $L = 5.0fb^{-1}$ in dilepton channel at $\sqrt{s} = 7$ TeV. All differential cross section results presented, are evaluated at the parton level, otherwise the analysis described in this thesis has been performed at particle level.

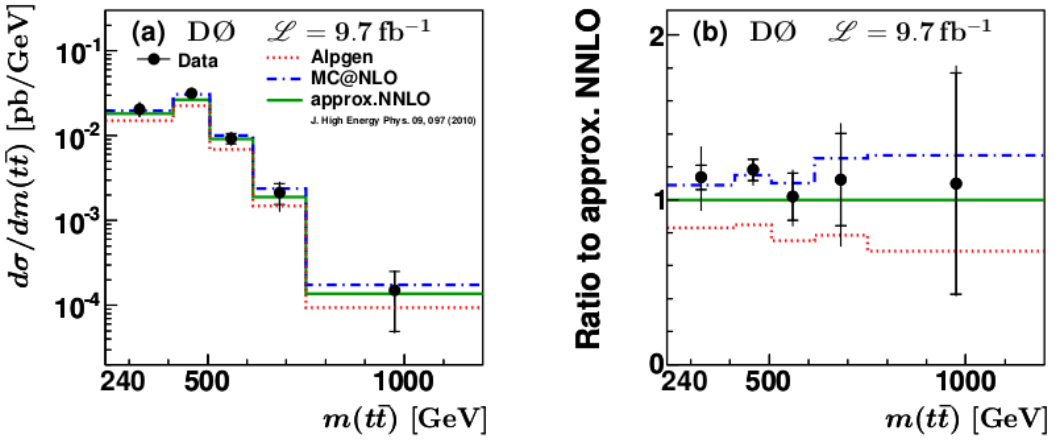

Figure 1.16: On the left is reported the differential cross section as a function of $m_{t\bar{t}}$ measured by D0, the inner error bar corresponds to the statistical uncertainty, whereas the outer error bar reflects the total uncertainty. On the right is reported a ratio of the cross section and the predictions to the NNLO.

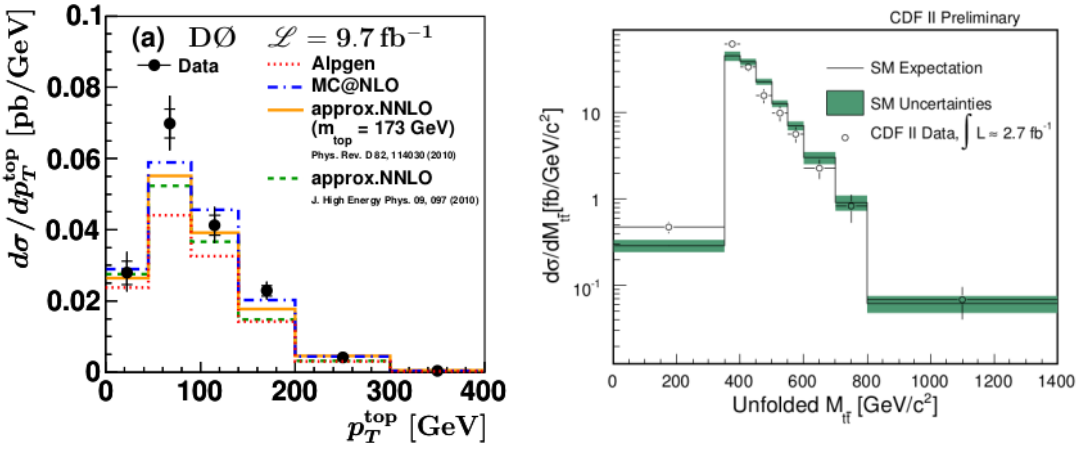

Figure 1.17: On the left is reported the differential cross section as a function of p_t^T measured by D0, the inner error bar corresponds to the statistical uncertainty, whereas the outer error bar reflects the total uncertainty. On the right are reported the CDF measurements of the differential cross section as a function of $m_{t\bar{t}}$ compared to the standard model expectation. The SM uncertainty reflects all systematic uncertainties, except for the luminosity.

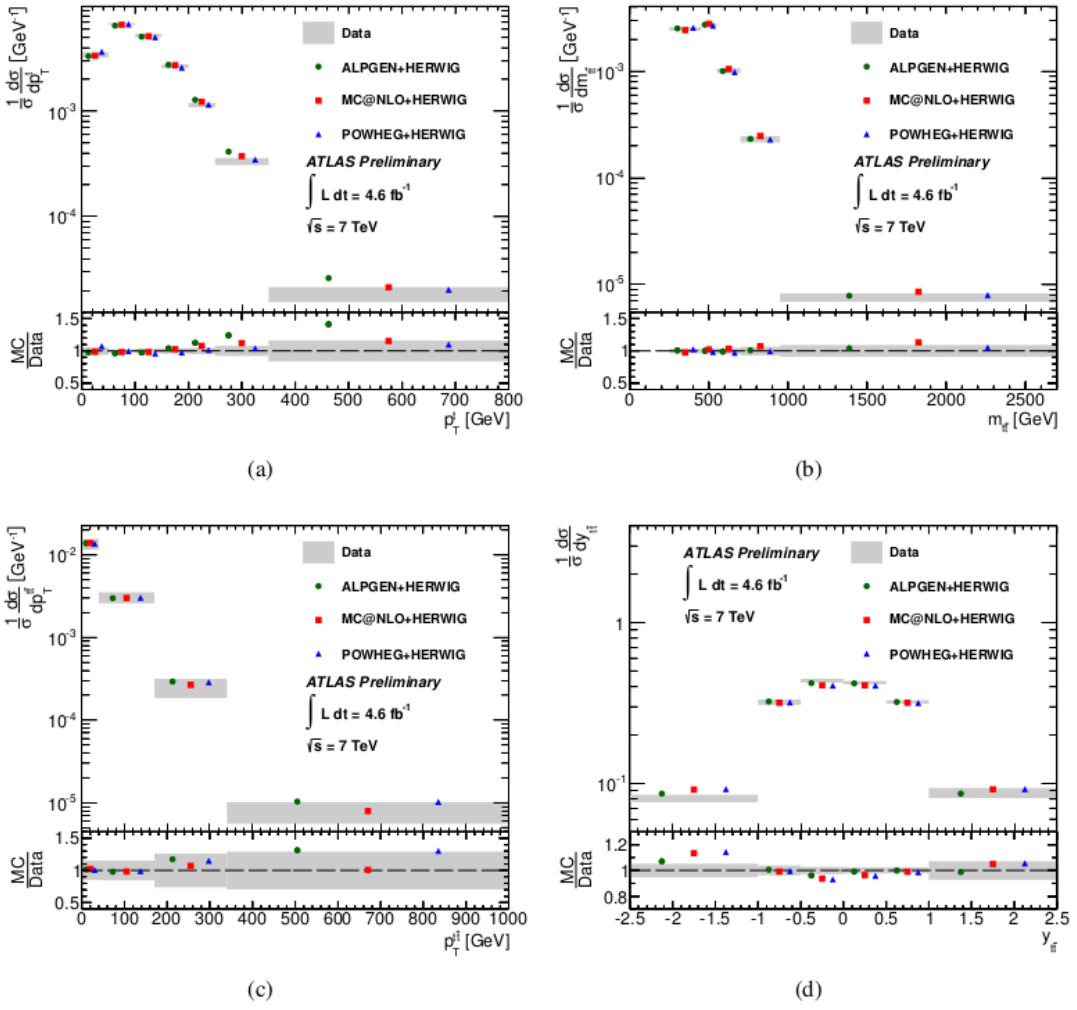

Figure 1.18: Normalized differential cross-sections with respect to p_T^t(a), $m_{t\bar{t}}$ (b), $p_T^{t\bar{t}}$ (c), $y_{t\bar{t}}$ (d) at $\sqrt{s} = 7TeV$. Comparisons to several generators are shown with points corresponding to Alpgen+Herwig, MC@NLO+Herwig and Powheg+Herwig. The lower part of each figure shows the ratio of the generator predictions to data. The gray bands indicate the total uncertainty on the data in each bin.

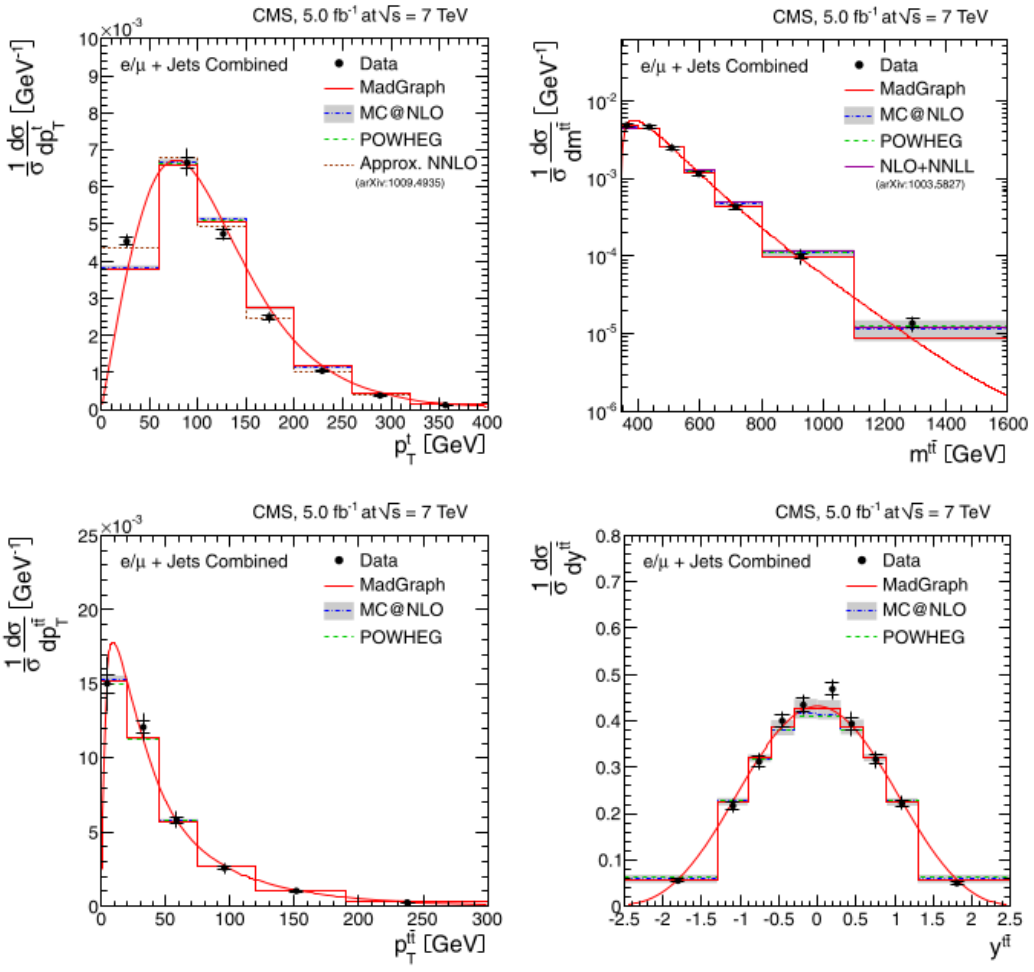

Figure 1.19: ATLAS preliminary results of normalized differential $t\bar{t}$ production cross section as a function of p_T^t (top left) and $m_{t\bar{t}}$ (top right) of the top quarks, the $p_T^{t\bar{t}}$ (bottom left) and the $y_{t\bar{t}}$ (bottom right). The inner (outer) error bars indicate the statistical (combined statistical and systematic) uncertainty. The measurements are compared to predictions from MADGRAPH, POWHEG, and MC@NLO, and to NLO + NNLL and approximate NNLO calculations.

Chapter 2

LHC and the ATLAS detector

The CERN, the European Organization for Nuclear Research, is the world largest research center and it is located astride the Franco-Swiss border, near Geneva. It was founded in 1954 and now comprises 21 member states that cooperate to probe the fundamental structure of the universe. This is possible through a succession of machines that accelerate particles to increasingly higher energies. The last and newest of the CERN accelerator complex is LHC, Large Hadron Collider, the world' s largest and most powerful particle accelerator. Along the accelerator ring are positioned four particle detectors: ATLAS, CMS, ALICE e LHCb.

2.1 LHC

LHC is situated about 100 m under Geneva and it consists of a 27-kilometer ring of superconducting magnets with 16 radiofrequency cavities to boost the energy of the particles, figure 2.1. Inside the accelerator two high energy particles beams travel in opposite directions in separate beam pipe kept at ultrahigh vacuum (10^{-10}torr). The beams collide in four points, in correspondence of the four detectors. LHC is designed to accelerate protons up to an energy of 7 TeV producing collisions at a center-of-mass energy of 14 TeV at a maximum of an instantaneus peak luminosity of L $L = 10^{34} cm^{-2} s^{-1}$. This high performance will be reached after a long shut down period in 2013-2014, so far it worked at a lower energy. LHC started operations in 2008 and during the period 2010-2011 it foresees a beam with a center of mass energy of 7 TeV and a peak luminosityof $L = 4 \cdot 10^{33} cm^{-2} s^{-1}$. In 2012 the center of mass energy has been increased to 8 TeV and to a luminosity peak of

The accelerating system consists of 16 radiofrequency cavities with a maximum electric field of 5.5 MV/m. The two beams are structured in 2808 bunch separated

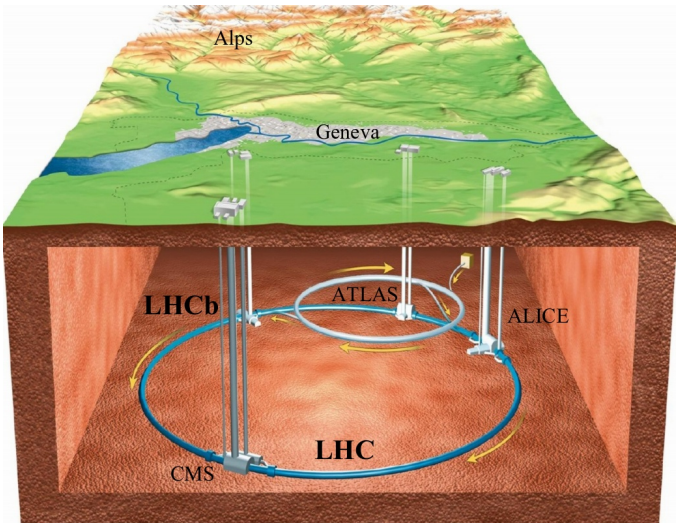

Figure 2.1: The LHC apparatus located in the underground of Geneva.

by 50 ns (in the next data acquisition the frequency and the bunch number will be increased to the LHC designed value of 25 ns and 3564).

The beams are directed around the accelerator from a complex magnetic system. The 1232 electromagnets are built from coils of special electric cables that, operating in superconducting state (temperature of $1.9K$) could endure a circulating current of 11.85 kA that generates a magnetic field of 8.4 T. The focusing system consists of 392 superconducting magnets quadrupoles producing a 6.8 T field. The complex of accelerators that precede LHC and accelerate particles until 300 GeV is composed by a chain of a linear accelerator (Linac2) and three synchrotrons, Proton Synchrotron Booster (PSB), Proton Synchrotron (PS) and Super Proton Synchrotron (SPS), figure 2.2.

The four experiments situated in beam interaction points have different detector structure that correspond to different physical purpose:

- **ATLAS (A Toroidal LHC Apparatus)** is a multipurpose experiment which works mainly at high luminosity to discover signature of new physics and performs precise measurement of Standard Model. Together with CMS recently observed the new boson candidate to be the Higgs particle

- **CMS (Compact Muon Spectrometer)** is the second multipurpose experiment that pursues the same physics goals as ATLAS using different and complementary technologies. It works mainly at a high luminosity.

- **LHCb,** has a completely different structure with respect to the other experiments and performs accurate measurements on flavour physics of B meson

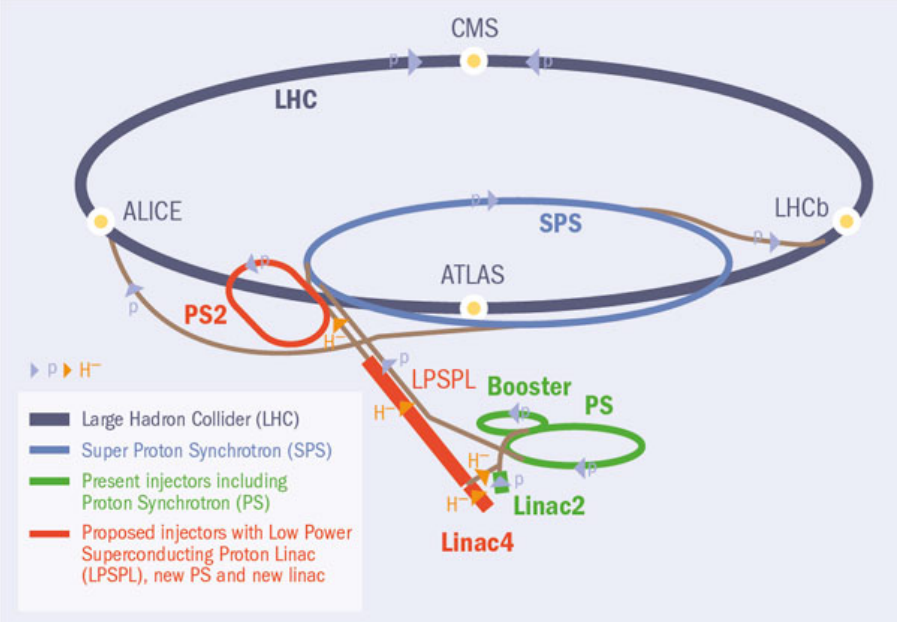

Figure 2.2: The complex of apparatus that proceeds LHC in the acceleration chain.

to explain the problem of CP violation.

- **ALICE (A Large Ion Collider Experiment)** is dedicated to the study of quark-gluon plasma formed in heavy nucleons collisions, that can be accelerated in LHC instead of the proton in some dedicated runs.

2.2 The ATLAS detector

ATLAS is a general purpose detector [35], designed to exploit the full discovery potential of LHC and with a total length of 42 m, a radius of 11 m and a weight of 7000 tons is the largest LHC experiment (figure 2.3). The detector has an overall cylindrical symmetry and it is composed by the inner detector (ID), the calorimetric system the muon spectrometer (MS) and the forwards detectors.

The combination of all different instrumentation composing ATLAS detector allow to reconstruct also the most complex final states, as the one of the semileptonic decay of the $t\bar{t}$ that contains several jets, a neutrino and a lepton. Different objects are reconstructed merging measurements coming from different part of the detector: the particle trajectory and the momentum of the charged particles are evaluated by the inner detector (ID) through the bending given by the magnet, while the energy and the particle identification have been provided by the calorimeters and the muon spectrometer. Forward detectors accomplish the luminosity measurement task.

In the ATLAS standard coordinate system the beam direction corresponds to the z axis while the x-y plane is on the transverse plane to the beam pipe, in particular the x axis points the center of the ring and the y points towards the surface. Polar coordinate are also largely used: the azimuth angle (ϕ) is measured around the beam axis and the polar angle (θ) is the angle that a particle forms with the beam axes. The θ coordinate is usually substituted with the pseudorapidity:

$$\eta = -\ln(\tan\frac{\theta}{2}) \tag{2.1}$$

that, in limit of a mass-less particle, is equal to the rapidity:

$$y = \frac{E + p_z}{2(E - p_z)} \tag{2.2}$$

This coordinate was introduced since the cross section is symmetric in η.

In the following sections will be presented the main features of the ATLAS detector. This is important to understand how is possible to reconstruct the complex final state used in the analysis, starting from the signal detected in ATLAS.

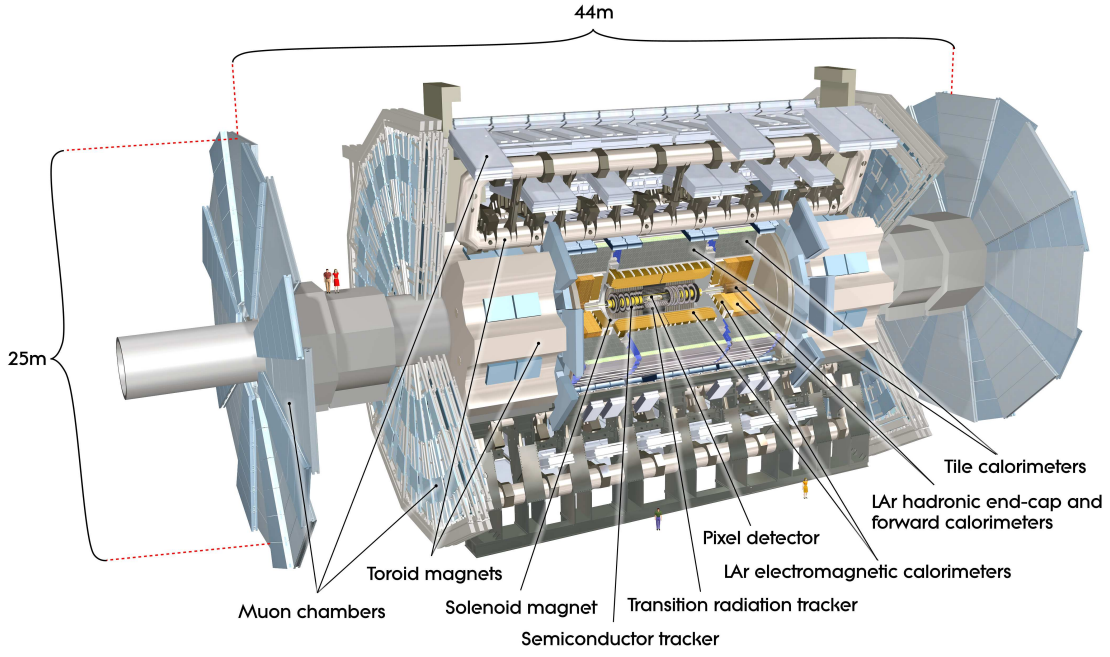

Figure 2.3: The overall view of ATLAS detector.

2.2.1 The inner detector

Approximately 1000 particles will emerge from the collision point every 25 ns within $|\eta| < 2.5$, creating a very large track density in the detector. To achieve the momentum and vertex resolution requirements imposed by the benchmark physics processes, high-precision measurements must be made with fine detector granularity. The Inner Detector is embedded within a 2 T thin superconducting solenoid magnet of 2.5 m diameter and it is composed by a silicon pixel, a micro-strip detector and (SCT) a straw-tube tracking detector (TRT), figure 2.4. The precision tracking detectors (pixels and SCT) are arranged on concentric cylinders around the beam axis while in the end-cap regions they are located on disks perpendicular to the beam axis. The highest granularity is achieved around the vertex region using silicon pixel detectors. ID provides tracking and vertexing capabilities within $|\eta| < 2.5$, as well as electron identification within $|\eta| < 2.0$. The Inner Detector globally provides a transverse impact parameter resolution of $\simeq 35$ ($\simeq 10$) μm for pions with $p_T = 5$ (100) GeV and a transverse momentum resolution of about 4% for 100 GeV muons.

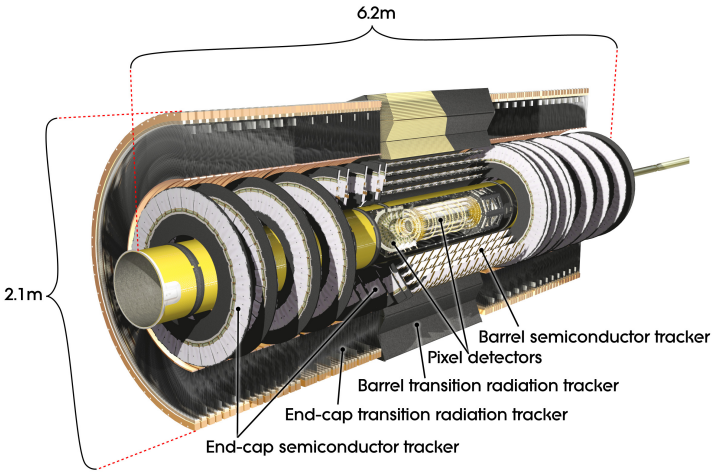

Figure 2.4: Cut-away view of the ATLAS Inner Detector.

Pixel Detector

The pixel detector [36], figure 2.5, is the nearest to the collision point and measures the particle impact parameters and the decay vertexes of short living particles. The pixel detector consists of 1744 pixel modules organized in three barrel layers, containing approximately 67 million of pixel, complemented by three end-cap disks

on each side, containing 13 million of pixel. The system covers a total active area of about 1.7 m^2. To counterbalance the effect of the Lorentz deviation the pixel modules in the barrel region are tilted 20° with respect to the cylinder's tangent.

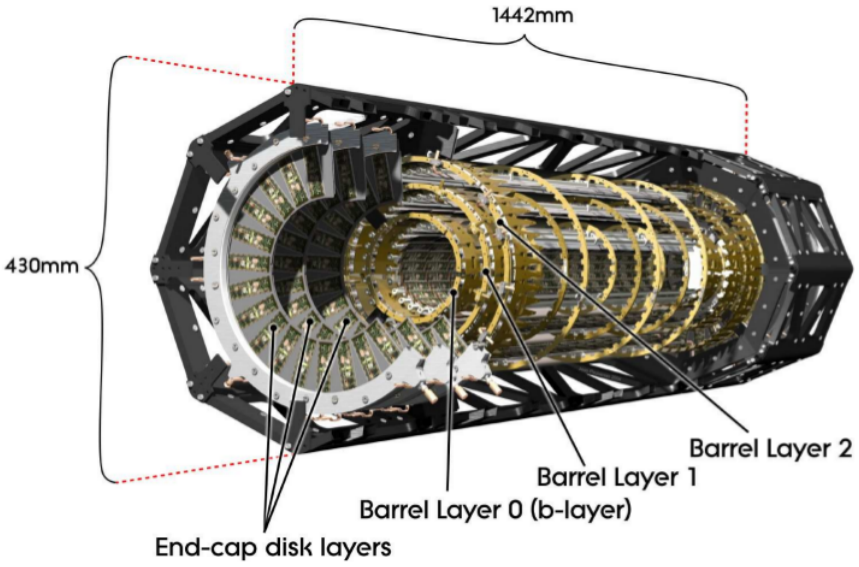

Figure 2.5: A schematic view of the active region of the pixel detector consisting of barrel and endcap layers

The pixel layers are segmented in R-ϕ and z with typically three pixel layers crossed by each track. All pixel sensors are identical and have a minimum pixel size in R-ϕ x z of 50 x 400 $(\mu m)^2$. The intrinsic precision in the coordinate measurements in the barrel and the end cap are respectively 10 μm (R-ϕ) 115 μm (z) and 10 μm (R-ϕ) 115 μm (R).

Each readout channel of the pixel detector is radiation hard to withstand over 300 kGy of ionizing radiation and more than $5 \cdot 10^{14}$ neutrons per cm^2 over ten years of operation of the experiment.

The micro-strip detector

The Semi Conductor Tracker (SCT) system is designed to provide track precision measurements and contribute to the measurement of of momentum, impact parameter and vertex position in the intermediate radial range. The SCT barrel consists of four concentric layers of radius ranging from 30 to 52 cm from the beam axis. On both sides of the barrel layers, there are 9 SCT disks that cover up to $|\eta| < 2.5$. Each SCT module is made of two strip layers, each of which consists of two 6.4

cm long sensors with a strip pitch of 80 μm rotated of 40 mrad with respect one each other. Each track passing in the SCT cross eight strip layers, corresponding to four space points. The intrinsic accuracies per module in the barrel are 17 μm (R - ϕ) and 580 μm (z) and in the disks are 17 μm (R - ϕ) and 580 μm (R). The total number of readout channels in the SCT is approximately 6.3 million.

The straw-tube tracking detector

The Transition Radiation Tracker (TRT) combines drift tube chamber tracking capabilities with transition radiation detector power of electron/pion discrimination. A single TRT component is composed by Polyimide drift (straw) tubes of 4 mm diameter that contains the anodes: 31 μm diameter tungsten wires plated gold, directly connected to the front-end electronics and kept at ground potential. The cathodes are operated typically at –1530V to give a gain of $2.5 \cdot 10^4$. The gap between the straw and the wire is filled by a mixture of gases (70% Xe, 27% CO_2 and 3% O_2). The passage of ionizing particle induce a low energy signal on the anodes. At the same time, some particles crossing polypropylene fibers cause transition radiation emission (in the X-ray spectrum) which is absorbed by the Xe present in the gas mixture. This last process leads to an high energy signal in the TRT electronic that can be distinguished from ionization signal by the voltage intensity. A large number of hits (typically 36 per track) is provided by the straw tubes of the TRT, which enables track-following up to $|\eta| = 2.0$. The TRT only provides R - *phi* information, for which it has an intrinsic accuracy of 130 μm per straw. In the barrel region, the straws are parallel to the beam axis and are 144 cm long, with their wires divided into two halves, approximately at $\eta = 0$. In the end-cap region, the 37 cm long straws are arranged radially in wheels. The total number of TRT readout channels is approximately 351,000.

2.2.2 Magnetic system

ATLAS features a system of four large superconducting magnets that provide, over a volume of approximately 12,000 m^3, a magnetic field that is required for the identification of charged particles and momentum measurements. The system, illustrated in figure 2.6 consists of a solenoid and three toroids (one barrel and two end-cap).

- The **solenoid** is aligned with the beam axis and provides a 2 T axial magnetic field for the inner detector as can be seen in figure 2.7. The layout was

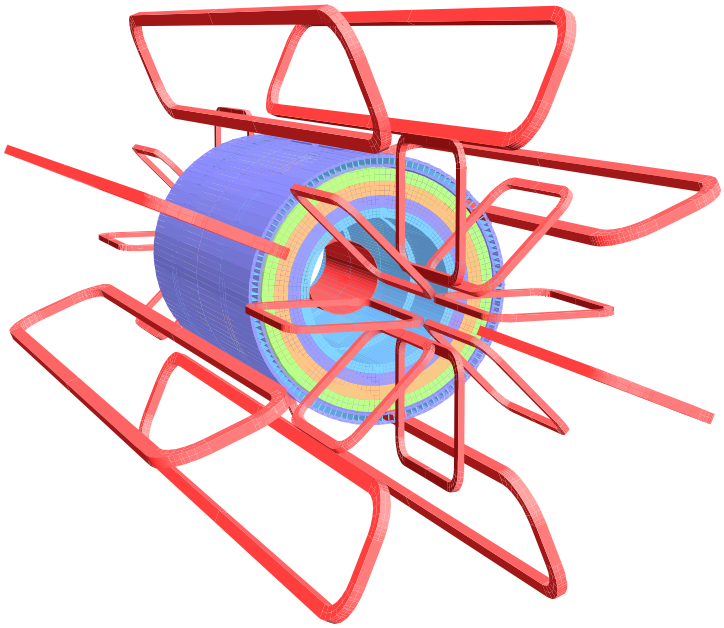

Figure 2.6: Geometry of magnet windings and tile calorimeter steel. The eight barrel toroid coils, with the end-cap coils interleaved are visible. The solenoid winding lies inside the calorimeter volume.

carefully optimized to keep the material thickness in front of the calorimeter as low as possible, resulting in the solenoid assembly contributing a total of $\simeq 0.66$ radiation lengths at normal incidence.

- The large super-conducting air-core **toroid system** is constituted by eight Barrel Toroids (BT) 25 m long, with an inner core of 9.4 m and an outer diameter of 20.1 m, and two End-Cap Toroids (ECT) 5 m long (inner core 1.64 m, outer diameter 10.7 m). The toroid structure is open to minimize the uncertainty on the momentum measurements due to multiple scattering. The BT provide the particle bending in the region $\eta < 1$ while in $1.4 < \eta < 2.7$ charged tracks are bent by the ECTs. Over $1 < \eta < 1.4$, usually referred to as the transition region, magnetic deflection is provided by a combination of barrel and end-cap. This magnets configuration provides a $\simeq 4$ T field mostly orthogonal to the muon trajectories.

The magnetic field components in the inner detector cavity is illustrated in figure 2.7, the field is almost entirely in the beam direction with a constant intensity of 2 T.

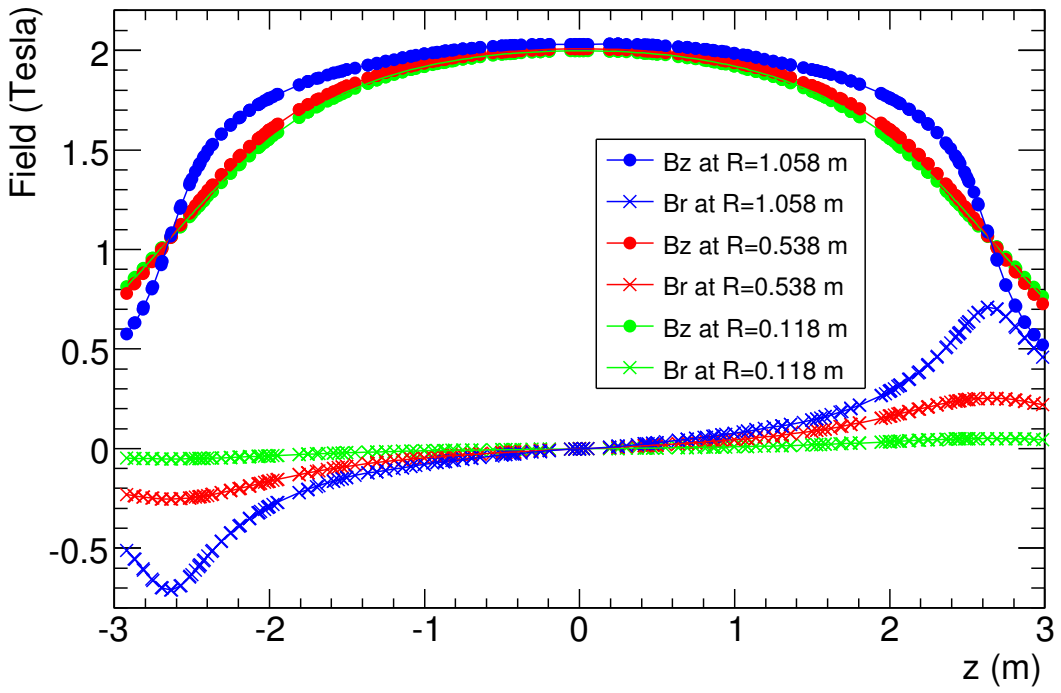

Figure 2.7: R and z dependence of the radial (Br) and axial (Bz) magnetic field components in the inner detector cavity, at fixed azimuth. The field is almost constant at 2 T and direct along the z axis.

2.2.3 Calorimetric System

The calorimetric system [37] plays a central role in ATLAS.

In the difficult environment created by the high luminosity of the LHC machine, the calorimeter is designed to trigger and to provide precision measurements of electrons, photons, jets, neutrons and missing E_T. The system, composed by an Electromagnetic (EM) and an Hadronic calorimeter, cover the range $|\eta| < 4.9$ using different techniques, suited to the widely varying requirements of the physics processes of interest and of the radiation environment. The finer granularity is reached in the EM, over the region matched to the inner detector, and is ideally suited for precision measurements of electrons and photons. The missing E_T and jets measurements are performed in the Hadronic calorimeter with less granularity, even so sufficient to satisfy the physics requirements.

The foremost design feature of the calorimeters is their depth since they must provide good containment for particles showers and must also limit punch-through into the muon system. The total thickness of the EM calorimeter is higher than 22 radiation lengths in the barrel and 24 in the end-caps and the 9.7 interaction

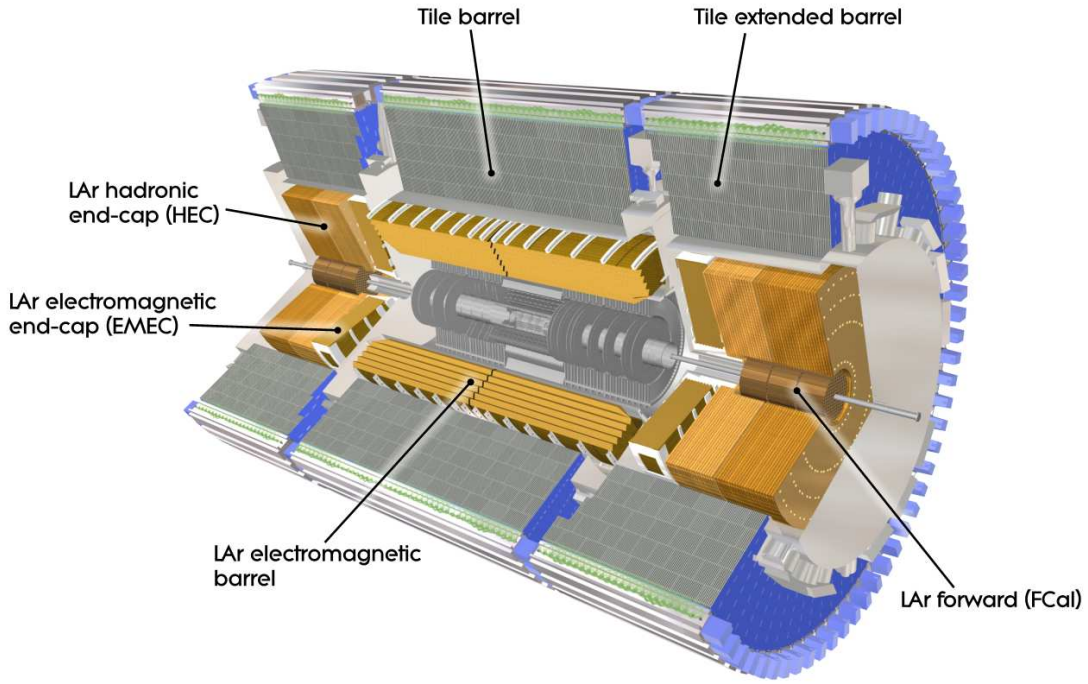

Figure 2.8: Cut-away view of the ATLAS calorimeter system. Both the EM (barrel and end-cap) and the hadronic calorimeter (Tile, HEC and forward) are shown

lengths of active calorimeter in the barrel (10 in the end-caps) are sufficient to provide good resolution also for higher energy jets. The overall structure of the calorimetric system is reported in figure 2.8

The energy resolution of a calorimeter may be calculated as:

$$\frac{\sigma(E)}{E} = \frac{a\%}{\sqrt{E}} \oplus \frac{b}{E} \oplus c\% \tag{2.3}$$

where E is in considered in GeV, a is the sampling term, b the noise term, and c the constant term reflecting local non-uniformities in the response of the calorimeter.

The resolutions of both EM and Hadronic Calorimeter are reported in table 2.1.

Electromagnetic Calorimeter

The EM calorimeter is a lead-LAr detector with accordion-shaped kapton electrodes and lead absorber plates over its full coverage, it is composed by two end-caps and by two identical barrel part separated by a small gap at z=0, located outside the

Detector Component	Energy Resolution
EM calorimetry	$\sigma(E)/E = 10\%/\sqrt{E} \oplus 1\ GeV/E \oplus 0.7\%$
Hadronic calorimetry	
barrel and end-cap	$\sigma(E)/E = 50\%/\sqrt{E} \oplus 1\ GeV/E \oplus 3\%$
forward	$\sigma(E)/E = 100\%/\sqrt{E} \oplus 1\ GeV/E \oplus 10\%$

Table 2.1: General performance of the calorimetric system of the ATLAS detector.

solenoid magnet. Each end-cap is formed by two coaxial wheels: an outer one covering the region $1.375 < |\eta| < 2.5$, and an inner wheel covering the region $2.5 < |\eta| < 3.2$.

The amount of material and the presence of a magnetic field imply the necessity to correct for the energy lost in front of the calorimeter, this is possible thanks to a presampler, consisting of an active LAr layer.

This calorimeter allow to achieve an energy resolution for photons of p_T of 100 GeV better than 1.5%.

Hadronic Calorimeter

The central part of the Hadronic calorimeter is the **tile calorimeter**, placed directly outside the EM calorimeter envelope. Overall it covers a range $|\eta| < 1.7$ and it is a sampling calorimeter using steel as the absorber and scintillating tiles as the active material. The tile calorimeter measures jet energies with a resolution $\Delta(E)/E = 65\%/\sqrt{E} \oplus 2.5\% \oplus 5/E\%$.

The **Hadronic End-cap Calorimeter** (HEC) is located directly behind the EM calorimeter end-caps and has the same structure, consisting of two independent wheels per end-cap. The HEC extends out to $|\eta| = 3.2$, thereby overlapping with the forward calorimeter.

The **LAr forward calorimeter**(FCal) is integrated in the end-cap nearby the beam axis with the front face recessed by about 1.2 m with respect to the EM calorimeter and it has two main purposes: to increase the uniformity of the coverage of the calorimetric system and to reduce radiation background in the muon spectrometer. FCal consists of three modules in each end-cap: the first in copper, is optimized for electromagnetic measurements and the other two made of tungsten useful predominantly for the measure of the energy of hadronic interactions. The coverage reached from FCal is $3.1 < |\eta| < 4.9$.

2.2.4 Muon Spectrometer

The outer part of the ATLAS detector is formed by the muon spectrometer. This is designed to detect charged particles exiting the calorimeters and to measure their momentum in the pseudorapidity range $|\eta| < 2.7$. The second purpose of the muon spectrometer is to trigger on these particles in the region $|\eta| < 2.4$. The layout of the muon spectrometer is shown in figure 2.9

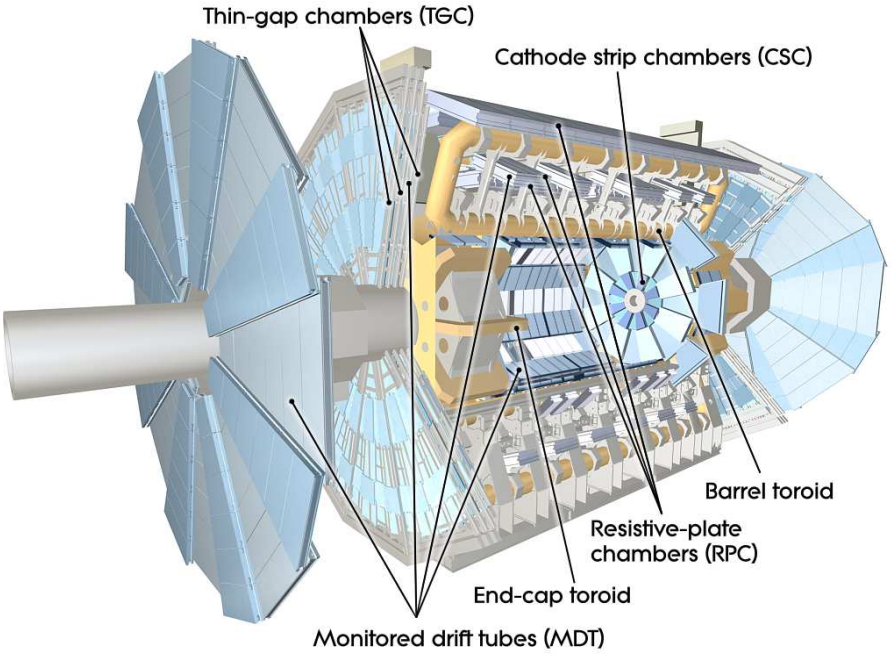

Figure 2.9: View of the overall muon spectrometer.

The momentum is measured thanks to the deflection provided by the barrel and end-cap magnet toroid. The magnetic system configuration provides a field which is mostly orthogonal to the muon trajectories, minimizing the multiple scattering, that would imply a degradation of resolution. In the barrel region, tracks are measured in chambers arranged in three cylindrical layers around the beam axis; in the transition and end-cap regions, the chambers are installed in planes perpendicular to the beam, also in three layers.

The Monitored Drift Tubes (MDT's) are installed on the most of the η range and provide a precision measurement of the track coordinates in the principal bending direction of the magnetic field. At large pseudorapidities are located multiwire proportional chambers with cathodes segmented into strips, called Cathode Strip Chambers (CSC's). These have an higher granularity to withstand the demanding rate and background conditions. The stringent requirements on the relative

alignment of the muon chamber layers are met by the combination of precision mechanical-assembly techniques and optical alignment systems both within and between muon chambers. The trigger system is made by the Resistive Plate Chambers (RPC's), used in the barrel region, and Thin Gap Chambers (TGC's), in the end-cap regions. The trigger chambers serve a threefold purpose: first of all provide bunch-crossing identification than provide well-defined pT thresholds, and lastly measure the muon coordinate in the direction orthogonal to that determined by the MDT. The performance for a stand-alone transverse momentum resolution varies between 3% and 12%, for p_T values between 10 GeV and 1000 GeV. The spectrometer can measure muon momenta with adequate momentum resolution and excellent charge identification in the range between $\sim$ 3 GeV and $\sim$ 3 TeV.

The coverages and functions of all muon spectrometer chamber are summarized in the following table:

Muon Spectrometer	Performance
Monitored drift tubes	
- Coverage	$\|\eta\| < 2.7$ (innermost layer:$\|\eta\| < 2$)
- Function	Precision tracking
Cathode strip chambers	
- Coverage	$2.0 < \|\eta\| < 2.7$
- Function	Precision tracking
Resistive plate chambers	
- Coverage	$\|\eta\| < 1.05$
- Function	Triggering, second coordinate
Thin gap chambers	
- Coverage	$1.05 < \|\eta\| < 2.7$
- Function	Triggering, second coordinate

2.3 Luminosity measurement

To obtain the cross section from the measured number of event, as shown in equation 1.8, the luminosity has a main role and error on this quantity represent one of the major systematic uncertainties for cross-section measurements. Consequently

the measurement of this quantity is fundamental for this thesis and it is, in general, a key component of the ATLAS physics program. In ATLAS there are four detectors LUCID (LUminosity Cherenkov Integrating Detector), BCM (Beam Conditions Monitor), FCAL and TILECAL that determinate luminosity [38]. Here will be given only a brief description of the LUCID, where the INF and the Universityof Bologna have a leading role in the projection, development and mainteinance

2.3.1 LUCID

LUCID is a Cherenkov detector formed by two identical part each one composed by 16 aluminum tubes filled with the C_4F_{10} gas and distant 17 m from the interaction point. It is located around the beam pipe and covers a pseudorapidity range $5.6 < |\eta| < 6.0$.

The Cherenkov photons emitted by a charged particle passing through the gas inside the detector are reflected by the tube walls and finally recoiled by photomultiplier (PMT) situated at the bottom of each tube. Additional Cherenkov photons are produced in the quartz window of the PMT, for a total of about 100 photoelectrons per incident charged particle. From the number of tubes giving a signal over threshold (equivalent on average to 15 photoelectrons) is possible to evaluate the average number of interactions per bunch crossing and then the luminosity. LUCID monitors bunch crossing in order to compare its results with the BCM detector and to indicate eventual problem of LHC.

2.4 Trigger System

The huge amount of data collected by LHC in each collision can't be completely acquired and analyzed. ATLAS through a complex Trigger and Data Acquisition (collectively TDAQ) systems manages the selections of events, needed to reduce the information rate from $\sim$ 40 MHz to approximately 200 Hz.

The trigger system has three distinct levels: L1, L2, and the event filter. Each trigger level refines the decisions made at the previous level and, where necessary, applies additional selection criteria.

The first level trigger is implemented using custom-made electronics and uses only a limited part of the information given by the detector. The decisions on each collision must be really fast to avoid the loss of information on the following bunch crossing: L1 decision must reach the front-end electronics within 2.5 μs after the

bunch-crossing with which it is associated. The L1 trigger searches for signatures that could signal an interesting event: high p_T muons, electrons, jets and also large missing transverse energy and large total transverse energy. For each event the L1 trigger also defines one or more Regions-of-Interest (RoI's) that are the η and ϕ coordinates of the detector regions where had been identified interesting features. Aside this geographical information the data on each RoI include also the type of feature identified and the criteria passed. The maximum L1 rate acceptable for the detector readout systems is 75 kHz. The diagram of the L1 operation is reported in figure 2.10.

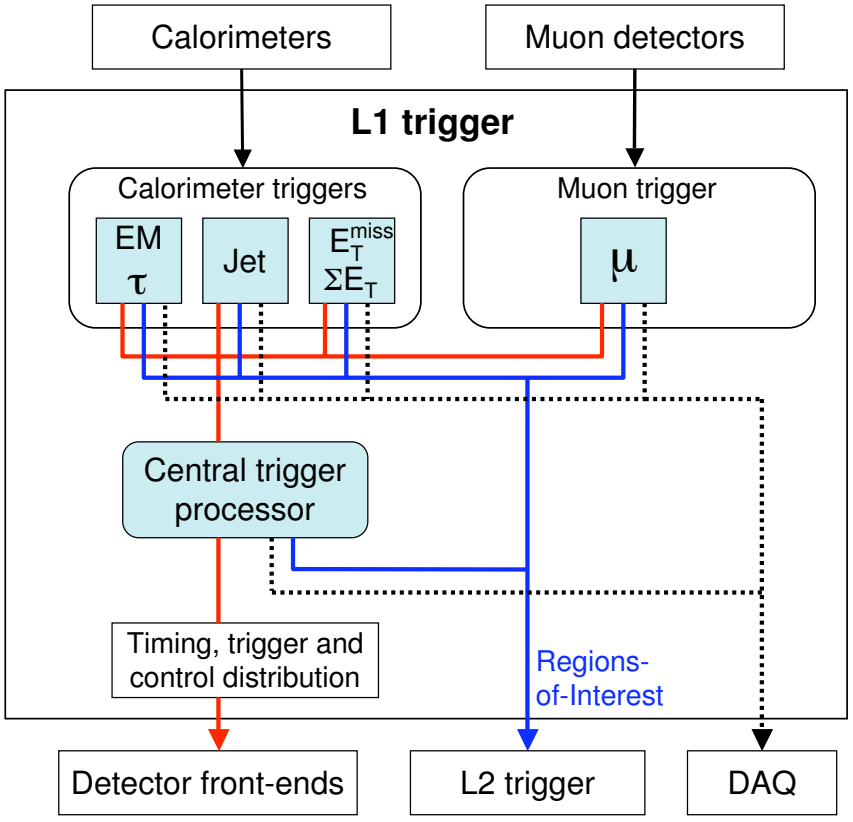

Figure 2.10: Block diagram of the L1 trigger. The overall L1 accept decision is made by the central trigger processor, taking input from calorimeter and muon trigger results. The paths to the detector front-ends, L2 trigger, and data acquisition system are shown from left to right in red, blue and black, respectively.

The L2 and the event filter are called together High Level trigger (HLT), they access to a larger amount of detector information and are almost entirely based on commercially available computers and networking hardware. The L2 trigger uses the RoI's data to reduce the event rate to below 3.5 kHz, with an average

event processing time of approximately 40 ms. The event selected by L2 then are transferred to the event-builder system that merges information from different sub-detectors and subsequently transfer the events to the event filter for the last selection.

The event filter uses offline analysis procedures on the events entirely built to reduce the rate to a size which can be recorded for subsequent analysis. The rate that passes the HLT selection is reduced to approximately 200 Hz and the events selected, with size of about 1.3 Mbyte, are moved to permanent storage.

The event selected from the TDAQ refers to all the ATLAS analysis so it's necessary to apply further selection cuts during the offline analysis to reject events selected by triggers not specific for particular analysis on which we are interested for. The event selection applied in this thesis will be presented in Chapter 3.

Chapter 3

Analysis Method and Event Selection

In this thesis the quark top is studied in a particularly high momentum regime that requires the application of innovative techniques, mainly regarding the jet reconstruction. These techniques will be illustrated in this chapter together with the main steps of the analysis chain. The definition of the physical objects and the main features of the high p_T regime will be presented in Paragraph 3.1. Then in Paragraphs 3.2 and 3.3 the characteristics of data and MC samples used for signal and background will be described. In Paragraph 3.4 and 3.5 respectively the event selection and the comparison between data and MC samples will be presented. In the following the main features of the unfolding procedure, consisting in the correction of the final distributions for the distortion effects introduced by the finite acceptance and resolution of the detector will be explained. In conclusion an overview on the analysis framework used to perform the entire analysis will be given in Paragraph 3.7.

3.1 Object definition

The LHC allows to reach center-of-mass energy that far exceeds the masses of all the known particles and consequently there is a high probability that also heavy particles as the bosons (W,Z,H) or the top quarks can be produced with a large transverse momentum, becoming *boosted particles*. In these conditions the decay products tend to be more collimated due to relativistic effects and the three jets coming from the hadronic decay of the top quark could result partially or totally

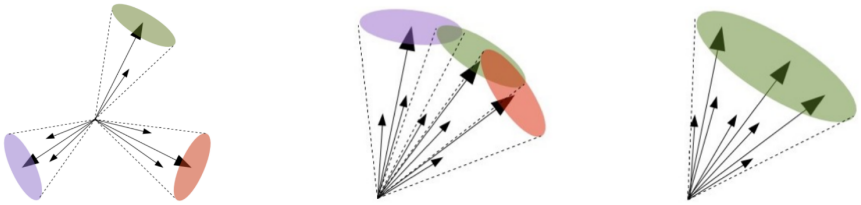

Figure 3.1: Graphical representation of hadronically top decay: the jet are well separated in case of low p_T (left) and partially overlapped in case of high pt (center). In this case is more convinient to reconstruct the system as one large-R jet (right).

overlapped, as shown in figure 3.1. In this case the three jets cannot be separated by the standard reconstruction algorithm and the original boosted top quark cannot be efficiently detected and reconstructed. In order to improve the top reconstruction efficiency new jet algorithms have been developed enlarging the opening angle variable R defined as R=$\sqrt{\Delta\phi^2 + \Delta\eta^2}$ where $\Delta\phi$ is the azimuthal angle difference between each couple of clusters belonging to the jet and $\Delta\eta$ is related to the polar angular difference θ ($\eta = -1/2 \log \tan(\theta/2)$). These concepts will be detailed in Paragraph 3.1.4. The standard jet reconstruction foresees to include in the jet all the clusters belonging to a radius of R=0.4. In case of top quarks with p_T higher than 300 GeV, the decaying products are often contained in a cone of R=1 (see figure 3.2) that allows to consider them as a single large-R jet. In figure 3.2 is represented the ΔR between the two products of the top decay, on the left, and of the two quarks deriving from the W boson decay (that derives from the hadroni top decay) with respect to the hadroic top p_T; in both the histograms when the top p_T is higher than 300 GeV the ΔR between the two objects represented is less than 1. There are algorithms and jet internal variables, illustrated in next section, that allow to discriminate if a large-R jet comes from a top quark decay.

The analysis described in this thesis aims to measure the production differential cross section of the $t\bar{t}$ pairs at high transverse momentum in the semileptonic (l+jet) channel. The searched event signature is composed by a lepton, a jet coming from a b quark, a large-R jet and large missing energy in the transverse plane indicating the presence of a neutrino. The event selection has been optimized to individuate the $t\bar{t}$ production signature and reduce the background contamination by similar final state events that could be misidentified as signal.

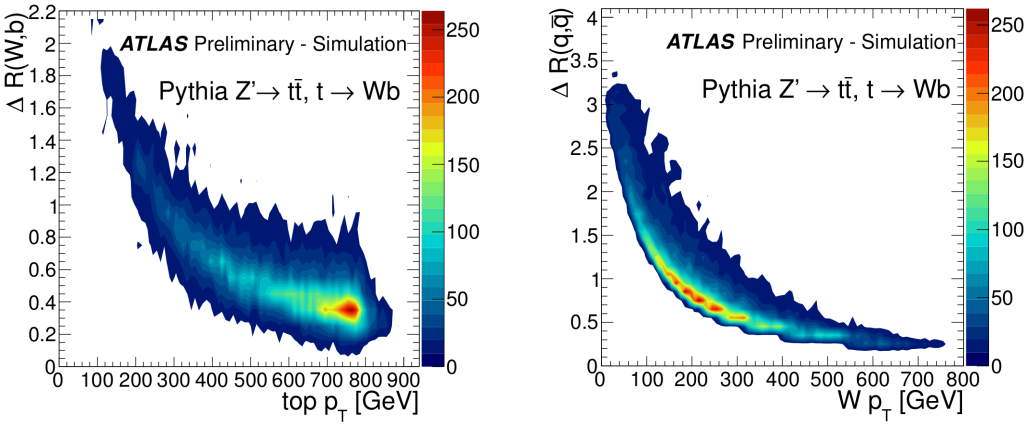

Figure 3.2: Sample simulated with Pythia for $Z' \to t\bar{t}$ ($m_{Z'}$ fixed to 1.6 TeV) events. In the two figure is reported with respect to the top p_T the ΔR between W and b (left) and the ΔR between the quark coming from W decay (right).

3.1.1 Electrons

The reconstruction of electrons in ATLAS is based on the matching between ID tracks, that give its trajectory, and EM calorimeter clusters (deposit of energy in a part of the calorimeter), that define its energy. In order to be classified as "good electron" and to reduce the probability of misreconstruction, the electrons must satisfy several requirements on track quality, geometric position inside the detector, isolation and matching between track and calorimeter information. The list of all needed requirements is:

- No errors in the Liquid Argon electromagnetic calorimeter during data taking;
- The objects should be identified as "tight++" electrons by a specific algorithm based on the shape of the shower deposit in the calorimeter, the matching of the clusters to the associated tracks and the number of hits in the pixel and micro-strip detectors. The algorithm working point efficiencies has been chosen to have an efficiency of 78% evaluated on a Z boson sample;
- Distance between the track impact parameter and the primary vertex projection on the z-axis ($|Z_0^{PV}|$) less than 2 mm;
- The transverse energy $E_T = \frac{E_{cluster}}{cosh(\eta_{track})} > 25$ GeV;
- Compatibility with the calorimeter acceptance requirements $0 < |\eta_{cluster}| < 1.37$ and $1.52 < |\eta_{cluster}| < 2.47$;

- Isolation cuts on "Etcone20" and "Ptcone30" variables. The "Etcone20" is the total E_T deposited in the calorimeter towers in a cone of radius R = 0.2 around the electron position. "Ptcone30" is an analogue variable built by summing the p_T of the tracks in the ID around a cone of $\Delta R = 0.3$.

3.1.2 Muons

Muons are reconstructed combining information coming from the Muon Spectrometer (MS) with ID tracks and calorimeter cells. The reconstruction reaches a great accuracy with an identification efficiency greater than 95% and a relative momentum resolution running from 3% to 10% depending on the muon p_T.

In ATLAS, depending on detector information used for their identification, four muon candidates (illustrated in figure 3.3) have been defined: combined, standalone, segment-tagged and calo-tagged.

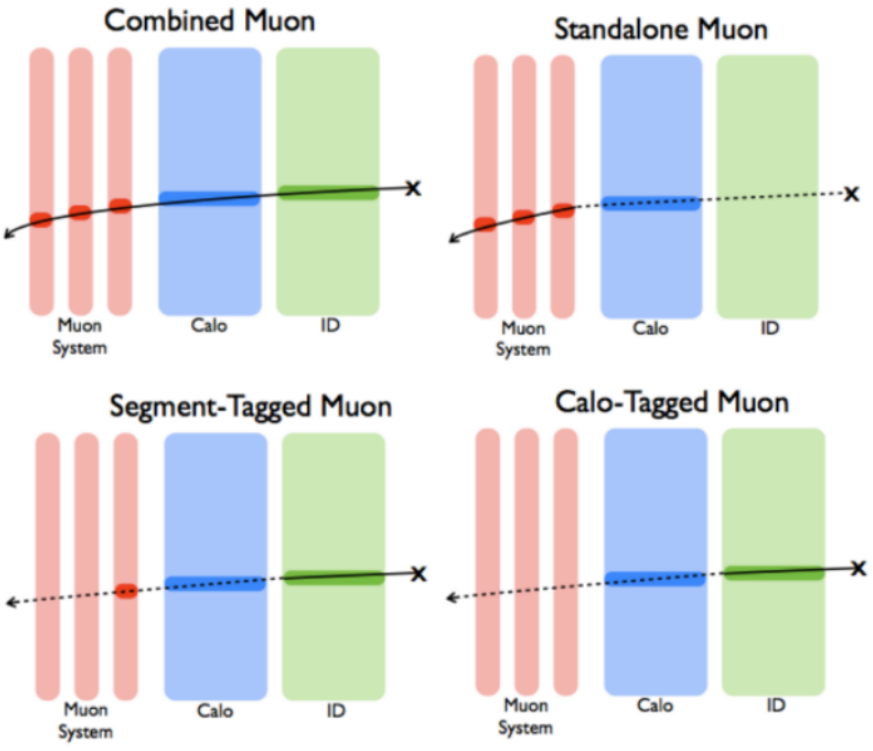

Figure 3.3: The four kinds of reconstructed muon candidates in ATLAS: Combined muon require information from all the three detectors, Standalone muons need MS identificaton.

The algorithm used to reconstruct the muons, called MuId [49], is based on a refitting procedure starting from the ID tracks to the MS ones. The "good muons"

must satisfy the following requirements:

- To be a combined muon or a standalone muon with at least three MDT+CSC hits;
- Distance between the track impact parameter and the primary vertex projection on the z-axis ($|Z_0^{PV}|$) less than 2 mm.
- Isolation requirement: Etcone20 $<$ 4 GeV and Ptcone30 $<$ 2.5 GeV

3.1.3 Missing transverse energy ($\not{E}_T$)

Neutrinos interact with matter only via the weak force and in ATLAS there are no detector capable to identify its transit. The quantity that allows to identify the presence of a neutrino is the missing transverse energy ($\not{E}_T$), that is calculated using the transverse momentum of all the physical products of the interesting event in the transverse plane. The detector configuration and the momentum conservation impose that the total transverse momentum should be zero, so if the sum of all detectable tracks p_T differs from zero then the missing p_T is attributed to the neutrino. The $\not{E}_T$ is calculated as:

$$\not{E}_T = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2} \tag{3.1}$$

where the ATLAS algorithms for the $E_{x,y}^{miss}$ reconstruction take into account several contributions:

$$E_{xy,tot}^{miss} = E_{xy,calo}^{miss} + E_{xy,cryo}^{miss} + E_{xy,\mu}^{miss} \tag{3.2}$$

The cryostat term $E_{xy,cryo}$ gets rid of the non-negligible loss of energy in hadronic showers due to the cryostat system installed between the LAr electromagnetic calorimeter and the Tile hadronic calorimeter. The $E_{xy,calo}^{miss}$ is evaluated only from cells belonging to topological clusters and included in the pseudo-rapidity range $\eta<$ 4.9.

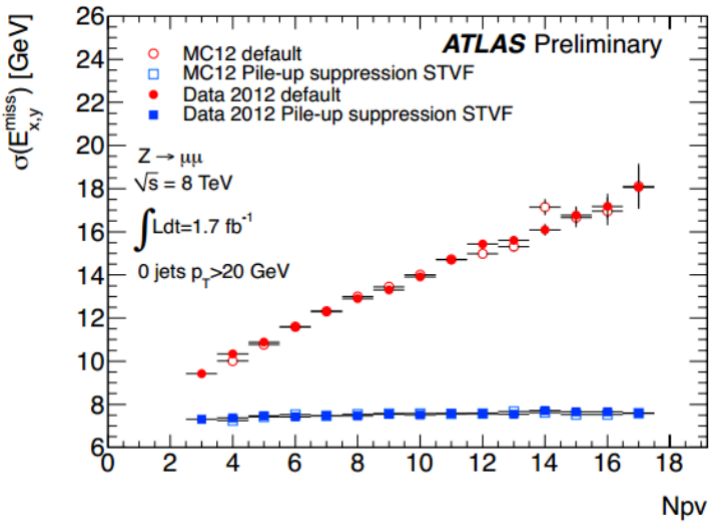

Figure 3.4: $\not{E}_T$ resolution with respect to the number of primary vertexes.

The muon term $E^{miss}_{xy,\mu}$ is evaluated from the ID and MS muon information

The $\not{E}_T$ performance and systematic uncertainties are estimated comparing the differences between data and simulations distributions in $Z \rightarrow ll$ and $W \rightarrow ll$ events.

The $\not{E}_T$ resolution is reported in figure 3.4.

3.1.4 Jets

The jets [50] [51] are collimated spray of energetic hadrons derived from the fragmentation and hadronization of gluons and quarks; by measuring their properties (energy, p_T, direction) one can get information on the parton that originated them. Jets can derive both from the hadronic decay of a heavy particles and from the radiative emission of a gluon by some other parton in the event; the first are usually harder while the second are generally considered a source of background.

In ATLAS the jet reconstruction process stars with calibrated three-dimensional clusters formed by calorimeter' s cells. The reconstruction starts with the identification of seed cells that should have energy significance at least 4σ above the noise level, defined as the square sum of the electric noise and pile-up. The neighbor cells with energy significance higher than 2σ are iteratively added to the seed individuated. In the end an additional ring of direct neighbor cells is added to the cluster. This topological clusters (topo-clusters) reconstruction is followed by a splitting algorithm that divides clusters in energy-categories using a local energy maximum criterion. Each cluster is calibrated using local properties such as energy density,

calorimeter depth and isolation with respect to nearby clusters. The jet is to be defined following a set of rules to group clusters into jets. These rules are provided by different jet algorithms and the one used for the analysis presented in this thesis is a sequential recombination algorithm, called anti-k_T.

This algorithm is based on the definition of the distance d_{ij} between two clusters (i,j) or intermediate objects of the reconstruction (pseudo-jet):

$$d_{ij} = \min(p_{Ti}^{2k}, p_{Tj}^{2k}) \frac{\Delta R_{ij}^2}{R^2} \tag{3.3}$$

$$\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2 \tag{3.4}$$

$$d_{iB} = p_{Ti}^{2k} \tag{3.5}$$

where d_{iB} represents the distance between the cluster and the beam and the parameter k of the anti-k_T method is fixed at -1. The algorithm is based on five steps:

- 1. Calculate d_{ij} of all the i, j combinations and the d_{iB}
- 2. Find the minimum between d_{ij} and d_{iB}
- 3. If the minimum is d_{ij} then recombine i and j into a single new pseudo-jet and return to step 1
- 4. Otherwise declare i to be a "final" jet and remove it from the list of particles. Return to step 1
- 5. Stop when no particles remains

The Cambridge Aachen (C/A) and the k_T algorithms are defined in a similar way with the parameter k=0,1 respectively. The anti-k_T algorithm is an infrared and collinear safe algorithm (IRC) for its distance definition. This means that if the event is modified by a collinear splitting or the addition of a soft emission the set of hard jets founded in the event remains the same. The anti-k_T algorithm favors clustering beginning from hard particles rather than clusterings that involves soft ones (k_T) or energy independent clustering (C/A). The resulting jet are almost circular shaped: a graphical comparison with the outcomes of other algorithms is reported in figure 3.5.

The jet algorithms take as parameter the jet radius R usually fixed in ATLAS to 0.4 for small-R jet and to 1 for large-R jet. A graphical example of the topoclusters

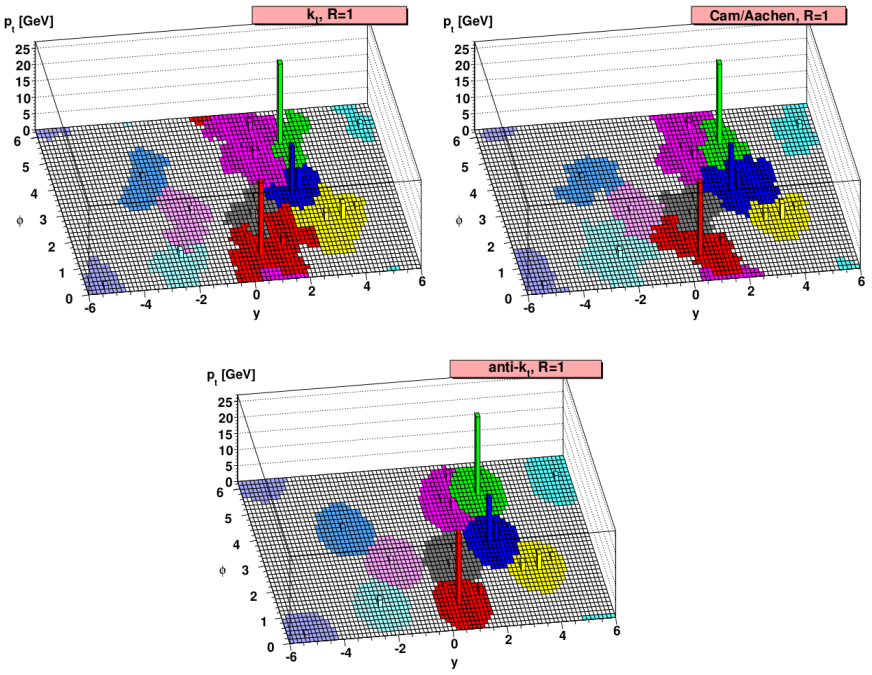

Figure 3.5: A sample parton-level event (generated with Herwig), clustered with four different jet algorithms, illustrating the "active" catchment areas of the resulting hard jets. The C/A and the k_T are more sensible to the softer clusters.

and the corresponding jets is reported in figure 3.6, here the calorimetric topoclusters are reported in the $\phi - \eta$ plane as green point while the final anti-k_T large-R jets are represented as ellipses whose border line color depend on the jet p_T. In the figure two random large-R jets configurations are shown, one of $t\bar{t}$ MC event (right) and one of a data random event (left).

A fundamental characteristic for a top quark analysis is to discriminate the top quark decaying jets originated by a b quark or by a lighter quark. This discrimination is performed by specific "b-tagging" algorithms based on the impact parameter and on the flight length (few millimeters) of the b quark; it can be measured due to the b quark mesons relatively long life-time ($\sim 10^{-12}$ s). The high luminosity of LHC leads to have events with many objects coming from secondary vertexes and underlying events (pile-up). These pile-up signals may be reconstructed as proper jets (fake jet) that can be rejected using the JVF quantity, defined as the ratio between the number of tracks matched with primary vertex and all the tracks from the hard scattering.

Large-R jets are more suitable in the boosted configuration where small-R jets starts to merge each other but, due to their greater dimensions, they largely suffer the presence of pile-up. Many algorithms (grooming algorithms) have been

developed in order to remove the pile up contribution.

Jet substructure variables and grooming algorithm

In boosted events where decay products are fully contained within individual large-R jets, a diminished mass resolution due to the high luminosity environment weakens the sensitivity to new physics and precise measurements so it is furthermore important to remove pile up tracks. This can be achieved through algorithms that selectively remove soft radiation during the iterative clusterization process. The jet grooming, in addition, leaves the hard substructure of a jet unchanged consequently allowing identification of jets coming from heavy particles. Large-R jets deriving from top quarks have indeed a peculiar substructure, given from the three high energy partons, that is absent in the light-quark and gluon jets.

The large-R jets used in this analysis have been obtained after the application of a particular grooming algorithm called trimming [52]. The trimming procedure uses the k_T algorithm to create subjets of radius R_{sub} (usually set to 0.2) the jet constituents. All the subjets with $\frac{p_{Ti}}{P_T^{jet}} < f_{cut}$ are removed, where p_{Ti} is the transverse momentum of the subjet i and f_{cut} is a parameter setted typically to few percent ($\sim$ 0.5 %). The remaining constituents form the trimmed jet, this procedure is illustrated in figure 3.7. Trimmed jets from light quarks or gluons typically lose 30-50 % of their mass, while jets containing the decay products of a heavy object usually loose only a few percentage of their mass and the removed part is mainly composed by pile-up. The trimming highlights the not-uniform internal structure of a large-R jet and the fraction of jet removed increases with the pile-up.

The trimming can also increase the discrimination power of some peculiar jet variables, called substructure variables:

- **Jet mass**: The jet mass is calculated from the momentum and the energy of all the sub-constituent i of a jet:

$$(m^{jet})^2 = (\sum_i E_i)^2 - (\sum_i p_i)^2 \tag{3.6}$$

  where E_i and p_i are the energy and three-momentum of the i^{th} jet constituent (each energy deposit is supposed to be given by massless particles). This variable is essential in the search of boosted high-mass particle and constitutes a powerful discriminant between signal and background.

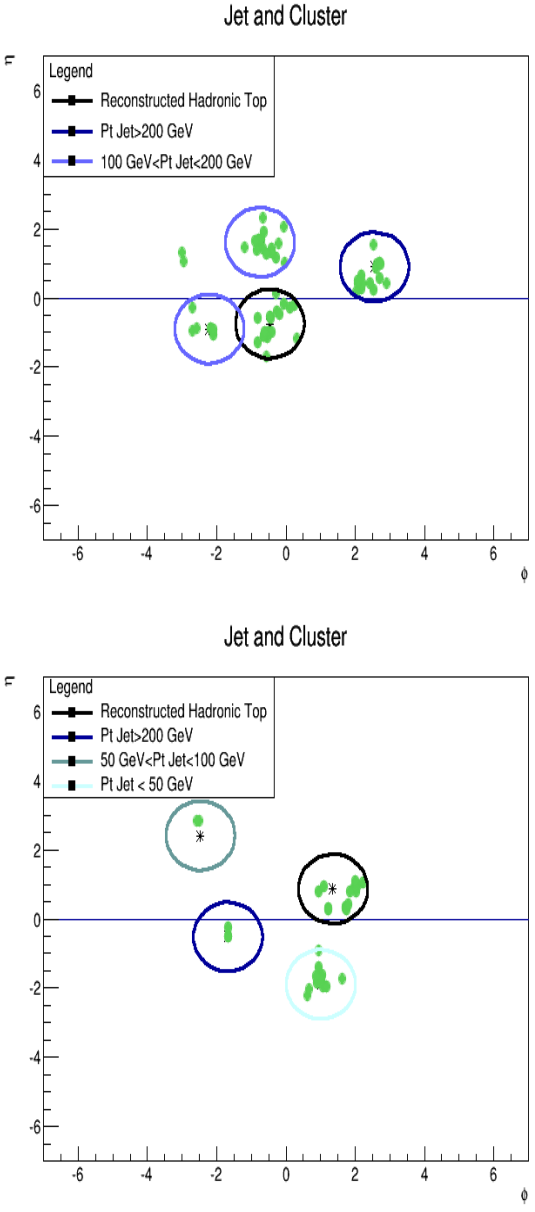

Figure 3.6: Representation of cluster and respective anti k_T jets. The jet's border line color depend on the jet p_T, in black there is the jet with highest p_T, reconstructed as top. The two plot reports two events from the analysis, on the left from the Powheg+Pythia simulated signal and on the right from observed data.

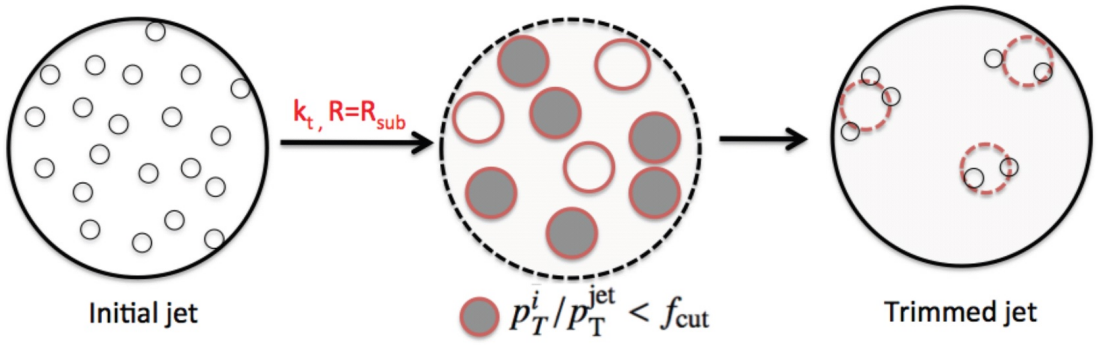

Figure 3.7: Graphical representation of the trimming algorithm

- **Splitting scale**: This method uses the k_T algorithm to reconstruct the jet. The observable considered as discriminant to individuate hard substructure is $\sqrt{d_{12}}$ (see below) defined as the splitting scale between the two proto-jet identified in the last step of the jet reconstruction of the k_T algorithm. The splitting scale variable is defined as:

$$\sqrt{d_{ij}} = min(p_{Ti}, p_{Tj}) \text{ x } \Delta R_{ij} \tag{3.7}$$

where ΔR_{ij} is the distance between two jet constituents. The distribution of W+jets background and MC signal with respect to$\sqrt{d_{12}}$ is shown in figure 3.8 (left) together with the signal efficiency versus the background rejection as a function of the cut on $\sqrt{d_{12}}$ (ROC Curve), figure 3.8 (right). Here is also indicated as green line the cut on $\sqrt{d_{12}}$ used in the event selection that allow to reach a 50% of signal efficiency and an 80% of background rejection.

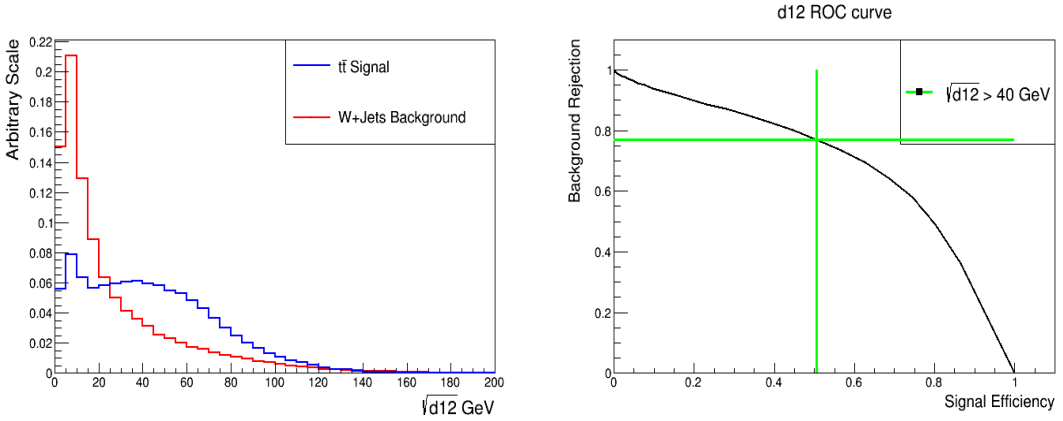

Figure 3.8: Normalized distributions of large-R jet of MC signal and W+jets background (left), the events reported passed analysis selection on jet p_T and η. The ROC curve of the $\sqrt{d12}$ cut is shown on the right. The green lines indicate the working point used in the analysis (after the application of the trimming procedure).

- **N-subjettiness**: This variable is correlated to the subjet multiplicity. The k_T algorithm is used to clusterize all the jet constituents with $d_{iB} < d_{ij}$, see definition 3.3, until there are exactly N proto-jet remaining. τ_N is defined as follows:

$$\tau_N = \frac{1}{d_0} \sum_k^N p_{Tk} \text{ x } \min(\delta R_{1k}, \delta R_{2k}, ..., \delta R_{Nk}) \text{ with } d_0 \equiv \sum_k p_{Tk} \text{ x } R$$

  where R is the jet radius, p_{Tk} is the p_T of proto-jet constituent k and δR_{ik} is the distance between two proto jets i and k.

The effect of the trimming on anti k_T large-R jet with p_T belonging to the range $600 < p_T^{large-Rjet} < 800$ GeV with respect to jet mass and splitting scale is shown in figure 3.9 where the two distributions are both normalized to 1. These studies had been performed on simulated sample of $Z' \to t\bar{t}$ ($m_{Z'}$= 1.6 TeV), used as signal (red lines) and multijet background (black lines). The trimming increases the separation between signal and Monte Carlo distributions for all the substructure variables considered, helping the discrimination based on these quantities.

3.2 Data and Monte Carlo Samples

The events considered in this analysis have been collected in proton-proton collisions acquired during 2012 by the ATLAS detector at the LHC, with a center

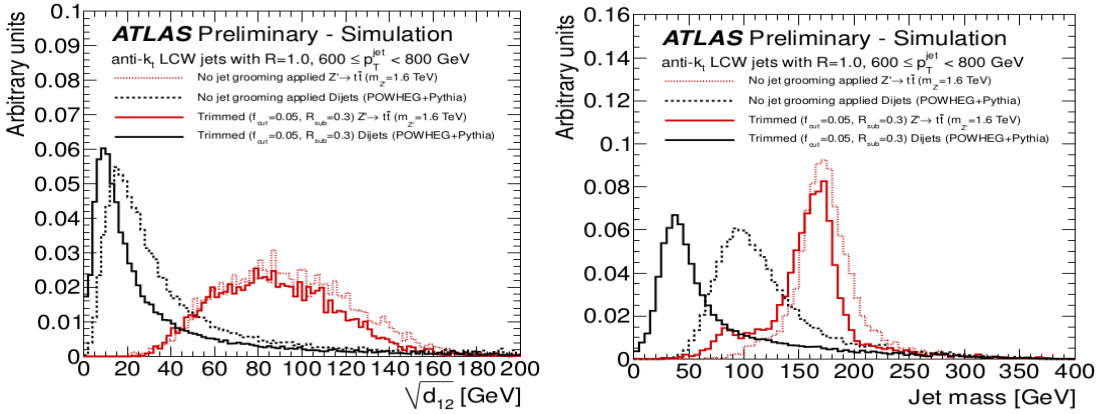

Figure 3.9: Leading-p_T anti-k_t large-R jet d_{12} (left) and mass (right) comparing $Z' \to t\bar{t}$ signal to POWHEG multi-jet background. The dotted lines show the ungroomed leading-p_T distribution, while the solid lines show the corresponding trimmed($f_{cut} = 0.05$, $R_{sub} = 0.3$) jets. The groomed distributions are normalized with respect to the ungroomed distributions, which are normalized to unity

of mass energy of 8 TeV and an integrated luminosity of 20.3 fb^{-1}. The average number of interactions per bunch crossing (pile-up) was about 21. The data are only used if acquired in stable beam conditions and with all sub-detectors fully operational. A trigger preselection has been imposed: for the electron a p_T threshold of 24 GeV for isolated electrons and 60 GeV for not isolated ones has been required while for the μ-channel the p_T thresholds are 24 GeV for the isolated muons and 36 GeV for the not isolated ones.

A Monte Carlo (MC) $t\bar{t}$ simulation has been used in order to evaluate efficiencies, acceptances, some systematic uncertainties and distortions introduced by the detector. The MC generators are based on theoretical calculations and on experimental results including all present knowledge on particle physics. The events generation proceeds through several steps starting from the calculation of the hard scattering matrix element to the stable and semi-stable particle creation. The generation chain is summarized in figure 3.10. The first step consists in the calculation, at a fixed perturbative order of the matrix element of the partonic interaction, followed by the QCD cascade generation that consisting in a soft gluon emission and of quark pair production (parton shower). Once the final partons are generated, phenomenological hadronization models are used in order to produce the stable particles. Most of generators do not simulate all the chain and it is necessary to interface two or more generators to produce the full $t\bar{t}$ event. The events are then passed to the detector simulation program, GEANT4 [39], that reproduce the full

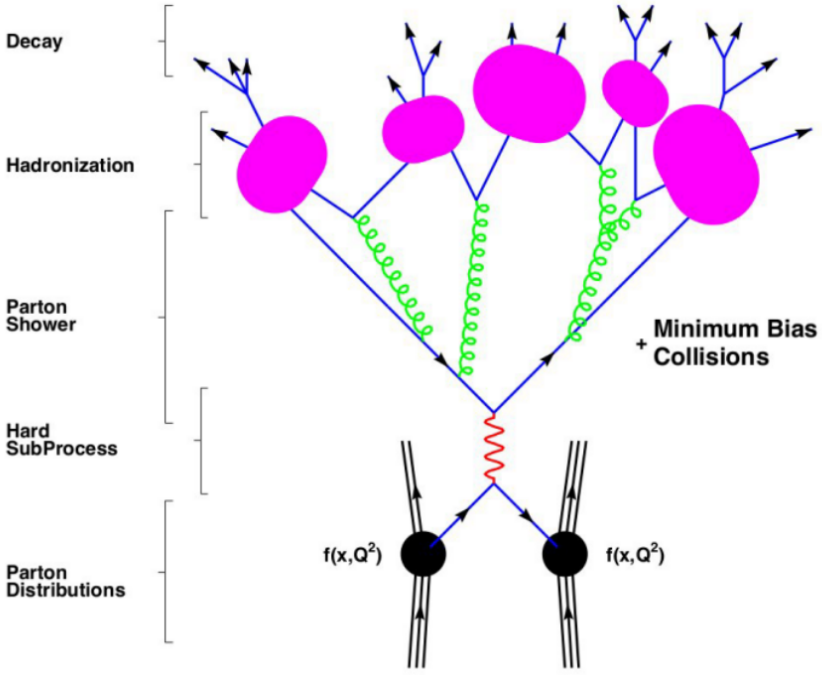

Figure 3.10: The several steps of simulation in Monte Carlo generators

detector response. The reference MC signal sample simulating $t\bar{t}$ events used in this analysis is generated using Powheg [40] interfaced with Pythia [41] for the parton shower and the hadronization simulation. The PDF set employed is the CT10 [24] one, with the top quark mass fixed to 172.5 GeV. In order to estimate the generator systematic uncertainty, the Powheg+Pythia predictions are compared with the ones obtained from other generators: Mc@Nlo [42], Herwig [43] and Jimmy [44] for the modeling of multiple parton scattering. The parton shower model uncertainty in particular is obtained comparing the samples generated by Powheg interfaced with Pythia and Herwig. The simulated events are weighted such that the average number of interaction in each bunch crossing agrees with the ones in data. They are also normalized to the latest QCD theoretical calculation at the next-to-next-to leading order (NNLO) for a quark mass of 172.5 GeV. The cross section considered at $\sqrt{s} = 8$ TeV is $\sigma_{t\bar{t}} = 253^{+13}_{-15} pb$ [46] comprehensive of the next-to-next-to-leading logarithmic (NNLL) soft gluon terms. The Monte Carlo simulations are of fundamental importance in every particle physics analysis representing the sum of the present theoretical and experimental knowledge. Comparing it with the obtained data result it is possible to test our comprehension of the physics processes and check if there are some discrepancies between the experimental outcomes and the expectations.

3.3 Background sources

Background events are represented by not $t\bar{t}$ events having the same signature of the $t\bar{t}$ ones or passing the selection cuts due to the detector misidentification. The greater background sources are: the single top, the leptonic decays of the W or the Z boson in association with high p_T jets (W or Z plus jets), the multijet events (can lead to a misedentification of a jet in an electron, also called fake leptons or QCD background) and the diboson (WW,ZZ,WZ) events. The $t\bar{t}$ pair events decaying completely leptonically are also a source of background for this analysis channel.

Background contributions are mainly estimated through Monte Carlo simulations but the fake leptons and W+jets background have been calculated directly via techniques that use the observed data (*data driven* method).

3.3.1 Monte Carlo Background

The decay of the Z and W bosons produced with up to five additional partons are produced with the ALPGEN [45] generator interfaced to Pythia for parton showering. W+jets samples are corrected in normalization and shape applying additional scale factors derived from data as described in the next Paragraph. The production of single top quark is simulated using a different generator in each channel: the t channel is realized using the AcerMc generator while production in the s channel and in association with a W boson are modeled with Powheg. Both these generators use the CTEQ6L1 set of PDF and are interfaced with Pythia for the parton showering modeling. Diboson production is modeled using Herwig and Jimmy with the CTEQ6L1 PDF set. The dilepton background is simulated using Powheg+Pythia in association with the $t\bar{t}$ semileptonic channel, the separation of the two samples is performed by the offline analysis.

3.3.2 Data Driven Background

Data driven algorithms allow to estimate partially or entirely some background sources from data. In this analysis the fake lepton background is fully calculated from data while regarding the W+jet background the ALPGEN samples are corrected both in shape and overall normalization via data driven methods.

W+jet scale factors

In the W+jets background two scale factors are calculated by data driven methods: the first one normalizes the different flavor component (W + $b\bar{b}$, $c\bar{c}$, c and light quarks) and the other is an overall normalization factor.

The fractions of different flavor component simulated using MC generator are rescaled by a set of scale factors extracted from a control region of the data where the selection requirements make the W+jets to be dominant. This region is determined applying on data the same selection cuts of the standard analysis except for the request of a b-jet and the requests on η, mass and $\sqrt{d_{12}}$ of the large-R jet. Furthermore the existence of exactly two jet is required. The W+jets background is composed by heavy flavour components (W + $b\bar{b}$, $c\bar{c}$, c) and a light flavour one (Wl). The number of b-tagged events $N^{W,tag}_{Njet=2}$ can be expressed as a function of the number of events before b-tagging $N^{W,pretag}_{Njet=2}$ multiplied with flavour-depending b-tagging probabilities $P_x(x = b\bar{b}, c\bar{c}, c, l)$:

$$N^{W,tag}_{Njet=2} = N^{W,pretag}_{Njet=2}(P_{b\bar{b},2}F_{b\bar{b},2} + P_{c\bar{c},2}F_{c\bar{c},2} + P_{c,2}F_{c,2} + P_{l,2}F_{l,2}) \tag{3.8}$$

where $F_x(x = b\bar{b}, c\bar{c}, c, l)$ are the flavour fraction (HFF) of pre-tagged events, the scale factors to be determined. The b-tagging probabilities are determined by MC studies while the $N^{W,tag}_{Njet=2}$ is determined from MC and data as follows:

$$N^{W,tag}_{Njet=2} = N^{data}_{Njet=2} - N^{MCnoW}_{Njet=2} - N^{QCD}_{Njet=2} \tag{3.9}$$

There are other two conditions that allow the determination of HFF:

$$F_{b\bar{b}} + F_{c\bar{c}} + F_c + F_l = 1 \tag{3.10}$$

$$F_{c\bar{c}} = k_{c\bar{c}\to b\bar{b}}F_{b\bar{b}} \tag{3.11}$$

The ratio $k_{c\bar{c}\to b\bar{b}}$ between W$b\bar{b}$ and W $c\bar{c}$ is determined from MC. The method first estimates HFF from 2-jet multiplicity, the one with higher statistic and lower uncertainties, and than extrapolate HFF for higher multiplicity. The factors used for this analysis have been derived from the group that research for $t\bar{t}$ resonance [47] and had been reported in table 3.1:

The normalization of the overall yields of W+jets are then performed by the comparison of the MC predicted charge-asymmetry with the charge-asymmetry of W boson production observed from data. The equation used to derive the factors

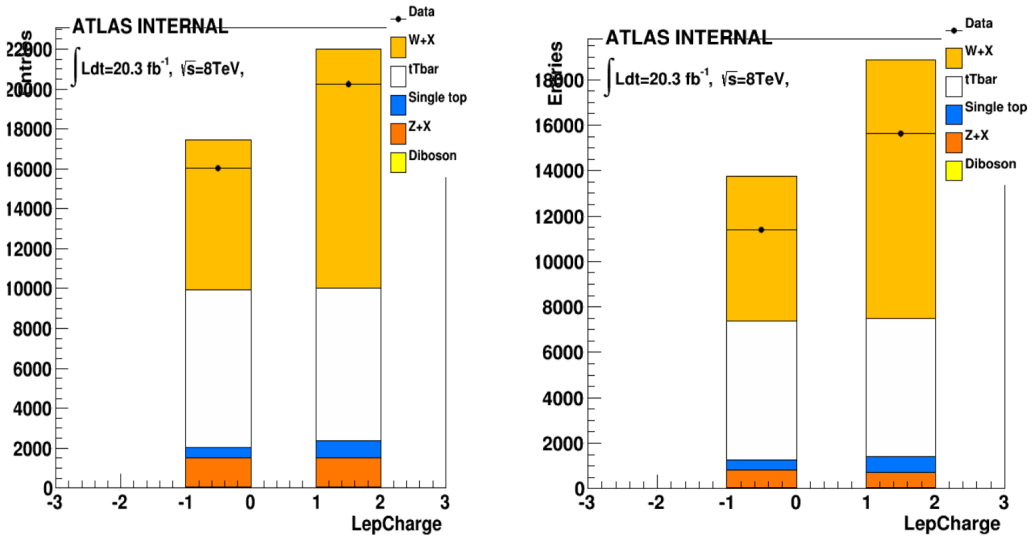

Figure 3.11: On the left the distribution of event in the electron channel, divided on the base of the leptonic charge. On the right the same for the muon channel.

is:

$$N_{W^+} + N_{W^-} = \left(\frac{r_{MC}+1}{r_{MC}-1}\right)(D_{corr+} - D_{corr-}), \tag{3.12}$$

where r_{MC} is the ratio between the number of positive W bosons to the negative, after the selection, calculated from the MC simulation and $D_{corr-}(D_{corr+})$ is the number of observed events with a negative(positive) lepton. All the symmetric contributions (Z+jets, fake leptons) are canceled in the difference while the asymmetric contribution deriving from the single top is estimated using MC simulation. For this factor, as for the flavor's one, the region selected for the comparison must contain a large sample of W+jets so the cuts applied are less selective than the ones applied in the full analysis, in particular there are no requirements on b-jet and on the internal characteristics of the jet with large R, it only must have p_T higher than 300 GeV. The histogram obtained for charge asymmetry distributions in both muon and electron channel are reported in figure 3.11,the values obtained are:

channel	scale factor
electron	0.88 ± 0.04
muon	0.79 ± 0.03

Fake Leptons background

The fake leptons (multi-jet) events are determined directly from data using a particular data driven method called *matrix method* [48]. The first step to apply this method is to define two regions, *loose* and *tight* that differ only on the definition on lepton cut. The *tight* selection is the same used in the analysis while in the *loose* selection is removed the isolation requirement. The number of events passing the two selections (N_{tight} and N_{loose}) can be expressed as the sum of the correspondent number of signal (real) events ($N_{real}^{\text{tight}(loose)}$) plus the number of events from the lepton misreconstruction ($N_{fake}^{\text{tight}(loose)}$):

$$N^{tight} = N_{real}^{tight} + N_{fake}^{tight}$$
$$N^{loose} = N_{real}^{loose} + N_{fake}^{loose}$$

Defining efficiencies as $\epsilon_{real} = \frac{N_{real}^{tight}}{N_{real}^{loose}}$ and $\epsilon_{fake} = \frac{N_{fake}^{tight}}{N_{fake}^{loose}}$ it is possible to derive the number of fake-lepton as:

$$N_{fake}^{tight} = \frac{\epsilon_{fake}}{\epsilon_{real} - \epsilon_{fake}}(\epsilon_{real} N^{loose} - N^{tight}) \tag{3.13}$$

The ϵ_{real} is estimated from a $Z \to e^+e^-$ events (that do not contains fake-lepton contamination) on which is applied the same selection of the analysis, except for jet dependent requirements. The ϵ_{fake} is evaluated on a $t\bar{t}$ sample as the ratio of events in which the selected *loose* lepton also pass the *tight* requirements, divided by the total number of *loose* events.

3.4 Event Selection

The $t\bar{t}$ event candidates both from collected data and from MC generators must satisfy the following requirements, summarized in figure 3.12.

- Events must belong to the so called good run list (GRL) of events acquired when all detectors work properly;
- A reconstructed primary vertex with five or more associated tracks and no LAr or Tile calorimeter corrupted data;
- One good lepton (electron or muon) with p_T>25 GeV and $|\eta|$<2.5
- No good leptons of different flavour;

El Channel				μ Channel			
Njets	F_{bb}/F_{cc}	F_c	F_{ll}	Njets	F_{bb}/F_{cc}	F_c	F_{ll}
1	1.55	1.33	0.91	1	1.74	0.96	0.97
2	1.47	1.26	0.86	2	1.66	0.92	0.92
3	1.42	1.21	0.83	3	1.58	0.88	0.88
4	1.38	1.18	0.81	4	1.52	0.84	0.84
5	1.33	1.14	0.78	5	1.45	0.80	0.81

Table 3.1: In the first column is indicated the number of jets, in the other columns is reported the scale factor used depending on quark flavor.

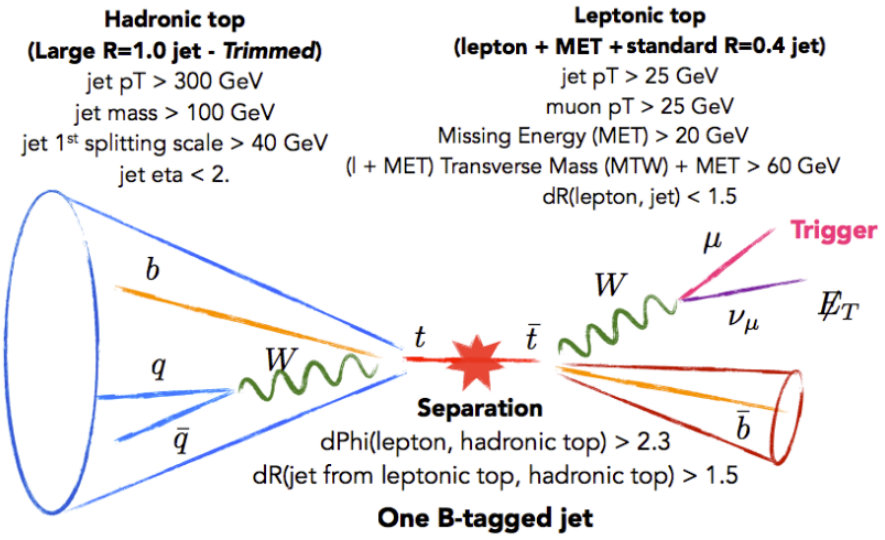

Figure 3.12: All the requirements for the signal signature used in the analysis

- $\not{E}_T > 20$ GeV and the sum between $\not{E}_T$ and the transverse mass ($m_T = \sqrt{E_T^2 + p_T^2}$) of the W boson must be: $\not{E}_T + m_T^W > 60$ GeV;
- At least one small-R jet with p_T >25 GeV close to the lepton, ΔR(lepton,jet)<1.5;
- At least one large-R jet with p_T>300 GeV, $m^{large-Rjet} > 100$ GeV, $\sqrt{d_{12}}$>40 GeV, ΔR(large-R jet,small-R jet)>1.5, $\Delta\phi$(large-R jet,lepton)>2.3. If in the event there are more than one small-R jet or large-R jet that follows the requirements on p_T and internal variables only the one with highest p_T is considered as candidate.
- At least one small-R jet b tagged jet matched with one of the two top quarks. The b quark jet identified is considered to match with the hadronic top if it is within a cone ΔR<1 from large-R jet axis while it is considered to match with the leptonic top if it correspond to the jet previously used to reconstruct the leptonic top quark. Following this definitions the event had been divided for some systematics studies in three categories:
  - **lep-btag** There is one b-jet only on the leptonic side
  - **had-btag** There is at least one b-jet only in the hadronic side
  - **two-btag** There is at least one b-jet on both sides

Electron Channel				
Sample	Events	two-btag	had-btag	lep-btag
Data	4114	1562	1518	1034
MC $t\bar{t}$ signal	4338	1767	1415	1156
W+jets	266	10	188	68
dilepton	243	76	65	102
stop	131	27	62	42
QCD background	91	35	34	22
Z+jets	48	2	39	7
dioboson	3	0	2	1

Table 3.2: In the first column are shown the events that passed the selection both from data and MC (signal and all background sources) in Electron channel, normalized to the same luminosity of 20.3 fb^{-1}.

The total amount of event passing the event selection for data, MC signal and background in both electron and muon channel is reported in the tables 3.2 and 3.3

Muon Channel				
Sample	Events	Both btag	Had btag	Lep btag
Data	3602	1411	1267	924
MC $t\bar{t}$ signal	3648	1501	1160	987
W+jets	248	9	180	59
dilepton	220	67	48	104
stop	131	32	55	44
diboson	2	0	1	1
Z+jets	29	2	18	9
QCD background	3	1	1	1

Table 3.3: In the first column are shown the events that passed the selection both from data and MC (signal and all background sources) in Muon Channel, normalized to the same luminosity of 20.3 fb^{-1}. In the other columns are reported the events divided for the b-tagging categories.

3.4.1 Particle level definition

The analysis results will be compared to the theoretical predictions in a fiducial particle level phase space instead to the full phase space. The particle level fiducial phase space is used in order to eliminate the theoretical and detector dependences; it can be defined by applying on truth MC stable objects a selection similar to the reconstructed one. The stable particle level objects are defined as:

- **leptons**: They are clusterized with all stable photons within a cone of radius R=0.1 (*dressed* leptons); the ones coming from a quark or a hadron are excluded.
- **jets**: They are reconstructed using the anti k_t algorithm starting from all stable truth particles, with the exception of those indirectly matched to a W boson.
- $\not{E}_T$: is given by the sum of all neutrinos coming from a W boson

The particle level selection follows:

- Only one lepton with $p_T > 25$ GeV and $|\eta| < 2.5$;
- No other flavour leptons;

- The $\not{E}_T$ must be higher than 20 GeV and $(\not{E}_T + M_T^W)$>60 GeV;
- A small-R jet with p_T higher than 25 GeV close to the lepton, $dR(lepton, jet) < 1.5$;
- A large-R jet with p_T higher than 300 GeV distant from the lepton, $dR(lepton, large-Rjet) > 1.5$, and from the leptonic jet, $dR(leptonicjet, large - Rjet) > 2.3$, identified as the hadronic top;
- One bjet matched to one of the two top quarks. The same three categories of the reconstructed level are obtained.

The particle level has the advantage to make the results independent from the MC generator used and to reduce the systematic uncertainties.

3.4.2 Physical Object definition

After the reconstruction of the physical objects it is necessary to reconstruct from the lepton the $\not{E}_T$ and the jets, the two top quarks initially produced. The hadronic top quark is considered to exactly correspond to the reconstructed large-R jet, that undergone the trimming procedure, with higher p_T among all the large-R jets individuated in the event. The leptonic top quark is instead reconstructed as the four-vectors sum of the W boson and the small-R jet with higher p_T in a $\Delta R < 1.5$ around the lepton. The W boson itself is reconstructed as a four vector sum between the lepton and the reconstructed $\not{E}_T$.

At the particle level the physical object are defined following the same prescriptions of the reconstructed level but starting from the truth stable objects presented in section 3.4.1.

3.5 Data/MC comparison

In this Paragraph the comparison between generated MC and data distributions are shown for some of the principal quantities of physical interest for both muon

and electron channel. This is done to check the agreement between the simulations and the real data before the final cross section estimation.

In the following plots, from figure 3.13 to 3.16, real data are represented as black dots while the background sources are reported in various color as indicated in the legend. The QCD background, completely *data-driven*, is reported only in the hadronic top p_T distributions. The background distributions are stacked with the MC signal, in white, to allow the comparison with data. In all figures the electron channel is reported on the left while the muon channel is on the right. The binning has been decided in order to reduce the event migration and to reduce statistical fluctuations. In the distributions on lepton, hadronic and leptonic top quark kinematic variables the systematic uncertainties are reported as gray band. In the bottom part of each plot the ratio between data and the sum of MC signal and background is represented, in case of perfect accordion it should be 1, as indicated by the red line.

All the distributions show an agreement in shape between the data and the MC expectations but a constant discrepancy of the order of 20-30% in normalization for all variable distributions. The origin of this discrepancy is still under investigation.

3.5.1 Study on b-tagging and leptonic top reconstruction

In order to understand the source of the discrepancy between data and MC predictions, I personally carried on some studies about the effect of the b-tagging and the leptonic top quark definition on the event selection. I started looking at the kinematics distributions obtained in three separated regions defined from different requirements on the b-tagging cuts, as described in the event selection 3.4:

- **lep-btag:** The only b-tagged jet in the event is the one selected to reconstruct the leptonic top quark.

- **had-btag:** There is at least one b-tagged jet in a cone of around $\Delta R = 1$ from the axes of the large-R jet candidate of the hadronic top quark. The jet used to reconstruct the leptonic top quark is not btagged.

- **two-btag:** There is at least one bjet in a $\Delta R < 1$ from the axes of the large-R jet candidate to be a top quark and the jet selected to reconstruct the leptonic top quark is b-tagged.

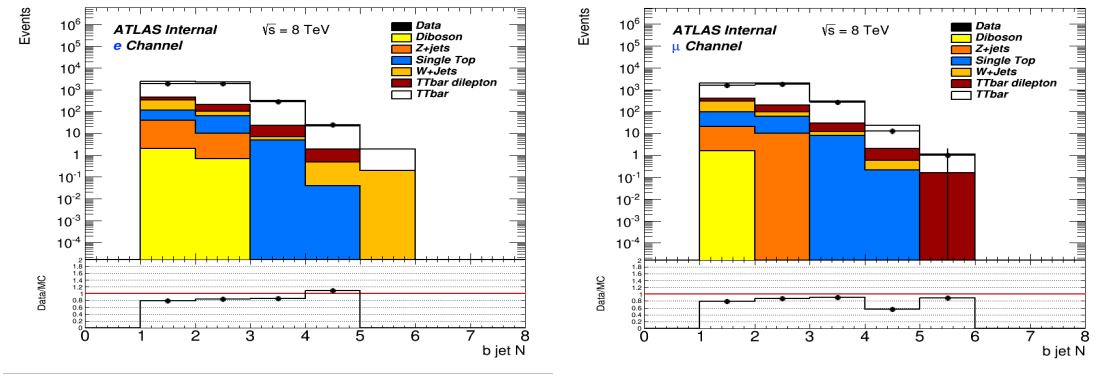

Figure 3.13: Distributions of the number of jets individuated as deriving from a b quark in electron (left) and muon channel (right). Data are compared with simulated background sources, data driven QCD and signal generated with Powheg+Pythia.

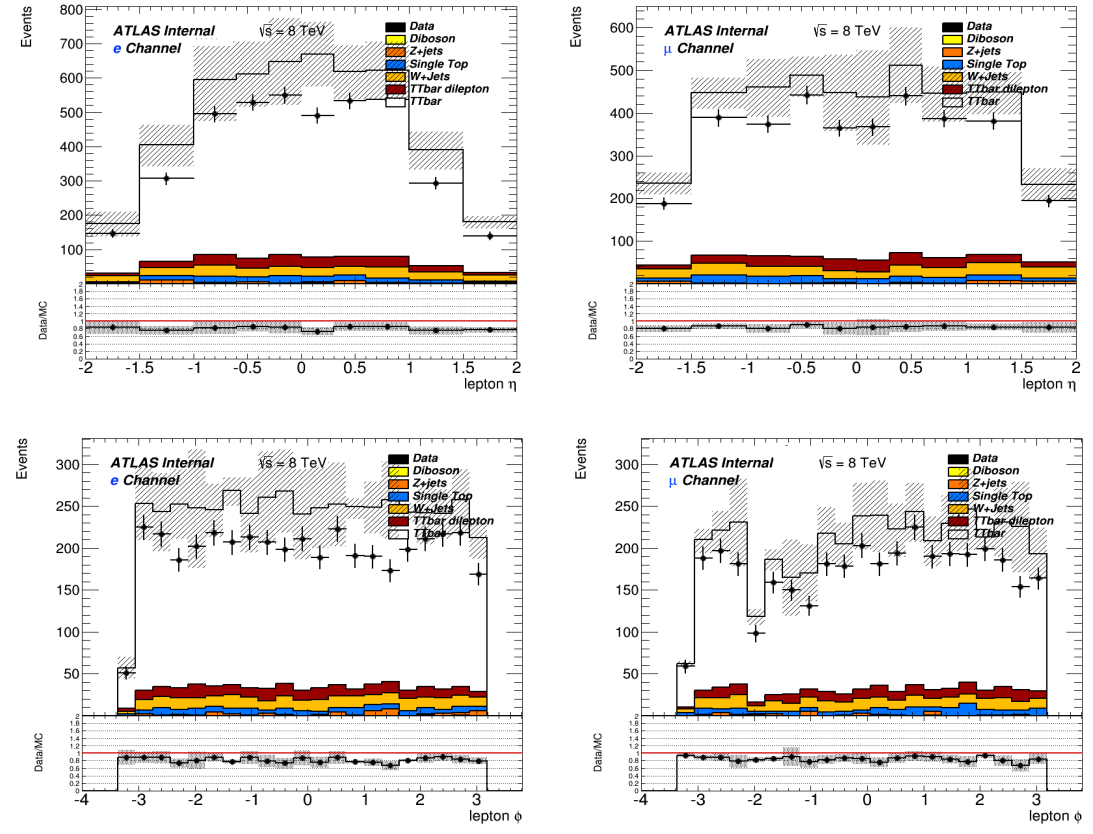

Figure 3.14: Distributions of the η (first row) and ϕ (second row) of the lepton. Data are compared with simulated background sources, data driven QCD and signal generated with Powheg+Pythia.

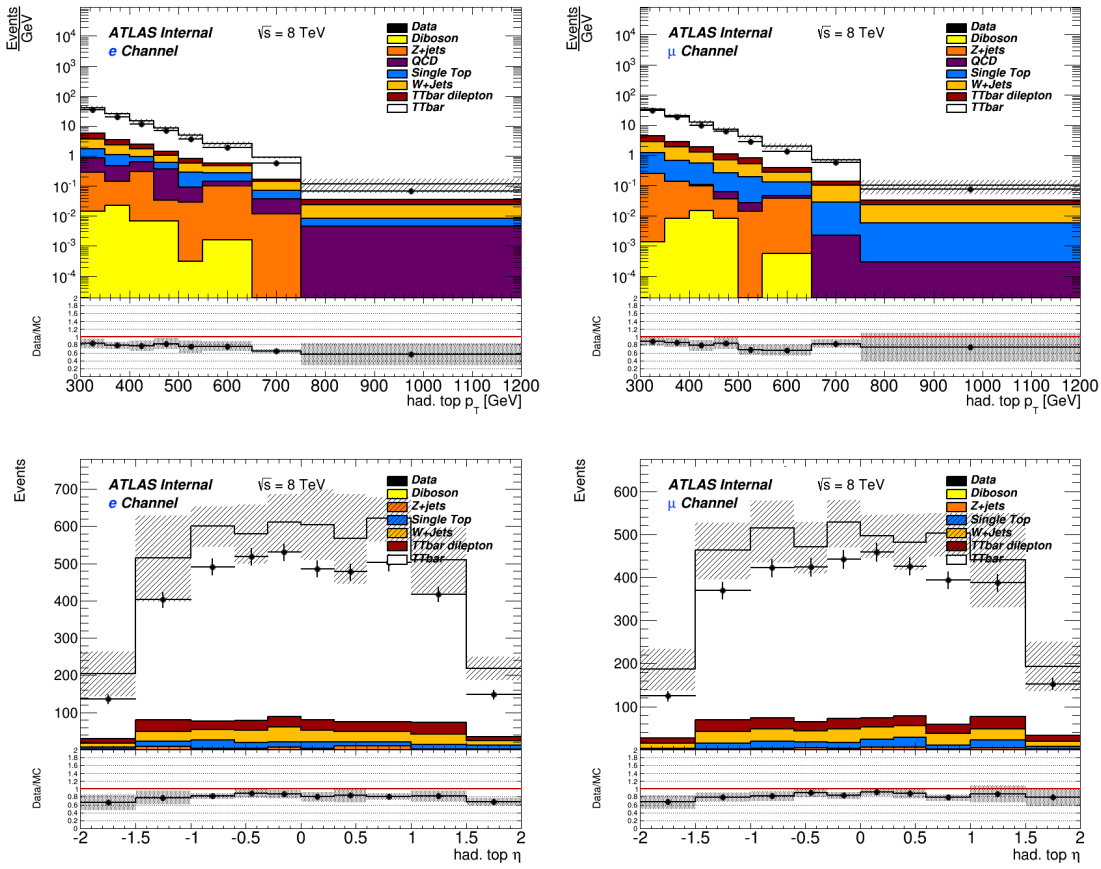

Figure 3.15: Distributions of the η and p_T of the top hadronically decaying. Data are compared with simulated background sources, data driven QCD and signal generated with Powheg+Pythia.

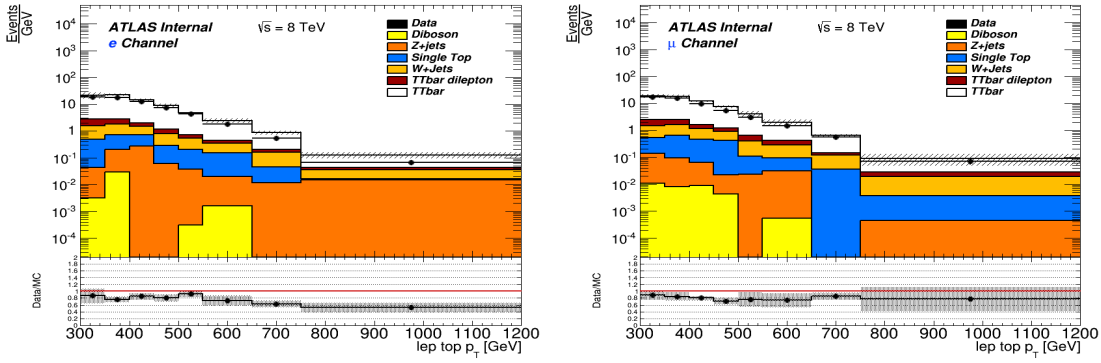

Figure 3.16: Distributions of the p_T of the top leptonacally decaying. Data are compared with simulated background sources, data driven QCD and signal generated with Powheg+Pythia.

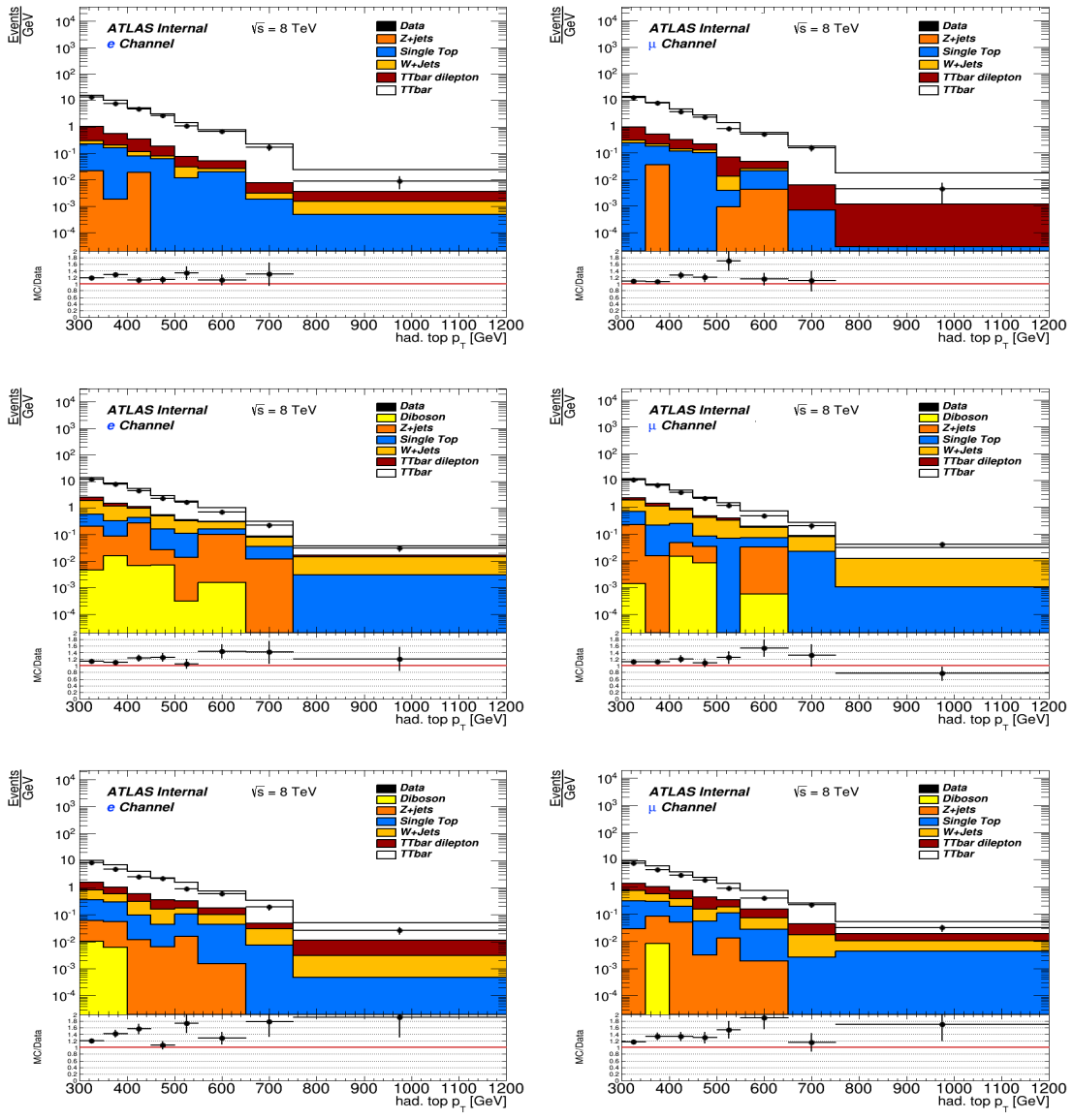

Figure 3.17: Distributions of the p_T of the top hadronically decaying with event devided in the three btag category: both tag (first row), had tag (second row) and lep tag (third row).

The MC/data comparison with respect to the hadronic top p_T in the three different categories are reported in figure 3.17. The better agreement is reached in the events belonging to the *had-btag* category; here the discrepancy decrease from $\sim$ 20 % to $\sim$ 10%. The worst agreement is shown by the *lep-btag* distribution; this can also be directly seen in the overall yield in tables 3.2 and 3.3, where the events in the three regions are reported separately. The fact that the MC signal contribution alone overcomes the amount of data, could indicate as possible source of the disagreement the definition of the *lep-btag* category or the mismodelling of b-tagging efficiency rather than other sources like for example the background estimation. Some control plots have been performed to suggest a new leptonic top quark definition:

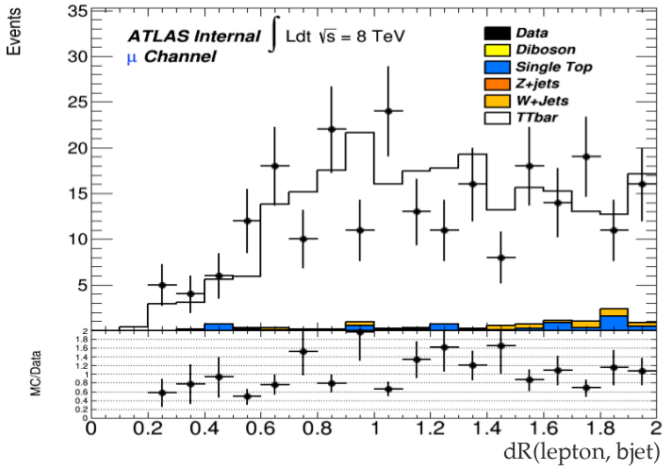

Figure 3.18: Distribution of event in had btag category with respect to the ΔR(lepton,nearest bjet in the event)

- Figure 3.18 shows the number of event in *had-btag* category with respect to the ΔR between the lepton and the nearest b-tagged jet in the event. Due to the fact that the events reported belong to the *had-btag* category, the jet used to reconstruct the leptonic top is not b-tagged. This distribution shows that in about the 12% of events in *had-btag* category there is a b-tagged jet close to the lepton but not selected because it is not the one with higher p_T within a range R=1.5 from the lepton.

- Figure 3.19 shows the number of events in the *two-btag* category with respect to the ΔR between the lepton and the jet used to reconstruct the leptonic top quark in the standard analysis. In this distribution there is a peak at ΔR

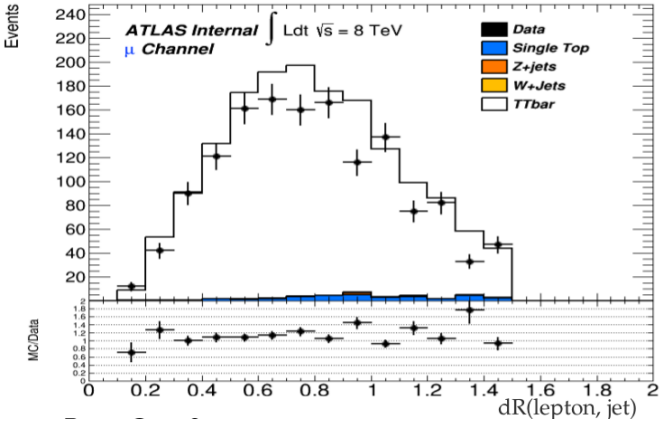

Figure 3.19: Distributions of events in in *two-btag* category with respect to the ΔR(lepton, leptonic top jet)

$= 0.8$ and the 80% of the events are in the $\Delta R < 1$ region. In the remaining part of the events a larger discrepancy than at small ΔR is observed.

- Figure 3.20 shows the number of events in the *two-btag* category and the

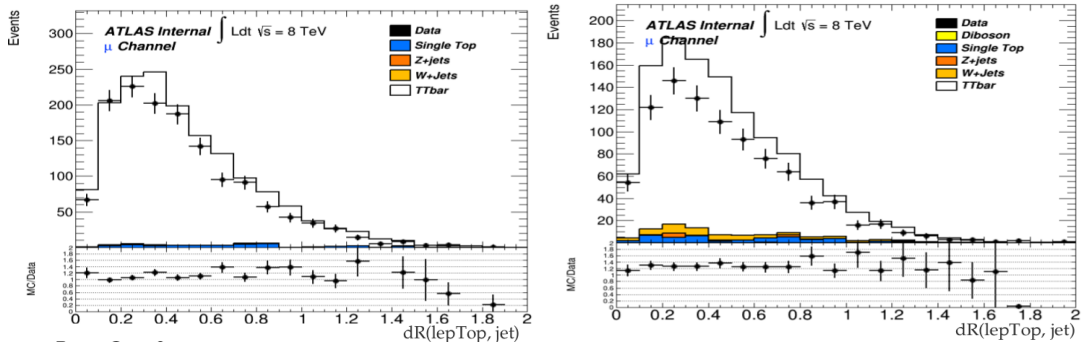

Figure 3.20: Distributions of events in in both btag category (left) and lep btag category (right) with respect to the ΔR(lepton, leptonic top jet)

lep-btag category with respect to the ΔR between the jet used to reconstruct the leptonic top quark and the reconstructed top quark itself. In the 90% of events the $\Delta R(lepTop, jet)$ is less than 0.8; in the remaining region there is an higher discrepancy between data and MC.

Starting from this control plots I have defined and compared four different ways to reconstruct the leptonic top checking if there are any improvements in data/MC agreement. The jet selected to reconstruct the top quark is:

- **Standard analysis**: the one with higher p_T in a cone around the lepton of dR(lepton, jet) <1.5

- **Lep Top Reco 1**: the nearest b-tagged jet to the lepton, with a maximum dR(lepton, jet) =1.5
- **Lep Top Reco 2**: is the one with higher p_T in a cone around the lepton of dR(lepton, jet) <1.0
- **Lep Top Reco 3**: the one with higher p_T that satisfy the condition dR(leptonicTop, Jet) <0.8

In figures 3.21 and 3.22 the comparisons between the observed data and the Monte Carlo expectations using the three leptonic top reconstruction discussed before are presented. On the left the events coming from all the *two-btag*, *had-btag*, and *lep-btag* categories are reported while on the right the distributions containing only the events from the *lep-btag* category (the one that gives the worst agreement in the standard analysis) are shown. In table 3.5 the percentage difference between the total amount of data after background subtraction and the MC signal obtained using the four different leptonic top quark reconstruction methods are reported.

Reconstruction mode	All events	tow-btag	had-btag	lep-btag
Standard analysis	21%	21%	2%	37%
LepTop Reco 1	18 %	13%	14%	26%
LepTop Reco 2	17 %	11%	20%	28%
LepTop Reco 3	17 %	10%	20%	26%

Table 3.4: The percentage of discrepancy between data background subtracted and MC is shown for the four different top reconstruction. The values reported are affected by a statistical error that goes from 2% to 5%, depending on the bin statistic.

All the methods implemented for the leptonic top reconstruction increase the agreement between MC and data both for the events in *lep-btag* category and in *two-btag* category, but enlarge the discrepancy in the *had-btag* one. The overall effect on the full sample using any new leptonic top reconstruction method is to improve of a few percent the agreement between the observed and theoretical events; instead there is not a great impact on the differential distributions.

The last check to identify which method is more efficient in the leptonic top quark reconstruction is based on the MC simulations. It consists verifying if the jet used to reconstruct the leptonic top matches a b quark at particle level (before detector simulation). This can be done by checking if there is a particle-level b quark in a cone R=0.4 around the reconstructed jet axis.

Reco. mode	St. analysis	Lep Top Reco 1	Lep Top Reco 2	Lep Top Reco 3
Matching %	84.89%	73.05%	71.53%	70.67%

Table 3.5: The percentage of matching between the leptonic top jet and the particle b for each leptonic top reconstruction method.

In table 3.5 the percentage of matching for each method is reported. The method used in the standard analysis individuates the leptonic top is 10% better than the ones of the other three methods.

The three methods studied at the moment do not show a great improvement in data MC agreement and but several tests are still under investigation to continue the study on this item. This studies verified that the leptonic top reconstruction method used in the standard analysis is consistent and the source of discrepancy between data and MC could derive from other source as a bad b-tagging efficiency simulation.

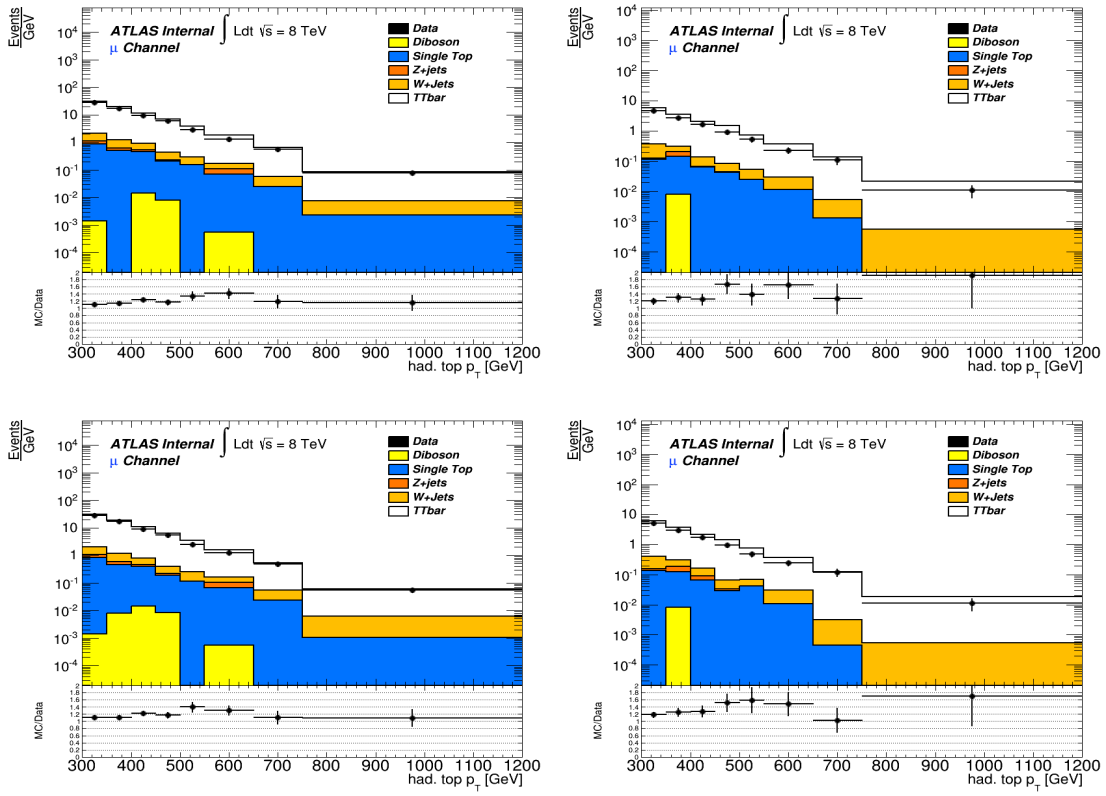

Figure 3.21: Distributions of events in all category (left) and *lep-btag* category (right) with respect to the hadronic top p_T. In the distributions in the first row is used the second leptonic top quark reconstruction while in the bottom row is used the third leptonic top quark reconstruction.

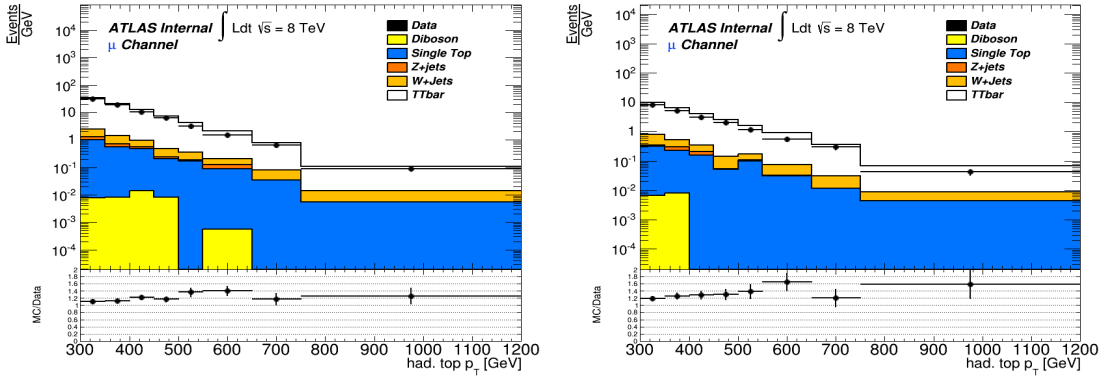

Figure 3.22: Distributions of events in all category (left) and lep btag category (right) with respect to the hadronic top p_T. In the distributions in the used the first leptonic top quark reconstruction.

3.6 Unfolding

The aim of experimental particle physics is to determine the distribution $f(x)$ of some physical variable x starting from a set of measurements.

In a experiment with a "perfect" apparatus the measured distributions can differ from the true distributions only by a statistical error $\epsilon(x)$ but a real apparatus gives rise to additional distortions. Two effects interfering with the measure arise from the limited acceptance and resolution: each detector can detect particles only on a limited phase space region (geometrical acceptance) and with a limited resolution that conducts to loose good events for trigger/ reconstruction/ selection efficiency. Generally the sum of all these contributions determines the overall detection efficiency of the experimental apparatus. It is necessary to deal with this effect by applying a bin dependent scale factor, evaluated from Monte Carlo, to rescale data to the predicted number of events.

The limited resolution implies a limited precision on the measurements and can cause a distorted reconstruction of the kinematic variables with respect to the real value. This leads to a statistical smearing between the true kinematic variable x and the measured quantity y. This effect can be mathematically represented by the Fredholm integral [53]:

$$g(y) = \int A(y,x) f(x) dx \tag{3.14}$$

where $g(y)$ represents the measured distribution and $f(x)$ the true one. The resolution function $A(y,x)$ represents the detector distortion. The unfolding consists

in evaluating the true distribution $f(x)$ from a given $g(y)$ and $A(y,x)$.

In particle physics applications the distributions $f(x)$ and $g(y)$ are discrete, represented as histograms, so the equation 3.14 may be written using a limited number of elements $f_j,\ ,j = 1...n$ and $g_i,\ ,i = 1...m$:

$$g_j = \sum_i A_{ij} f_i \tag{3.15}$$

where f is a n-dimensional vector, g is a m-dimensional vector and A is a matrix (m,n). The matrix A, called *Response Matrix* or *Migration Matrix*, usually is not diagonal due to some events that instead of be reconstructed in a certain bin j are reconstructed in a bin $i \neq j$, giving rise to a migration of events from one variable bin to another. The diagonal elements represent the bin by bin reconstruction efficiency of the quantity x in the bin j.

The inversion of a finite system of equation rarely admits an exact solution, so several techniques calculating approximate solutions have been developed. In order to overcome the intrinsic instability of these approximate solutions, some regularization conditions can be imposed, based on some a priori information (as the request of minimum curvature or that the solution must be strictly positive [54]). These methods usually allow the suppression of spurious oscillating components of the unfolded solution. The reference method used in this analysis is the Singular Value Decomposition (SVD).

3.6.1 SVD

The SVD method [55] is based on the decomposition of the (m,n) migration matrix A (with m≥n) in three matrices:

$$A = USV^T \tag{3.16}$$

where U and V are square orthogonal matrices with dimensions m and n respectively and S is a diagonal matrix of dimensions (m,n) with elements null or positive. The

three matrices have the following properties:

$$U^T U = UU^T = 1$$
$$V^T V = VV^T = 1$$
$$S_{ij} = \lambda_i \delta_{ij} \text{ e } \lambda_i \geq 0$$

This decomposition of the migration matrix makes the inversion of the matrix A easier, since only the diagonal matrix S must be inverted:

$$(A)^{-1} = (USV^T)^{-1} = (V^T)^{-1}(S)^{-1}(U)^{-1} = VS^{-1}U^T \tag{3.17}$$

The exact solution of the equation 3.16 leads in some cases to high oscillating distribution that can be suppressed using some a priori knowledge about the solution. Technically this can be achieved by adding a regularization term to the system and minimizing it:

$$(Ax - y)^\tau (Ax - y) + \tau \cdot (Cx)^\tau Cx = min. \tag{3.18}$$

where τ is the regularization parameter that determines the relative weight of the initial condition imposed on the solution and represented as the matrix C. In particle physics the C matrix is usually chosen from the theoretical distributions obtained from MC simulations, while the optimal value of τ is problem dependent and must be determined from case to case.

3.6.2 Response Matrix

The migration matrices obtained from MC distributions and used in this analysis are reported in figure 3.23 for electron channel and in figure 3.24 for muon channel for the variable transverse momentum of the hadronic top. On the y axis is reported the particle level variable while the reconstructed MC is reported on the x axis; inside each bin there is the number of events normalized to 100. The bins width has been evaluated in order to have about the 60% of statistics in the diagonal ($\pm$ σ). The matrices are almost diagonal, these means that there is a small migration.

Before the calculation of the response matrix the reconstructed MC sample has

been rescaled to the particle sample applying the acceptance, defined as:

$$\text{acceptance} = \frac{\text{number of events passing particle level and reco selection}}{\text{number of events passing reco selection}}$$

The distributions resulting from the application of the migration matrix will be furthermore rescaled by the ratio with the efficiency, defined as:

$$\text{efficiency} = \frac{\text{number of events passing particle level and reco selection}}{\text{number of events passing particle level selection}}$$

3.7 Analysis framework

The necessary requirements for a standard code in the ATLAS experiment is to be projected in order to share the information and all the analysis tools to all the collaboration.

All the program architecture has a high flexibility and allow to perform many studies complementary to the standard analysis. The program devoted to the analysis performed in this thesis is TopD3PDBoosted, that inherits all the tools from the main framework TRC (Top Root Core), builds to implement all the analysis involving top. This structure allows to easily maintain event corrections and scale factors up to date and to distribute useful tools among all the ATLAS collaboration. The programs included in TopD3PDBoosted allow to perform all the analysis chain, from the event selection until the histogram production.

I have collaborated to develop and implement all the different codes that allow to perform the entire analysis chain. In particular I have implemented part of the event selection at reconstructed and particle level, worked on the event reconstruction and implemented part of the code that actuate the Data/MC comparison beside implementing the code to evaluate b-tagging and leptonic top reconstruction efficiency.

The structure of the package is completely object based and the codes are implemented using highly hierarchical coding. The general analysis program architecture is the following:

- **D3PD2MiniSLBoosted** Reads the ntuple in the D3PD format, applies them the first steps of the event selection and convert them in RooTree.
- **HistMakerMaster** Reconstruct the four momentum of physical object as

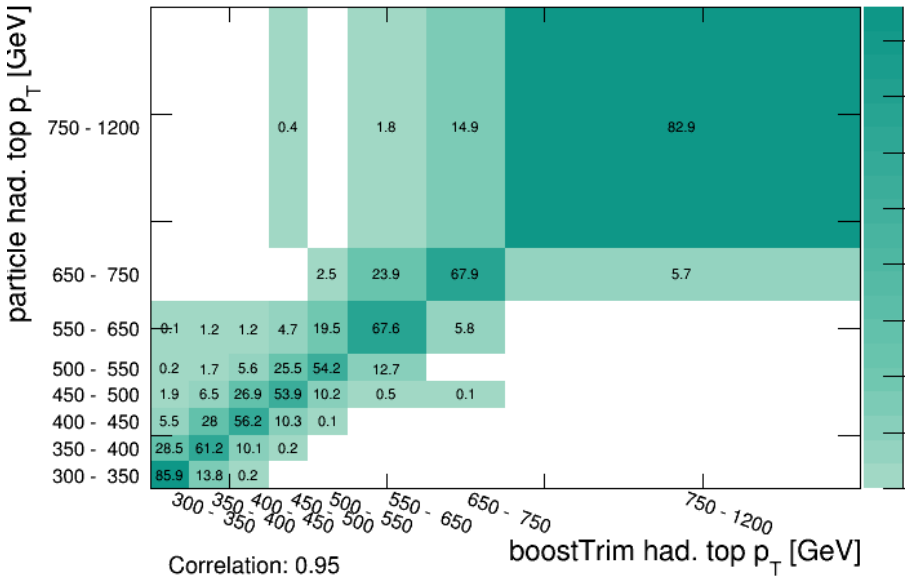

Figure 3.23: The migration matrices obtained from the Pythia+Herwig simulation relating the particle and reconstruction levels for in the electron channel. The linear correlation coefficient is given below each plot and all columns are normalized to unit

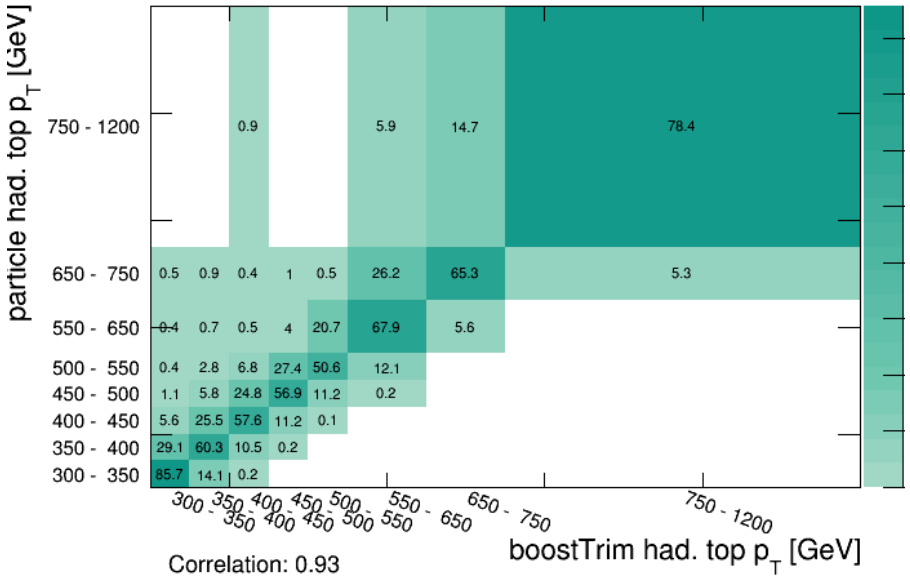

Figure 3.24: The migration matrices obtained from the Pythia+Herwig simulation relating the particle and reconstruction levels for in the muon channel. The linear correlation coefficient is given below each plot and all columns are normalized to unit.

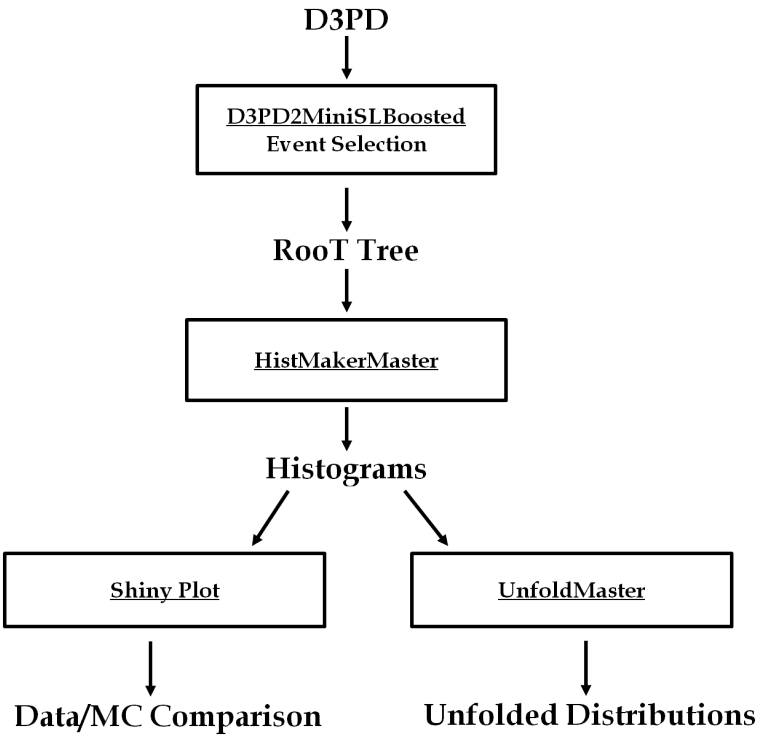

Figure 3.25: In the graph are summarized the main blocks of the analysis.

the neutrino, the W boson, and the top quarks, perform the last steps of the event selection and collect information in histograms.

- **ShinyPlot package** Change the histogram binning depending on the kinematic variable and the physical object and perform the comparison between data and reconstructed MC.

- **Unfold package** is a standalone package taking the histograms as input, perform the unfolding, and produce the comparison between the unfolded distributions and the particle-level distributions. This package provides also both the closure and stress test, I fully implemented the second one.

The packages treat and calculate also all the systematic uncertainties, that will be presented in Chapter 5, from the histogram until the propagation through the unfolding machine. The main blocks of the package are summarized in figure 3.25.

The TopD3PDBoosted package can be found on public web repository at the reference [56].

Chapter 4

PDF

Precise measurements, as the one performed in this thesis allow to test the underlying theoretical picture comparing experimental results with theoretical predictions. One of the crucial ingredients for any predictions involving hadrons interactions, is the partonic structure of the hadrons and, as described in the first Chapter, different possibilities (PDF sets) are available. Each Monte Carlo generation to simulate hadron interactions needs a PDF set as input and different sets can lead to different predictions. A precise measurement of the $t\bar{t}$ differential cross section can be sensitive to effects related to the different PDF sets, allowing a discrimination among the different possible choices.

The PDF modeling is one of the main source of systematic uncertainty and also one of the most complex to be treated. In the following section it will be illustrated how to calculate the PDF uncertainty using the MSTW2008NLO pdf. In this chapter will be also presented the comparison of the $t\bar{t}$ differential cross section measured at 7 TeV with NLO predictions obtained using different PDF sets as an example of the same study that will be performed in the 8 TeV analysis.

The measurement performed in this thesis cover a phase space region never studied before and the results could eventually be included in the fit to produce updated PDF sets with reduced uncertainties.

4.1 PDF systematic uncertainties

Each PDF set has one central distribution and several other distributions that slightly differ from the central one and are used to calculate PDF uncertainties. The Monte Carlo predictions, as illustrated in Chapter 1, also depend on the normalization and factorization scales that are involved in cross section calculation.

The signal has been generated setting both scales at the top mass (m_t= 172.5 GeV) and varying it to $2m_t$ or $m_t/2$ in order to estimate the corresponding uncertainty.

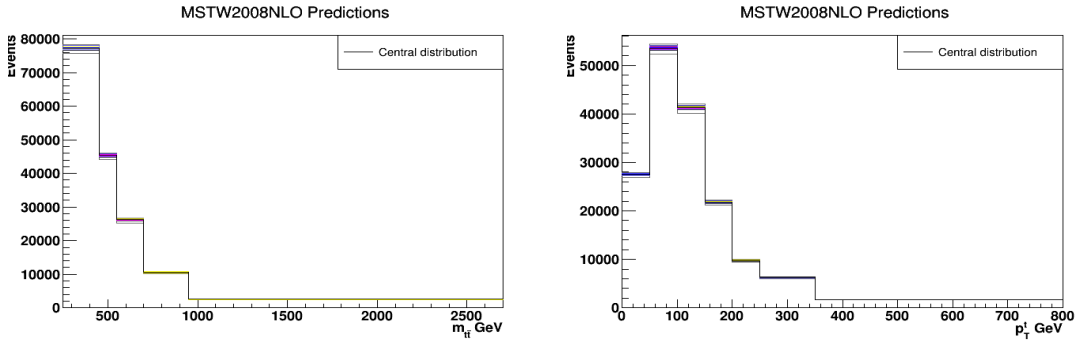

Figure 4.1: Events distributions of the $m_{t\bar{t}}$, on the left, and of p_T of hadronic top predicted using MCFM with the PDF set MSTW2008NLO. The central value is reported as black line while the band is formed by all the other MSTW2008NLO distributions.

All the sample used to perform PDF and scale error had been generated using MCFM [57], version 6.5. The program is designed to calculate cross-sections for various femtobarn-level processes at hadron-hadron colliders and takes as inputs scale values and the PDF family to be used. It performs predictions at the next-to-leading order QCD and does not include parton shower and hadronization.

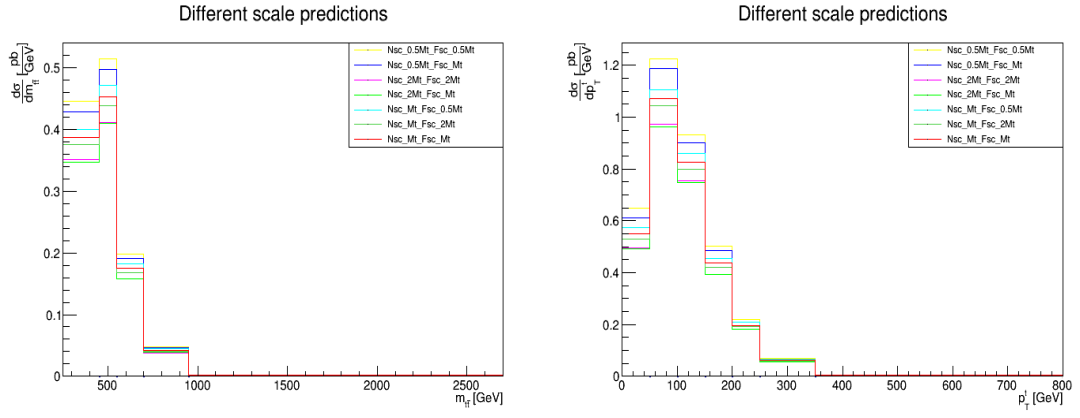

Figure 4.2: Distributions of the $m_{t\bar{t}}$, on the left, and of p_T of hadronic top predicted using MCFM with normalization scale (Nsc) and Factorization scale (Fsc) varied among m_t, $2m_t$, $m_t/2$

The MSTW2008NLO PDF set comprehends 40 distribution, 20 account for the positive fluctuations of the sources of uncertainty and 20 for the negative ones, aside the central distribution. An example of the predictions obtained using all the

distributions in MSTW2008NLO for hadronic top p_T and $m_{t\bar{t}}$ at 7 TeV is reported in figure 4.1.

Here the central distribution is reported as a black line while the band is composed from all the 20 sources of uncertainty (positive and negative variations corresponding to the same source are reported with the same color).

The predictions on kinematic variable distributions can be varied also modifying the normalization and factorization scale, an example of this effect is reported in figure 4.2, where distributions obtained varying the scale with respect to the nominal value are compared. The normalization and factorization scales can be varied independently or simultaneously. These two conditions lead to different errors as can be seen in figure 4.3, where errors on distributions obtained considering the scale dependent or independent have been compared. In the rest of the analysis the two scale will be considered to vary independently.

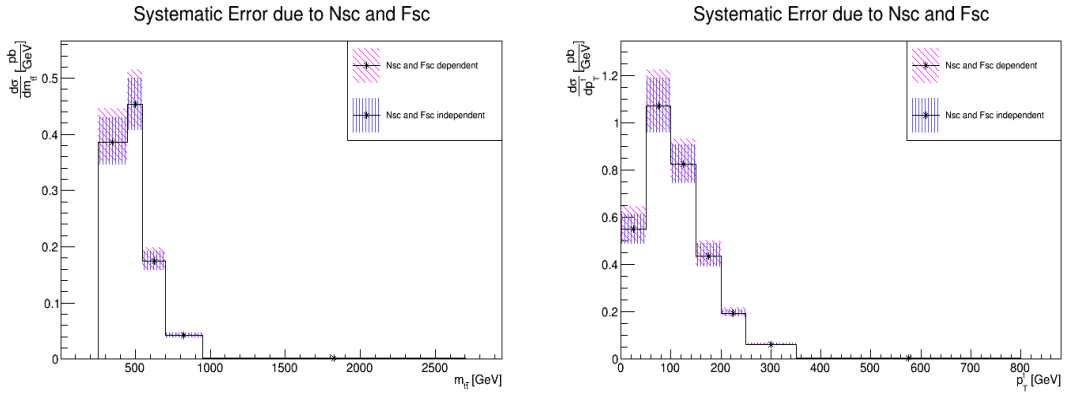

Figure 4.3: Distributions of the $m_{t\bar{t}}$ (left) and of p_T of hadronic top (right) predicted using MCFM. Error are derived varying the two scale Nsc and Fsc, dependently and independently.

The calculation of the PDF and scale systematic uncertainties has been performed in different steps:

- Firstly I evaluated the ΔX_i due to each MSTW2008NLO distribution and to the scale variations in each bin as:

$$\Delta X_{ij} = 0.5 \cdot |X_{i+} - X_{i-}| \tag{4.1}$$

where X_{ij} are the predictions of a particular kinematic value in the bin j and i runs on all 20 pairs of up and down errors sets and on the two scale contributions. For the factorization and the normalization scale X_{i+} and

X_{i-} correspond to the prediction obtained fixing one of the two scale at the nominal value and varying the other to $2m_t$ and to $m_t/2$.

- A set of 5000 pseudoexperiments have been performed; every "toy" spectrum was drawn by varying the nominal spectrum of the PDF set by a shape corresponding to:

$$X_j = \sum_i Rdm_i * \Delta X_{ij} + NomValue_j \tag{4.2}$$

where $NomValue_j$ is the central distribution value in the bin j and Rdm_i is a random number extract from a N(0, 1), the same for all bins, to consider bin correlation. The integer i varies on all the 22 ΔX.

- The error in each bin has been considered as the standard deviation of values X_j obtained from all Monte Carlo experiments. From these experiments the covariance and correlation matrix reported in figure 4.4 have been evaluated.

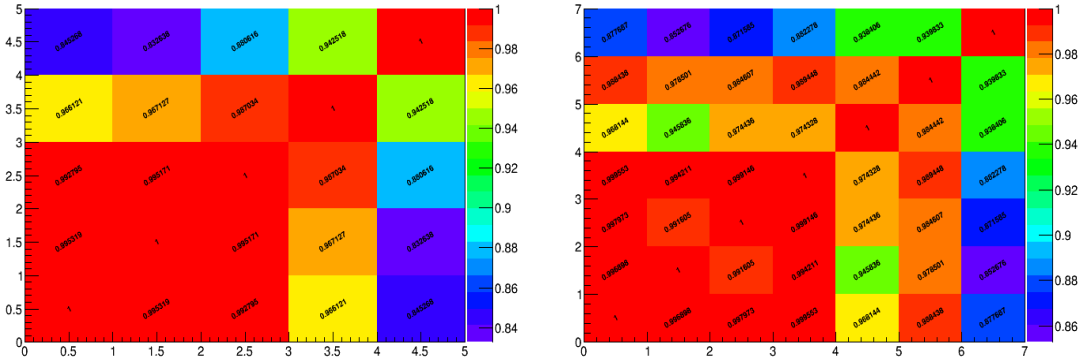

Figure 4.4: Matrices of correlation for distributions of $m_{t\bar{t}}$, on the left, and of p_T of hadronic top. The numbers on the axis represent the bin numbers and the off diagonal elements represent the bin correlations.

4.2 χ^2

The level of agreement between the measured distributions and theoretical predictions was quantified by calculating χ^2 values. In case of histograms the χ^2 can be performed using vector formalism:

$$\chi^2 \equiv (data - theory)^T \cdot [Cov_{data,syst}]^{-1} \cdot (data - theory) \tag{4.3}$$

where the vectors data and theory are formed by the difference between data and expectations and $Cov_{data,syst}$ is the covariance matrix obtained by the sum of the covariance matrix obtained from data and the one of systematic errors, calculated as illustrated in the previous section. The data covariance matrix is calculated performing 5000 pseudoexperiment, as in the PDF case, using as ΔX_i the uncertainties calculated from all systematic error source.

The results of 7 TeV analysis are presented normalized but this imply a complication in the calculation of χ^2. The normalization constraint, indeed, lowers by one unit the number of degrees of freedom of the system that is no more equal to the number of bins of the spectrum under consideration (N_{bin}). Consequently the covariance matrix calculated from normalized distributions become singular because it's dimension is higher than its rank and can't be inverted. An efficient solution to perform the χ^2 is to remove one value from data and theory vectors and to remove also the corresponding line and column from the covariance matrix. The sub-matrix obtained in this way is invertible and allows the computation of the χ^2 as:

$$\chi^2 = (data - theory)^T_{N_{b-1}}[Cov_{Nb_1}]^{-1}(data - theory)_{Nb-1}. \tag{4.4}$$

The value of χ^2 obtained in this way is consistent since it doesn't depend on the element discarded from the $(data - theory)$ vector and from the covariance matrix.

4.3 Results

In this section the results on the differential cross section measurement at $\sqrt{s}$=7 TeV and correspondent systematic error are presented; this are compared with predictions obtained using MSTW2008NLO PDF and other PDF sets. Figure 4.5 shows the normalized differential cross section with respect to the $m_{t\bar{t}}$ and the hadronic top p_T measured from data collected in 2011 from ATLAS with an integrated luninosity of $L = 4.6fb^{-1}$. In the same plot it is also reported the MC prediction obtained using the MSTW2008NLO PDF set and factorization and normalization scale fixed at m_t. Both the distributions are reported with errors calculated from Monte Carlo toys as illustrated in the previous section.

These studies can be extended to different PDF sets in order to study the sensitive of the $t\bar{t}$ differential cross section data in discriminating among the different PDF sets. Figure 4.6 shows the ratios between the measured $t\bar{t}$ differential cross section and several PDF sets for the mass of the $t\bar{t}$ system and the p_T of the top.

All the PDF sets describe reasonably the data in the low mass p_T range while there is a disagreement between data and expectations in the high-mass and high-p_T tail. This tension between differential cross section and theoretical predictions is under investigation and the boosted analysis will focus precisely to this high energy region in order to better understand this possible discrepancy.

The PDF set that better describe the data is the HERAPDF1.5.

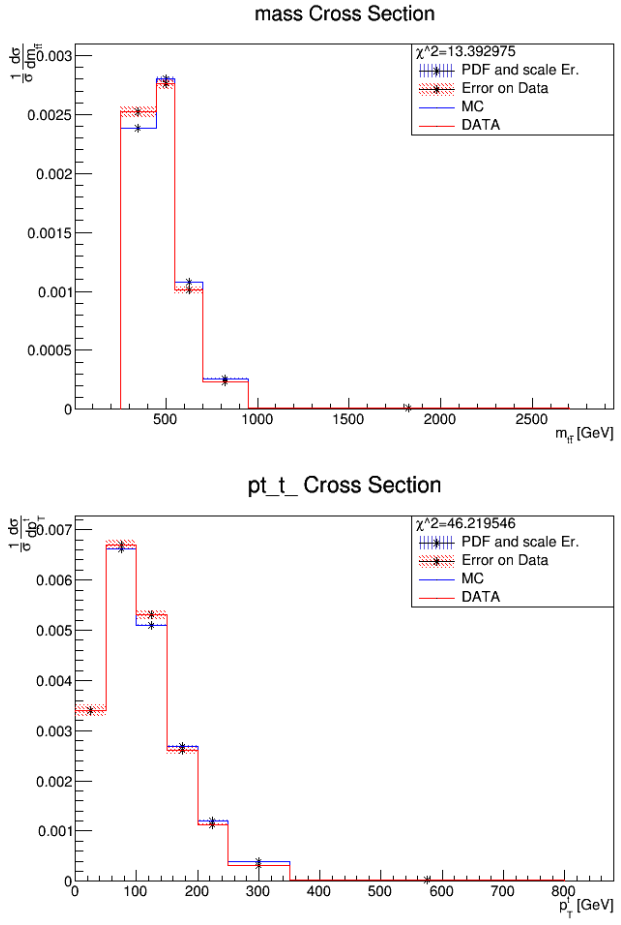

Figure 4.5: Differential cross section calculated from data (black line) and from MC predictions (red line). Both distributions are reported with systematic error and on the top of the legend is reported the chi^2, calculated as previously described.

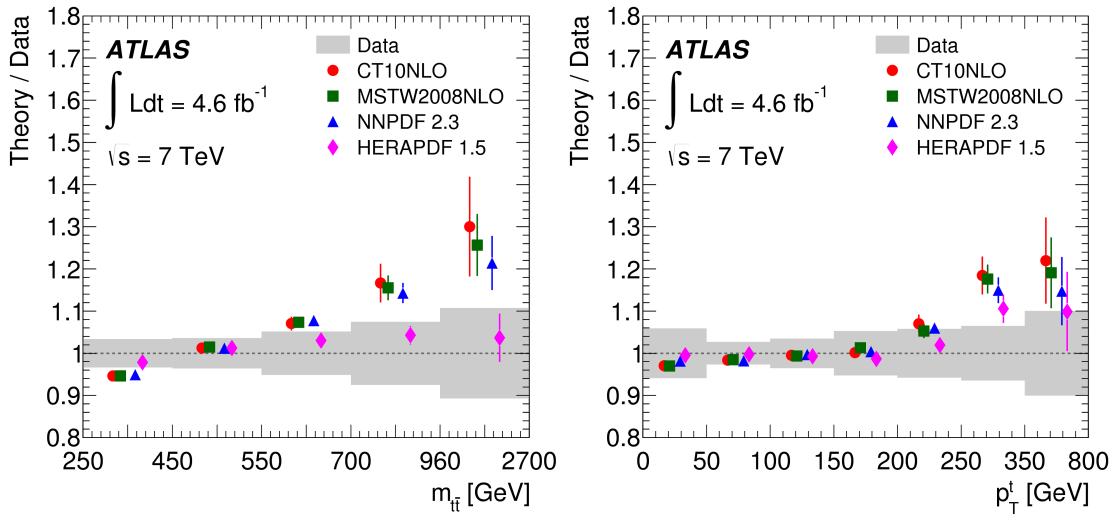

Figure 4.6: Ratio between several PDF set and observed data for the $t\bar{t}$ differential cross-section with respect to the $m_{t\bar{t}}$, on the left, and the hadronic top p_T^t, on the right. The gray bands indicate the total uncertainty on the data in each bin, while the error bars denote the uncertainties in the predictions, which include the internal PDF set variations and also scale uncertainties

Chapter 5

Results

In this Chapter the measurement of the $t\bar{t}$ production differential cross section with respect to the hadronic top p_T obtained with the full 2012 luminosity of 20.3 fb^{-1} at a $\sqrt{s}$ = 8 TeV are presented. In Paragraph 5.1 and 5.2 the techniques used in order to estimate the statistical and systematic uncertainties affecting these measurements are described. The differential cross section results are described in Paragraph 5.3.

5.1 Statistical Uncertainties

The uncertainty due to finite data statistics is determined using the collected data via a pseudo experiment method. One random weight, taken from a Poisson distribution with variance equal to the square root of the number of events, is assigned to each data event. Then the weighted events are used to produce a thousand replicas of data spectra. The distributions are then unfolded using the procedure described in Paragraph 3.6 obtaining a thousands of slightly different cross section distributions. These are used to construct the covariance matrix, whose diagonal elements are the square of the bin statistical errors. The means of the pseudo experiments results is consistent with the *nominal* (the original distribution, not shifted or re-weighted) data result, indicating no biases.

5.2 Systematic Uncertainties

The systematic uncertainties represent the errors on the measurement arising from object reconstruction and calibration, MC generator modeling and background es-

timation. Each systematic uncertainty is evaluated by varying the nominal distribution by a standard deviation on the considered error source, leading to a shifted distribution with respect to the nominal one. The uncertainty in each bin is calculated as the difference between the nominal and the shifted distribution, both considered after the event selection and the unfolding procedure. In this section the principal contributions to the total uncertainty are presented.

One of the main source of systematic uncertainty derives from the luminosity measurement; for the 2012 data set the total luminosity uncertainty is estimated to be the 2.8% [58].

The dominant contribution to the total uncertainty is given from large-R jet reconstruction. This uncertainty on is provided by the tool UJUncertaintyProvider [59] (recommended by the ATLAS top group) and include uncertainties on the jet energy scale (JES), the mass scale (JMS) and the $\sqrt{d_{12}}$ scale as a function of the large-R jet p_T. The uncertainty is derived using two different *data-driven* methods and the total uncertainty of the large-R jet ranges from $\sim 12\%$ to $\sim 20\%$ depending on the p_T.

Jet energy scale uncertainty (JES) for calibration of R=0.4 jets is calculated following the prescription given by the common ATLAS top working group [60]. This is derived combining results from simulations, test beams and in situ measurements. Additional contributions are given by the jet flavour composition and pile-up corrections.

The lepton reconstruction uncertainty includes contributions from the trigger efficiency, reconstruction efficiency and resolution uncertainty [61]. It has been found that this uncertainty has a small impact on the final measurement.

The efficiency of the b-tagging algorithm on real and fake b quarks jet is corrected in Monte Carlo events by applying a scale factor, extracted from dijet samples using *data-driven* techniques. The associated systematic uncertainty is computed by varying the scale factors within their uncertainties; for larger transverse momentum an additional MC-based uncertainty is extrapolated. The b-tagging contribution on the total systematic uncertainty runs from $\sim 2\%$ to $\sim 11\%$ depending on the bin considered.

The systematic uncertainties affecting the choice of the $t\bar{t}$ MC generator are evaluated comparing the distributions produced with different generators and considering the maximum shift between them and the nominal one. The MC generators considered in this calculation are: Powheg + Pythia, Powheg + Herwig, Alpgen + Herwig and MC@NLO + Herwig.

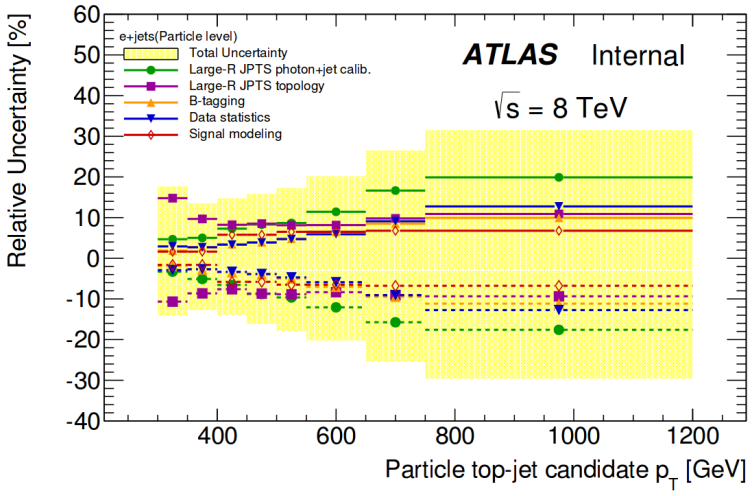

Figure 5.1: The main systematic uncertainties sources are reported as the relative percentage on the unfolded spectrum with respect to the hadronic top p_T, electron channel.

The systematic uncertainty accounting for the finite number of simulated $t\bar{t}$ events (*MC statistics*) is obtained using the *bootstrap method*; it consists on taking one random weight from a Poisson distribution with mean equal to one, to be assigned to each signal MC event. The weighted events are used to produce a thousand replicas of the truth and the reconstructed spectra. The data are now unfolded using the new MC samples and following the entire procedure described in Paragraph 3.6, obtaining a thousands of slightly different cross section distributions. These are used to construct the covariance matrix, whose diagonal elements are the square of the bin errors. This approach provides an estimate of the MC statistical uncertainty that properly accounts for all the statistical correlations between the unfolding, the efficiency and the acceptance. The resulting uncertainty ranges from $\sim 1\%$ (at low p_T) to 4% (in the highest p_T bin).

The backgrounds statistics uncertainties generally have a small impact on the final measurement error, main contributions derive from error calculated on *data-driven* backgrounds. The W+jets systematic uncertainty is evaluated varying the *data-driven* scale factors within their uncertainties. The fake lepton background uncertainty, instead, is estimated by varying the definition of loose leptons used in the background calculation.

In figure 5.1 and 5.2 are reported the main systematic uncertainties breakdown as relative percentage with respect to the total unfolded top quark p_T spectra for the electron and muon channel respectively. In the same figures are reported also the total systematic uncertainties, as yellow band.

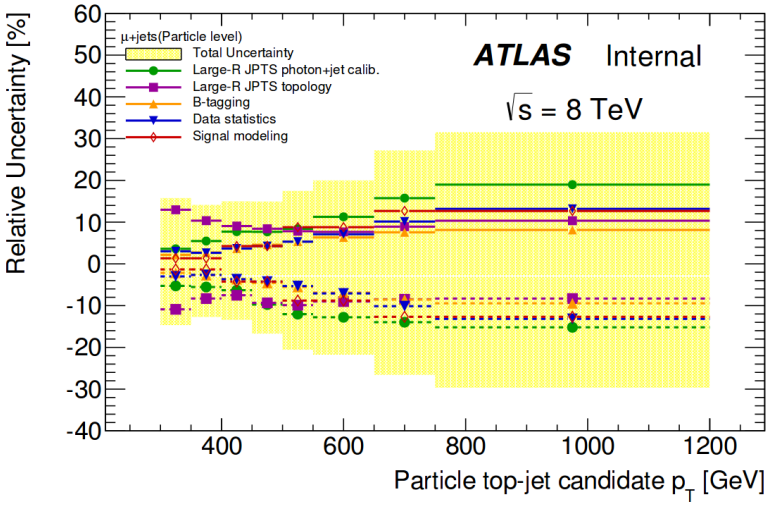

Figure 5.2: The main systematic uncertainties sources are reported as the relative percentage on the unfolded spectrum with respect to the hadronic top p_T, muon channel.

5.3 Differential cross section results

The $t\bar{t}$ production differential cross-section is measured as a function of the p_T of the particle-level top quark candidate defined as the leading p_T trimmed large-R jet; the cross section evaluation has been done in a specific fiducial region as defined in Paragraph 3.4.1. The specific selection tuned for boosted objects allow the measurement of the top quark p_T spectrum up-to 1200 GeV and to cover three orders of magnitude in cross section.

The differential cross section is obtained from the measured number of $t\bar{t}$ events through several steps, described in Chapter 3, that could be summarized in the following formula:

$$\frac{d\sigma_{t\bar{t}}}{dp^i_{T,corr}} = \frac{N^i_{part}}{\Delta X^i L f^i_{eff}} = \frac{1}{\Delta X^i L f^i_{eff}} \sum_j (M^{-1})^{part,i}_{reco,j} f^j_{acc} (N^j_{reco} - N^j_{reco,bkg}) \quad (5.1)$$

where $\frac{d\sigma_{t\bar{t}}}{dp^i_{T,corr}}$ is the final differential cross section with respect to the hadronic top quark p_T and N^i_{part} is the total number of events in the bin i satisfying the particle level selection requirements described in Paragraph 3.4.1. The N^j_{reco} represents the number of observed $t\bar{t}$ events in the bin j while $N^j_{reco,bkg}$ are the MC and *data-driven* estimated background, events normalized to the data integrated luminosity. The $(M^{-1})^{part}_{reco}$ factor is the inverted response matrix used in the SVD unfolding procedure described in Section 3.6.1, f^j_{acc} is the bin per bin acceptance , defined in

Paragraph 3.6.2. The result of the summation over all the bins j is then divided by the integrated luminosity L the efficiency f^i_{eff} defined in Paragraph 3.6.2 and the bin width ΔX^i.

The acceptance and efficiency distributions obtained with respect to the hadronic top p_T in the electron and muon channels are reported in figure 5.3 and 5.4.

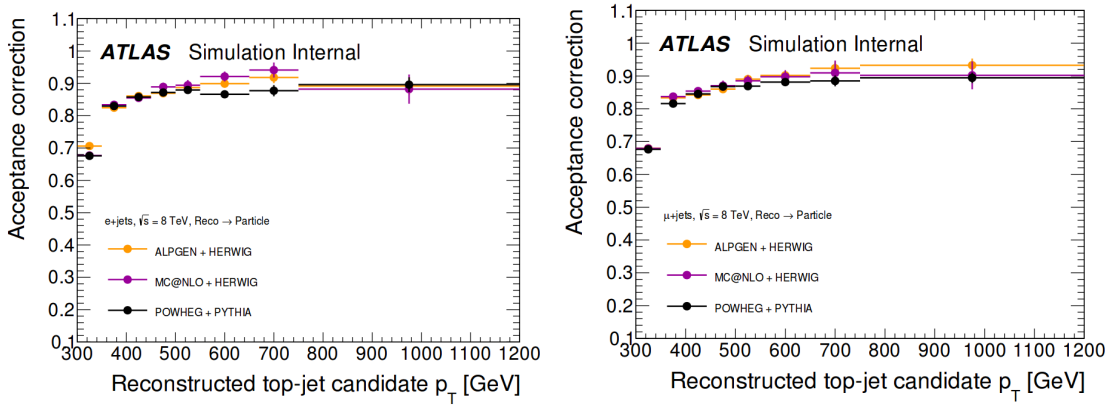

Figure 5.3: Acceptance with respect to the p_T of the reconstructed hadronic top in electron channel (left) and in muon channel (right)

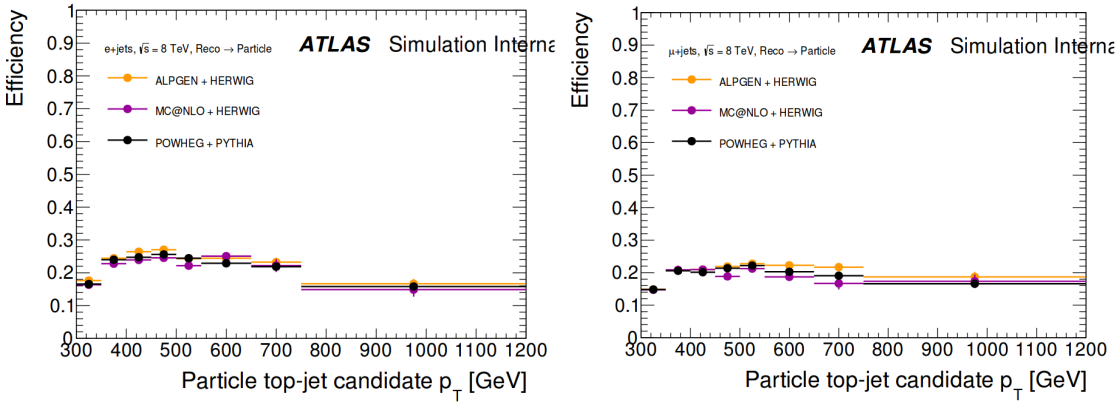

Figure 5.4: Efficiency with respect to the p_T of the particle top large-R jet in electron channel (left) and in muon channel (right)

5.3.1 $d\sigma/dp_T$ in electron and muon channel

In figures 5.5 and 5.6 the $t\bar{t}$ production differential cross section results are shown respectively for the electron and the muon channels. On the bottom part of the histograms is shown the ratio between the MC predictions and the results obtained from data. The obtained spectra are compared with four different MC predictions:

Powheg+Pythia, Powheg+Herwig, MC@NLO+Herwig and ALPGEN+Herwig. The results found are detailed bin per bin in table 5.1 and 5.2 respectively for the electron and muon channels.

The theoretical predictions from Powheg+Pythia, Powheg + Herwig and MC@NLO + Herwig generators are quite similar and their values differ from the final results of about 10-20 % or more depending from bin to bin. The discrepancy gets worst with increasing the top quark p_T value; here the predicted values are often outside the total uncertainty band. The ALPGEN + Herwig prediction, instead, completely disagrees with the found result in each bin.

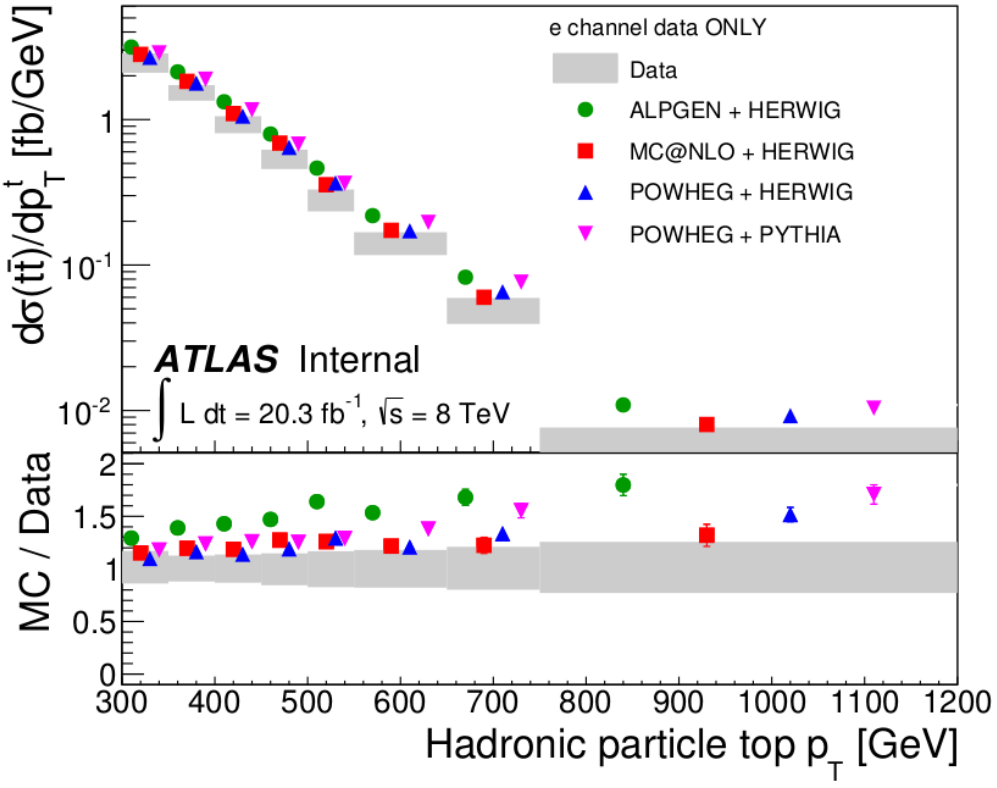

Figure 5.5: The differential cross section of $t\bar{t}$ in el+jet decay channel with respect to the hadronic top p_T, the data are reported as gray band including systematic and statistic error, the measured spectrum is compared to several theoretical expectations. On the bottom part the ratio between MC predictions/data is reported.

p_T^t [GeV]	$d\sigma_{t\bar{t}}/p_T^t$ [fb/GeV]	Stat.Unc.[%]	Total Unc.[%]
300-350	2.46	±2.9	±15
350-400	1.54	±2.7	±13
400-450	0.91	±3.3	±14
450-500	0.53	±3.9	±16
500-550	0.30	±4.7	±18
550-650	0.14	±5.9	±20
650-750	0.047	±9.1	±25
750-1200	0.0065	±12.7	±29

Table 5.1: Cross section obtained in each hadronic top p_T^t bin in the el+jet channel, on the third and fourth column are reported the statistical and total uncertainties.

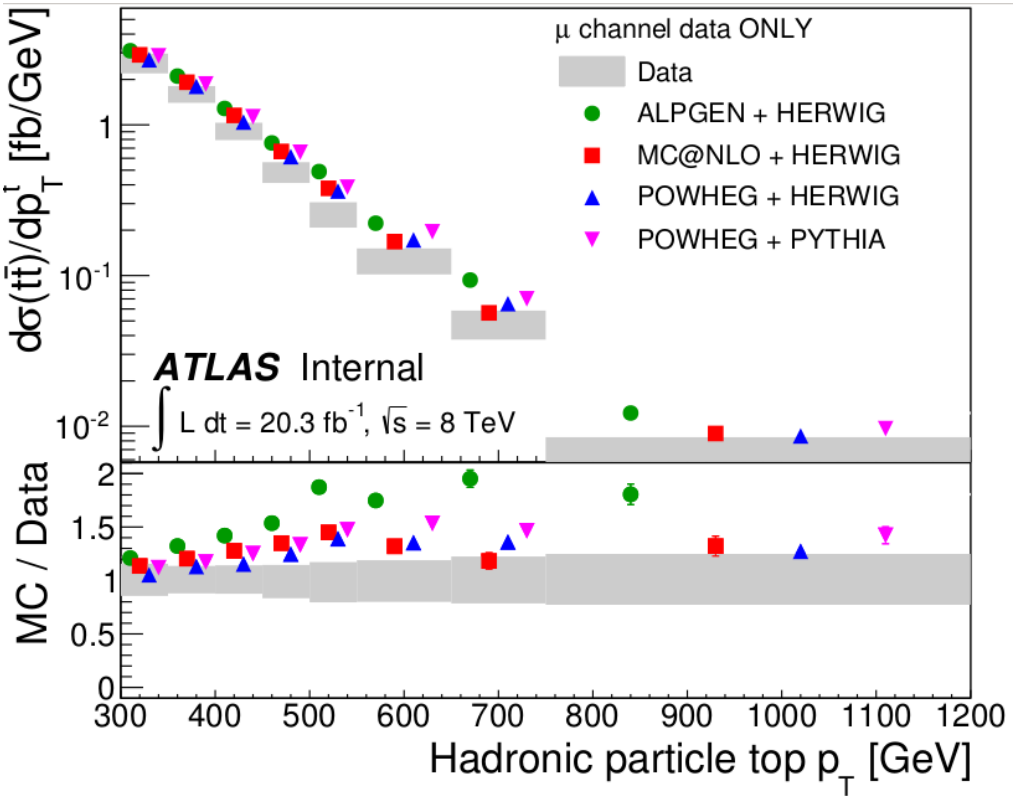

Figure 5.6: The differential cross section of $t\bar{t}$ in mu+jet decay channel with respect to the hadronic top p_T, the data are reported as gray band including systematic and statistic error, the measured spectrum is compared to several theoretical expectations, on the bottom part is reported the ratio between MC predictions/data.

p_T^t [GeV]	$d\sigma_{t\bar{t}}/p_T^t$ [fb/GeV]	Stat.Unc.[%]	Total Unc.[%]
300-350	2.57	±3.0	±15
350-400	1.6	±2.6	±13
400-450	0.9	±3.7	±14
450-500	0.5	±4.3	±16
500-550	0.27	±5.3	±18
550-650	0.12	±7.1	±20
650-750	0.045	±10.1	±25
750-1200	0.0065	±13.2	±29

Table 5.2: Cross section obtained in each hadronic top p_T^t bin in the mu+jet channel, on the third and fourth column are reported the statistical and total uncertainties.

5.3.2 $d\sigma/dp_T$ in lepton+jet channel

The channel-combination cross section result is performed by adding the events from the electron and muon channels both in data and MC at the reconstructed level. The unfolding procedure has then been applied to find the final cross section values for $t\bar{t}$ pairs decaying in the semileptonic (only electrons and muons) channel. The cross section is measured in a fiducial phase space corresponding to the sum phase spaces in the separated channels. This method is motivated by the fact that the event distributions at the particle level are comparable between the two channels within the statistical uncertainty. The uncertainties are evaluated on the summed sample as described in Paragraph 5.2, properly accounting for uncertainties correlations.

The resulting combined cross section is reported in table 5.3 and figure 5.7 [62]. The observed distributions are compared with Powheg + Pythia, Powheg + Herwig, MC@NLO + Herwig and ALPGEN + Herwig predictions as described in the previous Paragraph.

The agreement between data and MC is similar to the individual channels: the disagreement generally increases with the top quark p_T, reaching a MC/data ratio of approximately 1.3 for MC@NLO+Herwig and Powheg+Pythia and of 1.5 for Powheg+Pythia. The discrepancy reaches its highest value across the full p_T range with ALPGEN+Herwig.

At the moment this discrepancy is under investigation to understand if it is an

effect rising from a not perfect knowledge of the detector or of a not understood or new physics process.

p_T^t [GeV]	$d\sigma_{t\bar{t}}/p_T^t$ [fb/GeV]	Stat.Unc.[%]	Total Unc.[%]
300-350	5.0	±2.9	±15
350-400	3.1	±2.7	±13
400-450	1.8	±3.3	±14
450-500	1.0	±3.9	±16
500-550	0.57	±4.7	±18
550-650	0.26	±5.9	±20
650-750	0.093	±9.1	±25
750-1200	0.013	±13	±29

Table 5.3: Cross section obtained in each hadronic top p_T^t bin in the lepton+jet channel, on the third and fourth column are reported the statistical and total uncertainties.

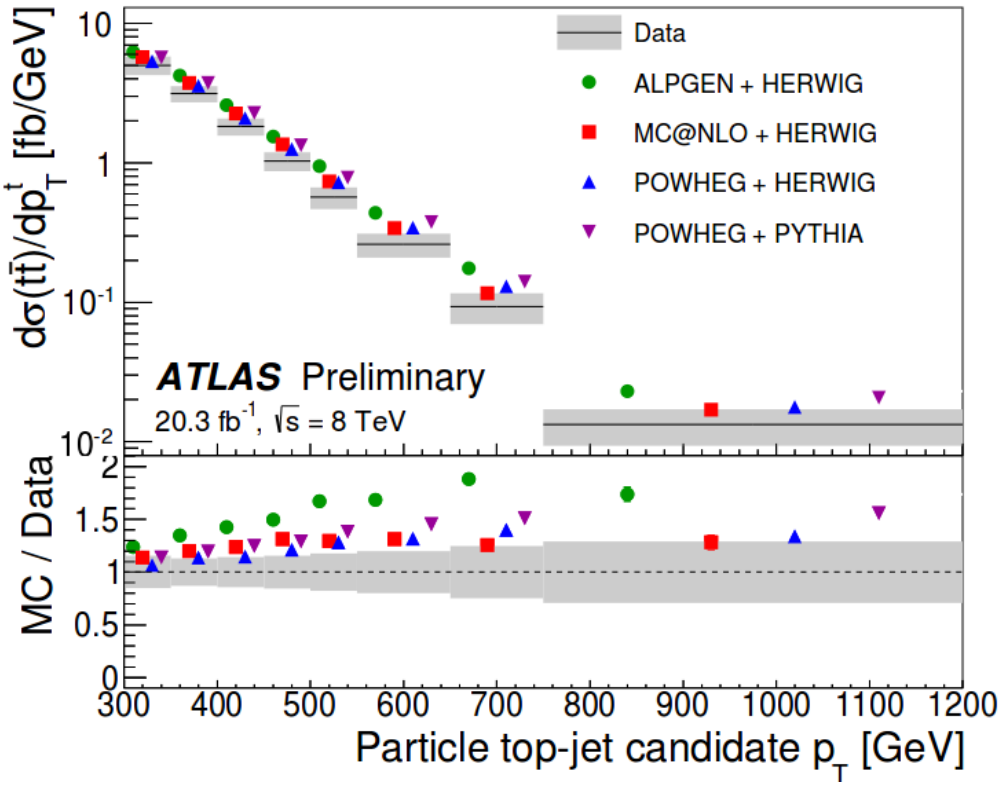

Figure 5.7: The differential cross section of $t\bar{t}$ in lepton+jet decay channel with respect to the hadronic top p_T, the data are reported as gray band including systematic and statistic error, the measured spectrum is compared to several theoretical expectations, on the bottom part is reported the ratio between MC predictions/data.

Conclusions

In this thesis the measurement of the $t\bar{t}$ production differential cross section at $\sqrt{s}$= 8 TeV has been presented as a function of the hadronic top p_T. The analysis is performed on 2012 data, corresponding to an integrated luminosity of L=20.3 fb^{-1}. The events selected contain $t\bar{t}$ decaying in the semileptonic channel and the measurement is performed in a high momentum phase space region ($p_T^t > 300$ GeV) where the top quarks are highly boosted and new techniques were needed for its reconstruction. This analysis measures a part of the quark top p_T spectrum never studied before and it is the first differential $t\bar{t}$ production differential cross section measurement performed at $\sqrt{s}$= 8 TeV. The results could be used to improve the theoretical simulations in this new phase-space region.

In this analysis the cross section is measured in a fiducial phase space (particle-level) specifically designed to be slightly dependent from different theoretical models, to reduce the detector dependencies and the systematic uncertainty. On the final measurement the total uncertainty ranges between approximately $15\% \div 29\%$ and is generally dominated by the large-R jets energy scale systematic source. The differential cross section measured has been compared with several MC generators and the measured spectra have shown a not satisfactory agreement with the predictions. All the theoretical expectations overestimate the measured cross section and the discrepancy generally increases with the p_T reaching approximately 30% to 70% in the highest-p_T bin, depending on the MC generator considered. At the moment this discrepancy is under investigation.

I'm currently one of the analysis code developer and I collaborated to all the analysis chain implementing codes involved in event selection, object reconstruction, error evaluation and Data/MC comparison, aside to perform a study on leptonic top reconstruction and b-tagging efficiency. All the details of the analysis have been published on a ATLAS note [62] of which I'm one of the authors.

The distributions of kinematic variable of the $t\bar{t}$ pair are really sensitive to the PDF set used in the simulation. In particular I have performed a study on the

differential cross section measured at $\sqrt{s}$=7 TeV comparing several PDF sets in order to understand which one better describes the measured spectra.

The results reached in this analysis will be start point for future studies and could give a contribution both from a physical point of view than from a technical one:

- Due to the proximity of the mass of the top quark to the electroweak scale, many extensions of the Standard Model (SM) give the top quark a more fundamental role in nature. Furthermore the rich phenomenology of top quarks, which involves several high-p_T jets, including b-jets, leptons and missing transverse energy $\not{E}_T$, makes top quark production a dominant background for a large class of searches for new physics. Consequently is fundamental to measure the $t\bar{t}$ differential cross section with high precision.

- The results of this analysis could be used to perform a study similar to the one I showed in this thesis on $\sqrt{s} = 7$ TeV measurement and the results could be included in the PDF global fit to provide new constraint to gluons PDF in the high momentum region.

- The next LHC run will collide protons at $\sqrt{s}$ =13 TeV and a large amount of particles at high p_T will be produced. In this new configuration the standard techniques won't lead to satisfying results and the boosted techniques and reconstruction algorithms will become the only way to measure top quarks and other heavy objects.

Bibliography

[1] Beringer J. et al. (Particle Data Group), *Phys. Rev. D86*, 010001, **2012**

[2] Higgs P. W., *Phys. Rev. Lett. 13*, **1964**, 508

[3] Englert F. and Brout R., *Phys. Rev. Lett. 13* **1964**, 321

[4] ATLAS Collaboration, *Phys.Lett. B*, **2012**, *716*, 1-29

[5] CMS Collaboration, *Phys. Lett. B*, **2012**, *716*, 30

[6] Halzen F. Martin A.D., *Quarks and leptons: An introductory course in modern particle physics*, John Wiley & Sons, Inc.,USA **1994**, pp. 82-85

[7] Perkins H. D., *Introduction to High Energy Physics*, Cambridge University Press, USA **2000**, pp. 181-187

[8] Bethke S., *Nuclear Physics B - Proceedings Supplements*, **2012**, *234*, 229-234

[9] Andrzej J. Buras,*Flavour changing neutral current processes, hep-ph/9610461*, **2006**

[10] F. Abe, et al. (CDF collaboration), *Phys. Rev. Lett. 74*, **1995**, 2626

[11] S. Abachi, et al. (DØ collaboration), Phys. Rev. Lett. 74, **1995**, 2632

[12] T.M. Liss, F. Maltoni and A. Quadt, The Top quark, *Phys. Rev. D86*, **2013**

[13] ALEPH and CDF and D0 and DELPHI and L3 and OPAL and SLD Collaborations, LEPEWWG, TEVEWWG, SLD EW and HF groups, *Precision Electroweak Measurements and Constraints on the Standard Model*, **2010**

[14] N. Kidonakis, *Top Quark Theoretical Cross Section, and p_T and Rapidity Distributions*, Proceedings of the DPF-2011 Conference, **2011**, Providence

[15] F. P. Schilling, *Top quark physics at the LHC: a review of the first two years*, International Journal of Modern Physics A, edited by World Scientific Publishing Company, **2012**

[16] N. Kidonakis , *Phys. Rev. D83*, 091503, **2011**

[17] N. Kidonakis , *Phys. Rev. D82*, 054018, **2010**

[18] N. Kidonakis , *Phys. Rev. D81*, 054028, **2010**

[19] The ATLAS Collaboration, *ATLAS-CONF-2012-132*, **2012**

[20] The CMS Collaboration, *CMS-TOP-12-038-003*, **2014**

[21] The ATLAS Collaboration, *ATL-PHYS-PROC-2013-276*, **2013**

[22] A. Martin, W. Stirling, R. Thorne and G. Watt, *Eur. Phys. J. C63*, **2009**, 189

[23] H1 and ZEUS Collaborations, F. Aaron et al., *JHEP 1001*, **2010**, 109

[24] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky et al., *Phys. Rev. D82*, 074024, **2010**,

[25] NNPDF Collaboration, R. D. Ball et al., *Nucl. Phys. B855*, **2012**, 153

[26] CDF Collaboration, *Conf. Note 9913*, **2011**

[27] The D0 Collaboration, *Fermilab-Pub-11-233-E*, **2011**

[28] The ATLAS Collaboration, *ATLAS-CONF-2011-100*, **2011**

[29] The CMS Collaboration, *CERN-PH-EP/2011-055 2011/05/31*, **2011**

[30] The ATLAS Collaboration, *ATLAS-CONF-2012-149*, **2012**

[31] The CMS Collaboration, *CMS PAS TOP-12-006*, **2012**

[32] A. W. Jung, *FERMILAB-PUB-14-071-PPD*, **2014**

[33] The ATLAS Collaboration, *ATLAS-CONF-2013-099*, **2013**

[34] The CMS Collaboration, *Eur. Phys. J. C 73:2339*, **2013**

[35] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST, 3, S08003,**2008**

[36] The ATLAS Collaboration, *ATLAS pixel detector electronics and sensors* JINST, 3, P07007,**2008**

[37] The ATLAS Collaboration, *Liquid Argon Calorimeter-Technical Design Report*, CERN-LHCC-96-041, **1996**

[38] The ATLAS Collaboration, *Eur. Phys. J. C 73:2518*, **2013**

[39] GEANT4 Collaboration (S. Agostinelli et al.), *Nucl.Instrum.Meth. A506 250-303*, **2003**

[40] S. Frixione, P. Nason, and C. Oleari, *JHEP 0711 070*, **2007**

[41] T. Sjostrand, S. Mrenna, and P. Skands, *JHEP 05 026*, **2006**

[42] S. Frixione, B. R. Webber, and P. Nason, *hep-ph/0204244 and hep-ph/0305252*

[43] G. Corcella et al., *JHEP 01 010*, **2001**

[44] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Z. Phys. C 72 637–646*, **1996**

[45] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *JHEP 07 001*, **2003**

[46] M. Czakon and A. Mitov, *arXiv:1112.5675*, **2013**

[47] The ATLAS Collaboration, *ATLAS-CONF-2013-052*, **2013**

[48] The ATLAS Collaboration, *Phys. Lett. B711 244–263*, **2012**

[49] T. Lagouri et al., *IEEE Transactions on Nuclear Science 51, 3030*, **2004**

[50] The ATLAS Collaboration, *ATLAS-CONF-2012-065*, **2012**

[51] Gavin P. Salam, *Towards Jetography*, arXiv:0906.1833v2, **2010**

[52] D. Krohn, J. Thaler, and L.-T. Wang, *JHEP 2010 20*, **2010**

[53] V. Blobel, *Unfolding methods in high-energy physics experiments*, *arXiv:hep-ex/0208022*, **2002**

[54] , M. Schmelling, *Nuclear Inst. and Methods in Physics Research, A 340 2, 400-412*, **1994**

[55] Andreas Höcker and Vakhtang Kartvelishvili, *SVD approach to data unfolding*, *MC-TH-95/15*,**1995**

[56] *https://svnweb.cern.ch/trac/atlasoff/browser/PhysicsAnalysis/TopPhys/TopD3PDBoosted/*

[57] J. Campbell, K. Ellis, and C. Williams, *MCFM - Monte Carlo for FeMtobarn Processes*, *http://mcfm.fnal.gov/*.

[58] The ATLAS Collaboration Collaboration, G. Aad et al., *Improved luminosity determination in pp collisions at* $\sqrt{s} = 7$ *TeV using the ATLAS detector at the LHC*, *arXiv:1302.4393 [hep-ex]*, **2013**

[59] *https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/UJUncertaintyProvider*

[60] *https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetUncertainties2012*

[61] The ATLAS Collaboration, *Tech. Rep. ATL-COM-PHYS-2013-1016*, **2013**

[62] Arguin J-F, Berta P., Fabbri F., Franchini M., Gabizon O., Leitner R., Massa L., Mengarelli A., Negrini M.; Ovcharova A., Shapiro M., Spanò F., Yao W. et The ATLAS Collaboration, *ATLAS-COM-CONF-2013-113*, **2014**